UNIVERSIDADE
BEIRA INTERIOR

Production Optimization Indexed to the Market Demand Through Neural Networks

Balduíno Patrício César Mateus

Tese para obtenção do Grau de Doutor em

Engenharia e Gestão Industrial

(3º ciclo de estudos)

Setembro de 2023

Production Optimization Indexed to the Market Demand Through Neural Networks

Balduíno Patrício César Mateus

Tese para obtenção do Grau de Doutor em

Engenharia e Gestão Industrial

(3º ciclo de estudos)

Orientador: Prof. Doutor António João Marques Cardoso
Co-orientador: Prof. Doutor José Torres Farinha

Provas de Doutoramento: 26 de julho de 2023

Júri:

PRESIDENTE:
Prof. Doutor Joaquim Mateus Paulo Serra

VOGAIS:
Prof. Doutor Diego Galar
Prof. Doutor Luís António de Andrade Ferreira
Prof. Doutor Luís António Fonseca Mendes
Prof. Doutora Isabel da Silva Lopes

Setembro de 2023

Declaração de Integridade

Eu, Balduíno Patrício César Mateus, que abaixo assino, estudante com o número de inscrição D2736 de/o Engenharia e Gestão Industrial da Faculdade da Engenharia, declaro ter desenvolvido o presente trabalho e elaborado o presente texto em total consonância com o **Código de Integridades da Universidade da Beira Interior**.

Mais concretamente afirmo não ter incorrido em qualquer das variedades de Fraude Académica, e que aqui declaro conhecer, que em particular atendi à exigida referenciação de frases, extratos, imagens e outras formas de trabalho intelectual, e assumindo assim na íntegra as responsabilidades da autoria.

Assinado por: **BALDUÍNO PATRÍCIO CÉSAR MATEUS**
Num. de Identificação: 32037376
Data: 2023.09.04 11:26:42+01'00'

Universidade da Beira Interior, Covilhã 04/ 09/ 2023

(assinatura conforme Cartão de Cidadão ou preferencialmente
assinatura digital no documento original se naquele mesmo formato)

Dedication

Life is full of difficulties, from our birth to our disappearance on this earthly plane. In the midst of these obstacles, there are always people who play the role of guardians and help you to have a broader vision to lift you to your unknown level. Intelligence is also learnable, but few people want to pass it on to others, but I thank God that I have met great academic and life instructors who have helped me reach that level. For this, I would like to thank these people from the bottom of my heart.

On a family level: my thanks go to God who made my existence possible; my thanks go to my parents who taught me the sacrifice of survival and adaptation. I also want to thank my beloved wife, my daughter Siara and my son Adam for giving me a unique and pure love that is a fuel for survival and perseverance in this adventure. To my sister Rosa, thank you for believing in my potential and for being an exemplary and great sister. To my dear nephew Fabio Marques, thank you for being a kind nephew and advisor to me. I would like to thank all the nephews, uncles, aunts, sisters and brothers of my family for the indirect help they have always given me during this journey.

On an academic level: my best wishes to my good professor José Torres Farinha, who pushed me beyond what I thought were my limits and was a great advisor, not only academically but also in life. Many thanks to Professor António J. Marques Cardoso for accepting the challenge to be my advisor. I am inwardly grateful for the immeasurable help he has given me. I would like to express my deepest gratitude to Professor Mateus Mendes for his great help and especially for providing me with knowledge that was important for the development of this thesis. To Professor Oliveira Santos, I thank you for the opportunity and for the trust placed in me and my colleagues that we are in the same boat. Thank you for your opinions and support. Thank you to Professor Rui Assis for the critical contributions that were essential to this work. I thank my colleagues Alexandre Martins, Edmundo Pais and João Rodrigues for being more than just colleagues (friends). Thanks to this fraternisation, it was possible to contribute pertinent ideas to the study.

Resumo

Conectividade, mobilidade e análise de dados em tempo real são pré-requisitos para um novo modelo de gestão inteligente da produção que facilita a comunicação entre máquinas, pessoas e processos, e usa a tecnologia como motor principal.

Muitos trabalhos na literatura tratam a manutenção e a gestão da produção em abordagens separadas, mas existe uma correlação entre estas áreas, sendo que a manutenção e as suas políticas têm como premissa garantir o bom funcionamento dos equipamentos de modo a evitar paragens desnecessárias na linha de produção.

Com o advento da tecnologia há uma corrida das empresas para solucionar os seus problemas recorrendo às tecnologias, visando a sua inserção nos conceitos tecnológicos, mais avançados, tais como as indústrias 4.0 e 5.0, as quais têm como princípio a automatização dos processos. Esta abordagem junta as tecnologias de sistema de informação, sendo possível fazer o acompanhamento do funcionamento dos equipamentos e ter a possibilidade de realizar o estudo de padrões de comportamento dos dados que nos possam alertar para possíveis falhas.

A presente tese pretende prever a produção da pasta de papel indexada às bolsas de valores. A previsão será feita por via das variáveis da produção da pasta de papel das prensas e das variáveis da bolsa de valores suportadas em tecnologias de *artificial intelligence* (IA), tendo como objectivo conseguir um planeamento eficaz. Para suportar a decisão de uma gestão da produção eficiente, na presente tese foram desenvolvidos algoritmos, validados em dados de cinco prensas de pasta de papel, bem como dados de outras fontes, tais como, de Produção de Aço e de Bolsas de Valores, os quais se mostraram relevantes para a validação da robustez dos modelos.

A presente tese demonstrou a importância dos métodos de tratamento de dados e que os mesmos têm uma grande relevância na entrada do modelo, visto que facilita o processo de treino e testes dos modelos. As tecnologias escolhidas demonstraram uma boa eficiência e versatilidade na realização da previsão dos valores das variáveis dos equipamentos, demonstrando ainda robustez e otimização no processamento computacional.

A tese apresenta ainda propostas para futuros desenvolvimentos, designadamente na exploração mais aprofundada destas tecnologias, de modo a que haja variáveis de mercado que possam calibrar a produção através de previsões suportadas nestas mesmas variáveis.

Palavras-chave

Manutenção, Produção, Redes Neuronais Recorrentes, Análise de Dados, Previsão.

Resumo alargado

Conectividade, mobilidade e análise de dados em tempo real são pré-requisitos para um novo modelo de gestão inteligente da produção, que facilita a comunicação entre máquinas, pessoas e processos, e usa a tecnologia como motor principal.

Actualmente vivemos na era digital, onde o principal foco é a digitalização das indústrias; um dos seus pilares corresponde à sensorização e aquisição de dados, visando a implementação generalizada da Manutenção, na vertente Preditiva, com o objetivo da maximização da Disponibilidade operacional dos equipamentos. Muitos trabalhos disponíveis na literatura tratam a manutenção e a gestão da produção em abordagens separadas; porém, existe uma correlação estreita entre estas áreas, sendo que a manutenção e as suas políticas têm como premissa garantir o bom funcionamento dos equipamentos, maximizando a sua disponibilidade, de modo a evitar paragens desnecessárias da produção.

A globalização enfatiza cada vez mais a importância de prever diferentes fenómenos que podem ocorrer e que têm implicações na competitividade dos mercados. Como resultado, as empresas procuram soluções tecnológicas incluindo algoritmos de previsão, que lhes permitam antecipar cenários e apoiar as decisões. Por outro lado, se estas previsões não forem bem-feitas, designadamente através de algoritmos de previsão, podem provocar uma falsa sensação de segurança e, por consequência, podem induzir perdas económicas.

Com o advento da tecnologia há uma corrida das empresas para solucionar os seus problemas recorrendo aos últimos avanços tecnológicos, visando a sua inserção nos conceitos tecnológicos mais avançados, tais como as indústrias 4.0 e 5.0, as quais têm como princípio a automatização dos processos. Esta abordagem é, em grande parte, suportada nas tecnologias de análise de dados, sendo possível fazer o acompanhamento do funcionamento dos equipamentos e ter a possibilidade de realizar o estudo de padrões de comportamento dos dados que possam alertar para possíveis falhas.

Ao longo do tempo surgiram novos desenvolvimentos no âmbito da Gestão de Operações, nomeadamente os que tornaram possível a compreensão e optimização dos sistemas de produção. O Japão, devido à sua cultura multidisciplinar, foi o país onde as indústrias foram mais bem-sucedidas na implementação destes novos desenvolvimentos. As tecnologias de automação, em combinação com a experiência humana, trazem vantagens significativas, levando a uma melhor resposta do mercado, proporcionando valor acrescentado à indústria, e visando uma maior flexibilidade na resposta aos desafios da economia. As tecnologias utilizadas para tal são suportadas em: IA; computação na nuvem; Big Data; realidade aumentada; IoT; automação robótica; impressão 3D; e nanotecnologia. Estas tecnologias precisam da experiência humana para que as suas respostas tenham uma validade robusta no âmbito do objectivo pretendido.

A presente tese apresenta abordagens que permitem prever a produção da pasta de papel indexada ao movimento das bolsas de valores, designadamente suportadas em tecnologias de IA, tendo como objectivo conseguir um planeamento eficaz. Para suportar a decisão de uma gestão da produção eficiente, no âmbito da presente tese foram desenvolvidos algoritmos, validados em dados de cinco prensas de uma indústria de pasta de papel, bem como de dados de outras fontes, tais como, de Produção de Aço e de Bolsas de Valores, os quais se mostraram relevantes para a validação da robustez do modelo.

O algoritmo de IA consegue realizar previsões multivariadas. Suportado no caso de estudo levado a efeito conseguiu-se fazer um *link* das variáveis das prensas da pasta de papel às variáveis da produção da pasta de papel, atendendo a que as mesmas apresentam uma correlação significativamente alta. Também se obtiveram correlações significativas entre as variáveis da produção da pasta de papel com os valores da bolsa, nomeadamente na referida empresa. Estas correlações reforçam a possibilidade de se conseguir implementar um modelo dinâmico que possa fazer uma leitura destes dados, tratá-los e usá-los como entrada do modelo preditivo, de modo a ter uma resposta de previsão que passa desde o departamento de manutenção ao de produção, com o objectivo de fazer um planeamento da manutenção em função da dinâmica do mercado.

A presente tese também demonstrou a importância dos métodos de tratamento de dados e que os mesmos têm uma grande relevância na entrada do modelo preditivo, visto que facilita o processo de treino e testes dos modelos. As tecnologias escolhidas apresentaram uma boa eficiência e versatilidade na realização da previsão dos valores das variáveis dos equipamentos, demonstrando ainda robustez e otimização no processamento computacional.

Para o tratamento de dados foi utilizada a biblioteca *Pandas* e outros recursos em linguegem *Python* que permitem manipular os dados de maneira mais eficiente. Para além destes métodos, também foram utilizados métodos clássicos de eliminação de *outliers*. Este tipo de tratamento de dados foi muito importante na análise de correlação e autocorrelação das variáveis, atendendo a que o mesmo permitiu demonstrar o quanto poderia ser possível esta relação entre as variáveis ou entre si mesmas.

A tese apresenta ainda propostas para futuros desenvolvimentos, designadamente na exploração mais aprofundada destas tecnologias, de modo a que haja variáveis de mercado que possam calibrar a produção através de previsões suportadas nestas mesmas variáveis. Atendendo a que a tecnologia tem vindo, cada vez mais, a ser aprimorada, tem-se como compromisso a sua aplicação nas políticas de manutenção preditiva, visando contribuir positivamente para a produção.

O contributo pode dar-se também na optimização dos processos de treino e teste das Redes Neuronais, uma vez que o mesmo afecta o tempo de processamento e a capacidade computacional. Minimizar ao máximo o fenómeno das explosões de gradiente também se torna um grande desafio, uma vez que, para taxas de amostragens elevadas, este fenómeno torna-se mais presente no processo de treino das Redes Neuronais.

O domínio da informação tornou-se um pilar na análise de padrões, quer na manutenção quer na produção, uma vez que, através do seu suporte, é possível traçar novas metas na gestão das empresas, cumprindo com os desafios acima citados; torna-se claro que a maximização da Disponibilidade dos Equipamentos e a racionalização dos *stocks* são objectivos exequíveis, tendo como resultado empresas mais competitivas e a diminuição dos desperdícios.

Abstract

Connectivity, mobility and real-time data analytics are the prerequisites for a new model of intelligent production management that facilitates communication between machines, people and processes and uses technology as the main driver.

Many works in the literature treat maintenance and production management in separate approaches, but there is a link between these areas, with maintenance and its actions aimed at ensuring the smooth operation of equipment to avoid unnecessary downtime in production.

With the advent of technology, companies are rushing to solve their problems by resorting to technologies in order to fit into the most advanced technological concepts, such as industries 4.0 and 5.0, which are based on the principle of process automation. This approach brings together database technologies, making it possible to monitor the operation of equipment and have the opportunity to study patterns of data behavior that can alert us to possible failures.

The present thesis intends to forecast the pulp production indexed to the stock market value.The forecast will be made by means of the pulp production variables of the presses and the stock exchange variables supported by artificial intelligence (AI) technologies, aiming to achieve an effective planning. To support the decision of efficient production management, in this thesis algorithms were developed and validated with from five pulp presses, as well as data from other sources, such as steel production and stock exchange, which were relevant to validate the robustness of the model.

This thesis demonstrated the importance of data processing methods and that they have great relevance in the model input since they facilitate the process of training and testing the models. The chosen technologies demonstrated good efficiency and versatility in performing the prediction of the values of the variables of the equipment, also demonstrating robustness and optimization in computational processing. The thesis also presents proposals for future developments, namely in further exploration of these technologies, so that there are market variables that can calibrate production through forecasts supported on these same variables.

Keywords

Maintenance, Production, Recurrent Neural Network, Data Analysis, Forecast.

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

ABR	Age Based Replacement
AE	Autoencoder
AI	Artificial Intelligence
ANN	Artificial Neural Network
ARMA	Autoregressive Moving Average
CM	Corrective Maintenance
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DT	Decision Tree
EDA	Event Oriented Architecture
GRU	Gated Recurrent Unit
HPP	Homogeneous Poisson Process
IoT	Internet of Things
IQR	Interquartile Range
JIT	Just in Time
LSTM	Long Short-term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MCO	Multi-Criteria Optimization
MLP	Multilayer Perceptron
MTTF	Mean Time To Failure
O&M	Operation and Maintenance
OM	Opportunistic Maintenance
PM	Preventive Maintenance
RBM	Restricted Boltzmann Machine
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SARIMA	Seasonal Autoregressive Integrated Moving Average
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
TPM	Total Productive Maintenance
WIP	Work in Progress

Chapter 1

Introduction

More and more studies prove the importance of having a more comprehensive view of the production system, specially related with maintenance management, as it brings added value to production systems. This chapter presents the link between maintenance management and production through existing studies, as well as contextualizing the problem studied in this thesis.

1.1 Background

Market information converted into data is an ally in production management, as from patterns in the data we can adjust the company's capacity with respect to production periods. Due to short lead times, production in small quantities, diversification of consumer needs, and irregular fluctuations in demand, companies show an interest in improving their internal processes, in order to achieve more flexible and anticipatory production [1, 2].

According to Mobley [3], the production capacity of a factory is limited in part by the availability of production systems and their utilities.

The level of adaptation of a competitive company means prioritizing key decisions related to structural and infrastructure investments, which are key to realizing its full operational potential as a competitive player [4]. According to Riis et al. [5], capacity the production process of a factory depends in part on the availability of production systems and their tools, and the main task of a maintenance team is to ensure that all equipment and systems in the factory are compliant and in good condition.

1.1.1 Data Processing

Modern processors, computers, and high-speed networks make it possible to collect, transmit, and store large amounts of data in real time. Collecting and combining data from various sensors provides an insightful look at the health of factories, industrial plants, and other facilities.

Information technologies such as Big Data, cloud computing, advanced computing and artificial intelligence tools can be used to create, store and process large data sets.

With the advent of new technologies, it is possible to monitor machine failures by means of predictive models supported by artificial intelligence. With machine history in data it is possible to perform an estimation of the next occurrence of faults with a reasonable degree of confidence.

Modern algorithms, data storage, and computing power make it possible, not only to analyze past behavior but also to predict future behavior of industrial equipment with reasonable certainty [6, 7, 8].

1.1.2 Maintenance

It is known that poor maintenance can lead to poor performance, causing breakdowns at inappropriate times and even poor performance. A plant manager does not want this to happen in his production line. Therefore, it can be said that maintenance is a fundamental pillar for the smooth running of the production line. The maintenance in a organization must ensure that all machinery, equipment, and plant systems are always on-line and in good working order.

Maintenance and production services must work together to achieve a common goal, which is to maximize the productivity of the plant [9]. Planning and maintenance play an extremely important role in production, as they can contribute to higher production efficiency and product quality.

According to [3], the production capacity of a factory is delimited in part by the availability of production systems and their utilities. The main role of the maintenance organization is to ensure that all equipment and plant systems are always on-line and in a good condition.

Maintenance costs can range from 15% to 70% of production costs [10]. With advances in industrial process technology, maintenance has also evolved and become more complex [10, 11]. This is particularly due to production systems that have numerous interactions and dependencies between components [12].

Predictive maintenance is designed to increase process efficiency and limit the optimal time window for maintenance work. With the help of sensory data and appropriate predictive algorithms, it is possible to determine the condition of equipment and predict the optimal time for maintenance intervention some time in advance, thus avoiding unnecessary costs and downtime due to lack of maintenance.

1.1.2.1 Model Prediction

Traditional forecasting algorithms have relied more on time series models, such as exponential smoothing [13] and seasonal autoregressive integrated moving average (SARIMA) [14, 15, 16].

Recently, however, artificial intelligence methods have become more popular. Artificial Intelligence is impacting society, politics, business, and industry [17].

Deep architectures are needed to learn the types of difficult functions that high-level abstractions can represent. Layer types, sizes, transfer functions, and other hyperparameters need to be thoroughly studied [18].

Modern time series and other data analysis techniques have been successfully applied to various tasks, such as highway traffic analysis [19] and additive manufacturing [20]. Various approaches have also been proposed in the field of predictive maintenance [21, 2].

1.2 Problem Description

The paper production market has gone through great social pressures in terms of sustainability, and its production requires the consumption of abundant water, but, on the other hand, the paper has the advantage of being biodegradable, which makes it a great partner in combating other

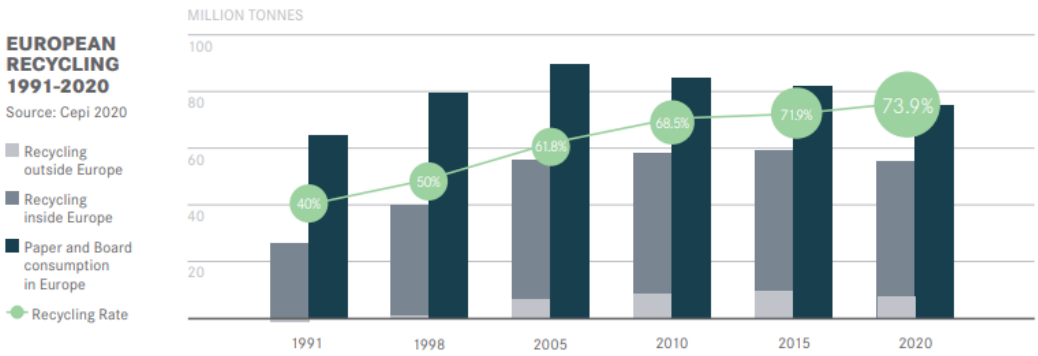


Figure 1.1: European paper recycling [23].

non-biodegradable materials [22]. This reuse presents a significant growth, as shown in Figure 1.1. With the reduction in the use and production of some materials that have a considerable environmental impact, such as plastic, many of these materials have been replaced by paper, which directly affects its production.

The strategies that support the production system of manufacturing industries are push and pull. Push's premise is to strictly monitor the growth of orders for the manufactured product in order to reduce the stocks of finished products, while pull's premise is to produce and then sell what, in fact, is widely used in industries until today. This rigorous monitoring of product requirement trends by the market is a great challenge.

With the advent of information technology, many industries have given more importance to information management in production chains, as this information has had great relevance in the answers to problems that arise during the production period. With historical data, it is possible to draw some conclusions from issues that would be impossible without it. This conclusion can be from the simplest to the most complex.

The information allows maintenance employees to have the most optimized intervention management possible, which makes it possible to strictly comply with the Total Productive Maintenance (TPM) concept of zero failure. This maintenance management can also be adjusted according to the industry's production flow while anticipating this flow, which leads us to implement predictive maintenance, as it allows us to carry out this premise.

There are many mathematical tools that allow us to perform forecasts with a very relevant precision, but for that the data must be processed, in order to eliminate possible mistakes in the forecasts and, consequently, to have a model adjusted to the forecast in a useful time. This processing starts right from the understanding of the variables and the relationship among them. From these studies, relevant conclusions can be drawn in order to optimize the time and computational capacity required to grant the model that best fits the respective predictions. Production lines often have to deal with a diverse set of issues. Different production lines consist of different problems that require numerous data and approaches [24].

Although machine learning has proven to be an effective tool for analyzing complex relationships and problems, it is still not clear which problems on production lines can be effectively solved with the help of machine learning approaches. Furthermore, the production line is a very broad concept and different industries have various production line configurations to deal with different issues during production. For example, some production lines can generate a large amount of data, in which case machine learning techniques can provide remarkable solutions; however, a similar solution may not be effectively applied to other production lines due to limited data [24].

1.3 Research Motivation

Industrial management aims to adopt measures that can bring a great balance between the process and decision-making. Based on this, many studies have been focused on finding optimization tools that can support short and long term decision making.

A long time ago, the information system did not present a robustness that nowadays is possible to achieve in several industries. This evolution was only possible thanks to the work that has been developed with the main focus on solving problems that cause losses in various industrial sectors. Having organized and accessible information is sharing the problem, which makes possible several solutions, which are only added benefits.

One of the biggest challenges for managing equipment maintenance is deciding which factors should be prioritized in upcoming interventions. Based on this, availability is a term that has great relevance in evaluating the effectiveness of any industrial plant, most of which are repairable systems [25].

With the help of preventive maintenance, it is necessary to predict the Remaining Useful Lifetime, which is difficult to measure in practice. However, preventive service can proactively reduce the number of breakdowns in production lines. Machine learning performance in the area of preventive service can be improved with data growth and new algorithms.

With the advent of computer technology, it was become possible to digitise data that explain the behaviour of real-world problems. The transformation into predictive mathematical algorithms was possible.

By combining the skills of different scientific fields, namely computer science, electrical engineering, mechanics and industrial management, it was possible to optimise problem solving.

1.4 Outline and Scope of Research

This PhD research used ten datasets, the first and second dataset is from the sensors installed on the pulp paper press machine. The first dataset has 1445760 sample size, with a one-per-minute sample rate. The second dataset has a 399745 sample size, with a five-per-minute sample rate.

The third, forth, and fifth of this dataset was obtained in internet site, the third is Steel Production in the world, the fourth is the Producer Price Index by Commodity, and the fifth is the Wold population.

The others datasets were obtained on the internet, these datasets were used to support the prediction of the first dataset, through variables correlation existence.

The data have constant sampling rates over a period of time, so they were treated as time series. These series were studied to get an idea of their seasonal behaviour and to obtain more information about their patterns; correlation and autocorrelation studies were also carried out.

This is reinforced with the idea that with databases and data processing tools it is possible to improve the asset monitoring system using the portfolios provided by the asset management system. This perspective, it is possible to automate the maintenance system by applying the concept that emerged in 1999, whose main objective is to make objects autonomous and intelligent enough that they do not need human intervention, this concept is supported by the IoT [26].

Industry 4.0 and IoT are gaining popularity [27, 28] as it makes industrial production a more flexible system, more adaptable to personalization and more traceable [29].

Industry 5.0 will complement Industry 4.0, which is based on three pillars aiming to humanize the use of AI [30]. Industry 5.0 is a value-based production paradigm and a revolution that emphasizes the importance of research and innovation to support industry placement. Worker well-being is at the heart of production processes [31].

Industry 5.0 research is still in its infancy, but is being officially promoted by the European Commission through an official document published in 2021 [32].

The circular economy and process engineering play an increasingly important role in recovering valuable components from highly fragmented material flows that leave users' warehouses after highly fluctuating periods of use [33].

These statements illustrate the need for a robust production process that is able to adapt to the market.

1.5 Research Questions

In this work, the following research questions (RQS) were answered:

- **Research question 1:** How Production can be optimized based on physical asset maximization availability?
- **Research question 2:** Which methods should be used in time series data analysis for predictive maintenance purposes?
- **Research question 3:** Which is the best algorithm, LSTM vs GRU to Predict the Condition of a Pulp Paper Press?
- **Research question 4:** What are the Pre-Processing Methods that improve GRU model in Prediction of Paper Pulp Press condition?
- **Research question 5:** How to do the prediction indexed to stock market?

1.6 Contributions of the work to the state of the art

The results obtained in the present work demonstrate the applicability of recurrent neural networks (i.e., GRUs) in predicting future behavior in the paper press industry. The same GRU architecture showed good results learning data from two different industrial pulp presses.

Data pre-processing can play a very important role in improving the predictions. In the present work, filtering out discrepant data and smoothing using a LOWESS filter reduced the MAPE errors for all variables.

The results show that it is possible to forecast future behavior of industrial paper pulp presses up to 30 days in advance with good degree of certainty. That can be a good opportunity for optimizing maintenance decisions, downtime and costs.

The case study it was possible to validate the importance of data treatment and the great relevance that data can have in a forecast of a single variable since these models allow us to extract information from the other variables in order to have a forecast with high accuracy.

The proposed data processing methods have so far been little explored for these types of applications. They have shown to be of great importance in making the training and testing process of the models more optimized and not only.

Training an encoder-decoder architecture model, consisting of a recurrent GRU neural network with the futures exchange data, becomes a contribution since our study shows this feasibility of the same.

1.7 Thesis Structure

The thesis continues as follows. Chapter 2 reviews, evaluates the relevant literature and summarizes and methods used in other studies applicable to our production management study. Chapter 3, discusses the methodology of data analysis and processing, the assumptions and the conceptual framework of our forecasting model. Chapter 4, presents different tests for tuning the hyperameters of each proposed forecasting model, focusing on predictive maintenance for the pulp press. In Chapter 5, they are the tests whose aim is to use the models and architecture created in the previous line and test them with production data and include external variables of the company that have an impact on the fitting of the models.

In Chapter 6 several discussions were presents on the results that each model presents in relation to the prediction data - these discussions are organised chronologically. Chapter 7 finally provides general observations and implications of our study, key findings, and suggestions for potential future research.

Chapter 2

State of the Art

Optimization is important because it allows us to find solutions to problems with few resources. Once the computational resources have been used to solve the prediction problem, optimization of the algorithm is required so that this task can be solved with computational resources. In this chapter, we show a review comparing the production optimization and plant availability approaches to understand how these two approaches work towards the same goal of maximizing the production capacity.

2.1 Production

Operations management involves the systematic management and control of the process that transforms resources (input) into finished goods or services for customers (output), as is shown in Figure 2.1 [34]. This basic model applies equally to manufacturing and service organizations, as well as to the private and nonprofit sectors.

Operations management is an emerging field, in production processes, i.e., undesirable interruptions that can affect the flow of production in real time [35].

A production cycle begins with a new system which is assumed to be in a state under control, producing items of excellent quality [36].

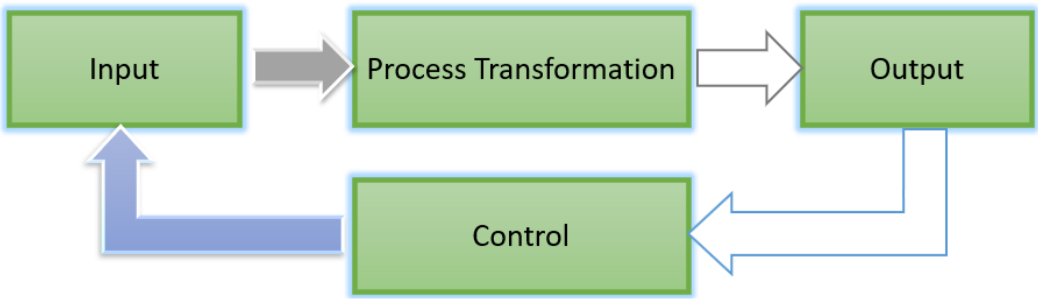


Figure 2.1: Operation management circuit.

This combination of human-machine knowledge allows the crossing of operators' knowledge with problem-solving algorithms, namely that of forecasting, making it possible to predict future failures [37].

Understanding a company's customer base for a particular set of products is critical to a company's decision on the required production process. Many authors suggest strategies for product positioning [38, 39, 27].

The closed-loop process planning and scheduling approach in Figure 2.2 uses dynamic feedback from the production schedule and information about the current availability of resources to generate process plans. In short, the planning phase communicates to the scheduling process the current availability of machines on the shop floor [40].

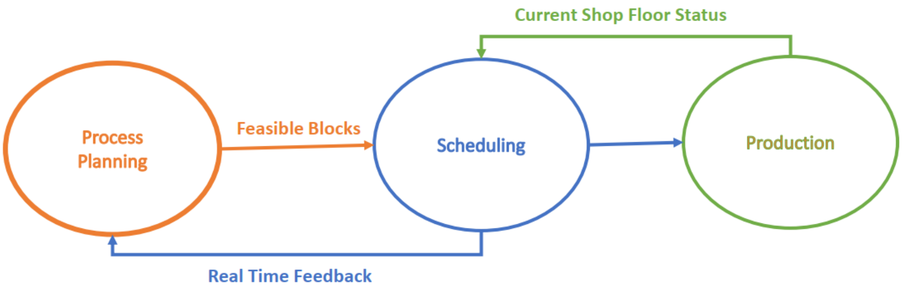


Figure 2.2: Closed-Loop Process Planning and Scheduling approach [37].

According to Tsang [41], a way to visualize the maintenance system based on equipment condition, operational load, maintenance actions (strategies), and business goals is affected by both operational load and maintenance actions.

The operational load depends on the production plans and decisions, which, in turn affect the commercial needs and market consideration. Therefore, maintenance planning must support production planning, such as maintenance decisions, legacy equipment reliability, and market and commercial requirements. It is shown in Figure 2.3.

For years, the relationship between production and maintenance was considered an antagonism in management decision-making. This situation has not changed because the scaling requirements of each role are not aligned [42]. Conflicts can lead to dissatisfaction in demand for production due to interruptions caused by preventive maintenance (PM) [9].
Availability is the most important term used to evaluate the effectiveness of an industrial facility. The steady-state availability of a plant for a given period of time is defined as the percentage of time that the plant is performing its properly designed production [25].

The degree of adaptation of a competitive organization implies priorities in its primary decisions regarding structural and infrastructural investments, which are key to promoting the full potential of its operations as a competitive weapon. Figure 2.4 illustrates a graphical model of the commonly

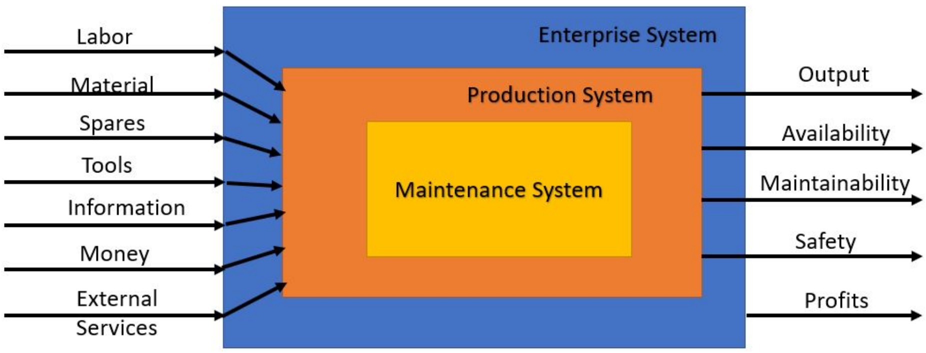


Figure 2.3: Production structure, and its surroundings Tsang [41].

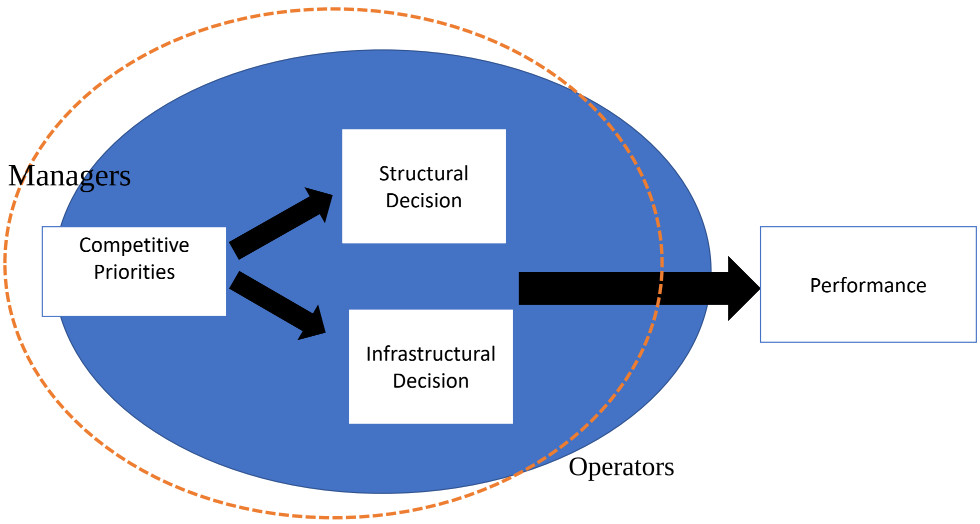


Figure 2.4: Production structure [43].

used operational strategy [43].

Figure 2.5 illustrates the internal and external factors that make up the production planning environment. In general, the external environment is outside the direct control of the production planner but, in some companies the demand for the product can be monitored. As for the internal factor, there are failures that can occur during the production period of a good. If there is a good maintenance strategy, it can facilitate resolution because there is a possibility that operators will do a lot of the maintenance since they are more familiar with their machines [44].

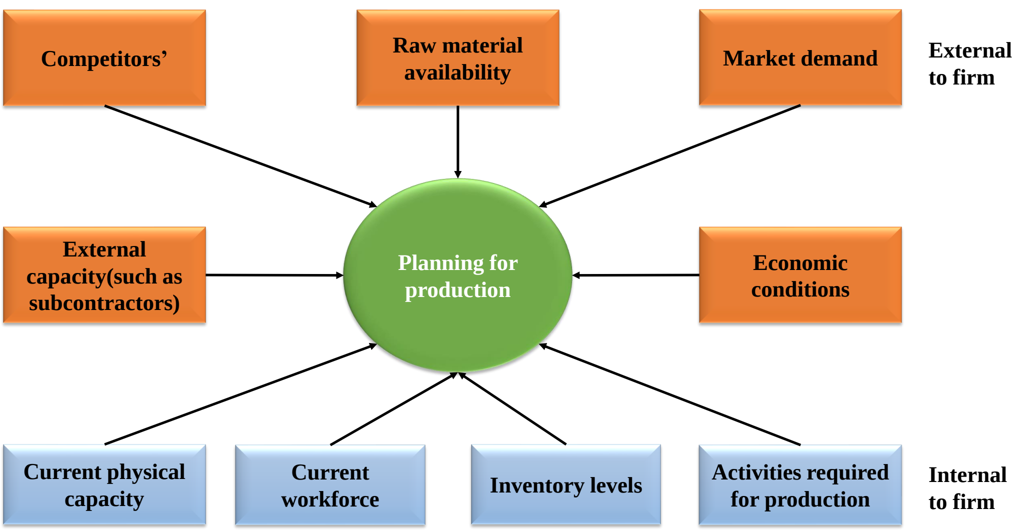


Figure 2.5: Factors required in the production planning system [44]

As machines become more complex, operators less skilled who maintain them may not recognize potential problems until the machine output is detrimentally affected. This will result in less operator participation in preventive maintenance and a reduction in the early detection of problems.
Bojanowski [45] presents the following rough distribution of factors causing machine failures:

- Failures caused by lack of awareness (caused by lack of inspection) of impending machine failure: 50%;
- Failures caused by lack of machine servicing at proper intervals: 20%;
- Failures caused by bad fit, use of substandard materials or accident: 15%;
- Failures due to normal wear: 15%.

2.1.1 Push and Pull System

Push and Pull are business terms that were originated in logistics and supply chain management, [46, 47], but are also used in marketing [48, 49] and the hotel distribution business.
Venkatesh et al. [50] described push and pull as operational paradigms. In a push system, the preceding machine produces parts without waiting for a request from a later machine. On the other hand, in a pull system, the preceding machine only receives products after the request of each machine.
According to De Toni et al. [51], to apply push-pull classification to manufacturing systems, they must be considered three subsystems: the manufacturing priority, material selection, and movement subsystems; and the production planning subsystem.
Demand forecasts form the basis of all managerial decisions in logistics and supply chain management. Regardless of a push or pull type of a supply chain system, demand forecasting is the starting point for all planning activities and execution processes.
Pull production eliminates the waste created by more traditional push production systems, where material is moved from upstream to the next downstream operation as it becomes available. This is a product-out philosophy of production and results in over-production and/or delivery delays [52].
Storage and logistics systems are classified as push or pull systems (Figure 2.6).
Consider the push process executed to forecast customer demand, purchasing, production, transport, activities and operational actions. All of them require demand forecasting as data input; the same applies to the pull process - required activities and inventory are planned, customer demand data should be a starting point [53].
Pull production brings all these methods together, and they entirely revolutionize production management.
Team [52] made a study on production management, namely the pull system and the optimal sequence to establish pull systems in a factory. In a nutshell, Figure 2.7 shows the relationship between the whole lean production structure and the various methods.
The fact that machine bottlenecks are unknown and machine loads are rarely updated makes production planning very unreliable. Nothing is more important than having prior knowledge of

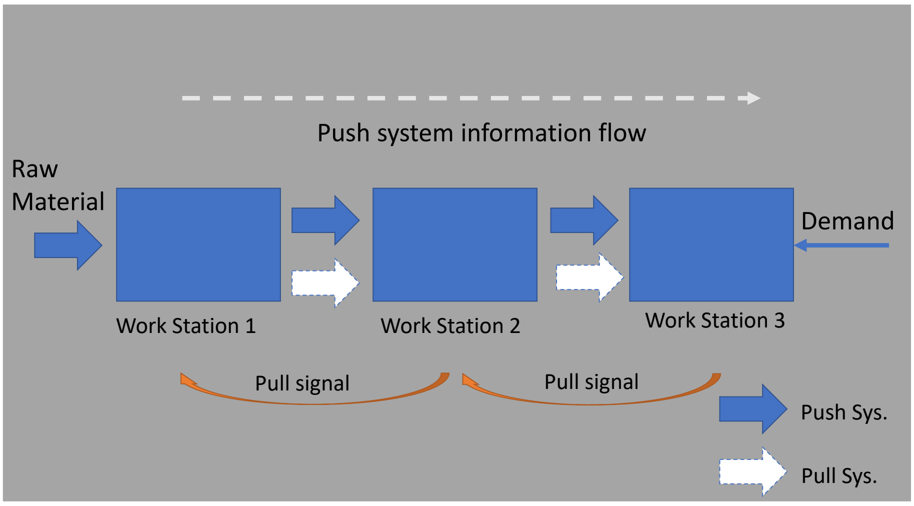


Figure 2.6: Pure push and pull systems [44].

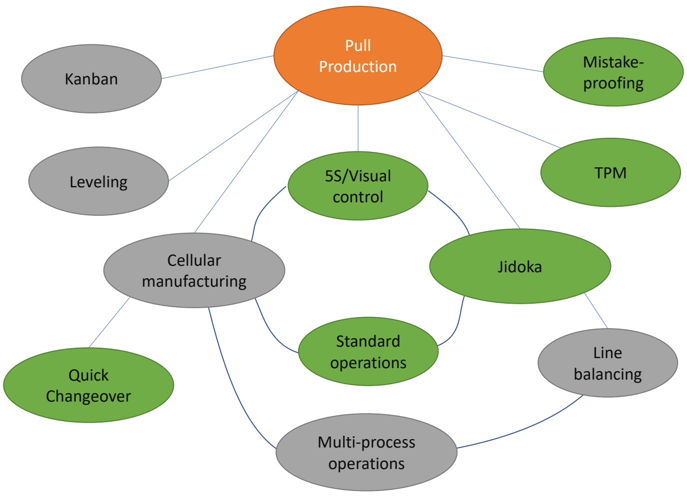


Figure 2.7: Connectivity in a Lean Manufacturing System [54].

what is happening with the machine. The performance of push-pull information flow systems is considered to be related to high quality levels, small setups and small batches, i.e., the conditions usually associated with a Just In Time (JIT), continuous improvement program [55].

Bonney et al. [56] demonstrated in their study the different results of the push-pull system. Assuming that the batch size is the same and orders are issued on a per order basis, the push-pull information system will perform better, which can be measured by rejected demand when no backlog occurs or average waiting time when demand is backlogged. A related finding is that push systems sometimes require lower inventory levels to achieve equivalent performance.

2.1.2 Production Just In Time

The growing interest in JIT management systems has led many companies to keep a close eye on their Work In Progress (WIP) inventory levels. Effort to reduce WIP inventory is becoming commonplace. The reduction in WIP inventory leads to a reduction in buffer inventory between workstations and makes equipment failures more severe than ever [57].
According to Rosenblatt and Lee [58], when the production process is subject to a random process deterioration it shifts the system from an in-control state to an out-of-control state.
Production management often view maintenance in the context of hours or days out of service and fails to realize the strategic importance of incorporating maintenance planning in the implementation of JIT manufacturing. Management for the maintenance function, on the other hand, attempts to impose constraints on production that it deems necessary to achieve complete equipment reliability. Thus, an issue that should be decided by the organization's strategic management often is settled at the operational level as a test of political clout [42].
Random failures of production cells are one of the main disturbances that production systems often suffer from. This disruption reduces the plant's effective capacity and results in higher operating costs, especially with JIT production [59].
In order to get the best out of expensive manufacturing processes, to meet the quality challenge, and to ensure the success of new management system strategies, such as the JIT approach, equipment must be maintained in good operating condition. This environment shifts the focus to maintenance and the need for effective maintenance policies [60].
According to Martin [61], costs associated with maintenance generally consist of the cost associated with maintenance labor, the cost of required material and spare parts, and the cost of production downtime when breakdowns occur.
Xie et al. [62] verified increased inventory and production costs related to frequent plan adjustments.

During the last decades, several production control and maintenance policies have been proposed in order to improve manufacturing system performance. Kimemia and Gershwin [63], and Akella et al. [64] have considered the production control problem for systems prone to failures.

In order to meet future demands with high level of uncertainty and to ensure the economic success of production systems, cost-effective and reliable production and maintenance planning is necessary [65].

Extensive existing literature shows that implementing preventive maintenance strategies in unreliable production facilities can effectively extend machine life and reduce operating costs [66]. According to Jonsson [67] the need to shift the focus of maintenance policy from traditionally focusing on short-term issues (resource consumption, costs, etc.) to long-term goals (competitiveness, sustainability and strategy). Figure 2.8 shows the goals and plans that emphasize on quality improvement, prevention and manufacturing capabilities.

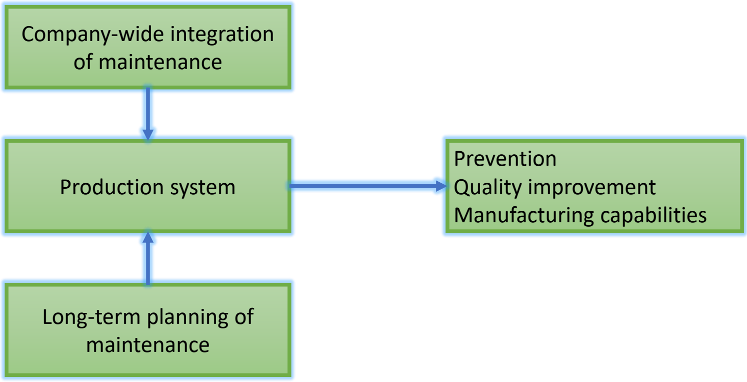


Figure 2.8: The links between integration and planning of maintenance and prevention, quality improvement and manufacturing capabilities [67].

Demand forecasting is especially important in businesses that involve mass production. As this requires a long lead time, a lot of forward planning has to be done. In addition, potential future demand must be estimated to avoid overproduction and underproduction.

2.1.3 Forecasting Production

Matsumoto and Komatsu [68] proposed forecasting methods for production planning at a food company. The method used demonstrates its simplicity and accessibility due to its low cost and ease of use. Due to these characteristics, this method can be used by small and medium-sized companies, where it is not possible to make large investments in planning their operations.
The energy production, distribution and consumption had been an important role for researchers during decades, as can be seen in valuable works in the field of forecasting energy consumption [69, 70].
Coarse modelling is used to develop a three stage electric energy load forecasting model to predict the yearly, weekly, hourly electric energy demand [71].
For example, for tourism,[72] applied Autoregressive Moving Average (ARMA)-based models to forecast the number of visitors to countries showed that the forecast numbers had an average error of less than 10 % (in terms of mean absolute percentage error (MAPE)).
The difficulty of forecasting demand has prompted companies to focus on improving supply chains [73, 74]. This is one of the success factors in the shortest time-to-market for brands like H&M and Zara.

According to Mir et al. [75] these forecasting methods on different time horizons show that time series modeling methods have been widely used for medium and long-term forecasting. AI-based techniques are still widely used in the literature for this purpose.

2.2 Optimization

Research in the field of optimization of maintenance is a priority [76], and there has also been a great trend in the field of optimization based on the simulation of maintenance [77, 78, 79]. This proves that research in maintenance optimization started decades ago.

An optimization process is required to determine optimal capabilities and operational strategy [2]. Traditional optimization for production processes usually involves simultaneous flowchart selection, as well as the corresponding operating conditions [80]. Preventive maintenance (PM) planning and production scheduling are among the most important problems in the manufacturing industries. The researchers began to investigate the problem of integrated optimization of PM and production programming with a single goal [81].

Cua et al. [82] say that optimization techniques seek the best solution for each problem (maximum or minimum measurable quantities in their definition domains); it is necessary tools in many areas of engineering, such as:

- Operational research - optimisation of technical and economic systems, stock control, production planning, etc.;
- Process control - system identification, optimal control, adaptive control, predictive control, state estimates, etc.

In the production process, optimization is not just about the first interaction. As an example study, Monostori et al. [83] describe each looping to optimize its production system to achievement goals for a predefined period, as is shown in Figure 2.9. The same study also states that the more simulations performed with improved parameters and conflicting with the actual situation on the shop floor, the better the model fits.

With strong competition resulting from global pressure, the dynamics experienced by the industry increases the critical need to economically optimise production. Achieving the latter is tantamount to shooting at a moving target.

In a global way if we reduce the problems of maintenance management we can reduce cost in production, and if in a way we can optimize production indexed to the market to predict possible variations in the market in pulled production it is possible to reach a precise adjustment in the realization of maintenance in the production lines. Unfortunately, only a few contributions address the problems under a view in the Optimization of production indexed to the market to date, there are three associated studies found in Scopus [85, 86, 87].

Figure 2.10 shows the results of papers in the individual areas, these are results from the ”Web of Science” website.

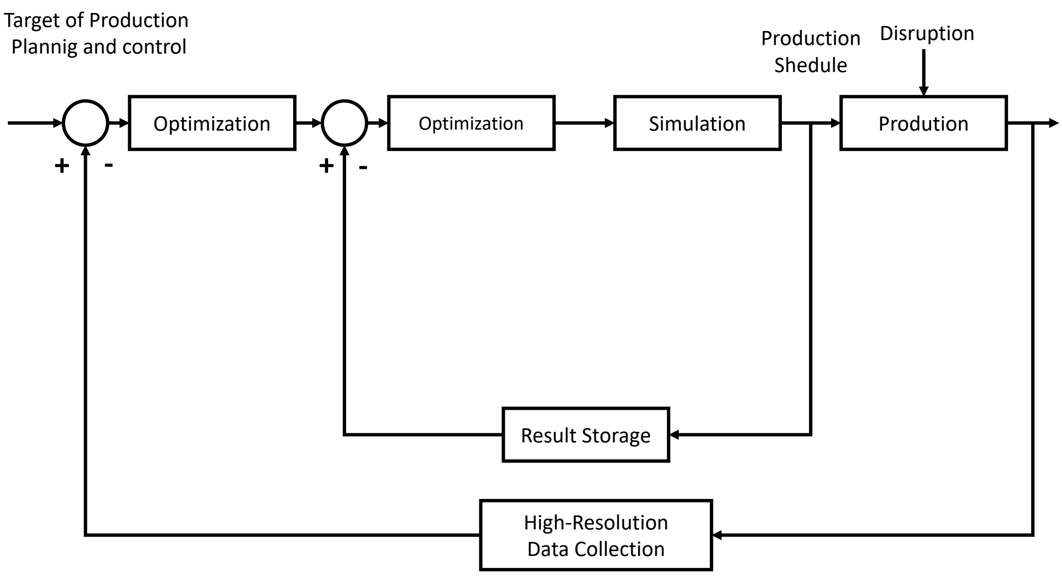


Figure 2.9: Architecture of a cyber-physical production control [84].

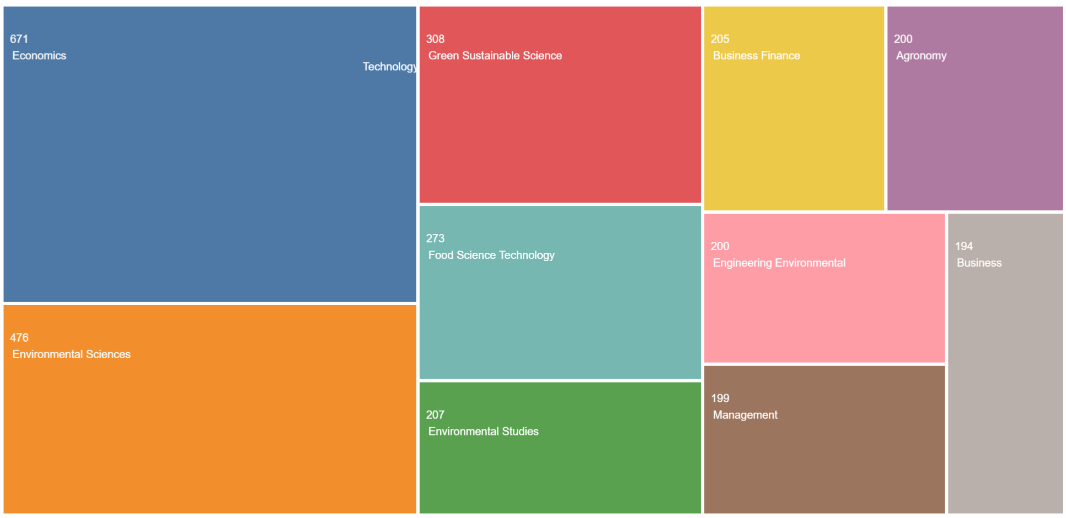


Figure 2.10: Documents, by research area, in the Web of Science database.

Wang et al. [88] assume that an optimization process is required to determine optimal capabilities and operational strategies. Great contribution to dynamic optimization by combining process simulators with meta-heuristic techniques to simultaneously optimize process flow diagrams with corresponding operating conditions [89].

Operations management is used for strategic and tactical applications, i.e., in the following areas: master planning; forecasting; positioning; scheduling; capacity planning; layout; process and product design; quality control; mission design; inventory control; maintenance and reliability [90].

In the industrial area, it is also possible to carry out an adaptation of the mathematic optimization model to solve complex problems.

Dorigo et al. [91] present an overview of recent work on ant algorithms, i.e., algorithms for discrete optimization that were inspired by the observation of ant colony forage behaviour and presenting the ant colony optimization metaheuristic.

Nocedal and Wright [92] noted that understanding the capabilities and limitations of optimization algorithms can lead to a better understanding of what they mean in different applications, and point the way for future research on algorithms and software to improve and extend them. In order to include the optimization of the digital industry, certain requirements must be met, such as: robustness of the database and reliability of the data/schema. According to Dekker [76], the maintenance optimization model has several applications, generally covering four aspects:

1. A description of a technical system (its function and importance);
2. A model for the deterioration of the system over time and possible consequences for the system;
3. A description of the information available on the system and the actions available to management;
4. An objective function and an optimization technique that helps to find the best balance.

These maintenance optimization models produce different results. First, policies can be evaluated and compared in terms of the characteristics of cost-effectiveness and reliability.

Wang [93] presents an extensive review of maintenance optimization policies. Maintenance optimization studies prior to 2002 mainly considered time-based maintenance configurations.

Syan and Ramsoobag [94] state that modern maintenance optimization decisions are complex problems that need to satisfy multiple and conflicting criteria. With the increase in applications and recent advances in Multi-Criteria Optimization (MCO) approaches, a review is needed to group and categorize these advances in the field of maintenance.

Jonge and Scarf [95] say that optimization applied to maintenance comprises the development and analysis of mathematical models that aim to improve or optimize maintenance policies. A study on the substantial developments in the field of maintenance optimization is fully demonstrated in [93].

In order to validate the effectiveness of decision models, Bousdekis et al. [96] prove that an event-driven proactive decision model is possible for joint predictive maintenance and optimization of

the spare parts inventory, which addresses the ”Detect” ”prevent-decide-act” model phase that can be incorporated into an Event Oriented Architecture (EOA) for processing time within the framework of the concept of electronic maintenance.
The techniques presented in Mobley [97], there are five non-invasive techniques used for the management of predictive maintenance, such as monitoring vibrations, monitoring process parameters, thermography, tribology and visual inspection. Predictive techniques can vary, as mentioned in [98]: lubricant analysis; vibration analysis; thermography; penetrating liquids; radiography; ultrasound; corrosion control; etc.
Zhou et al. [99] refer some drawbacks saying that the ideal maintenance policy is, in fact, a monotony, in which the limits decrease monotonously with the age of the system; but, other studies expose some solutions, like Zhao et al. [100], that propose a predictive maintenance policy based on process data, demonstrating that, when compared to traditional preventive maintenance strategies, their strategy has adaptability and effectiveness to the deterioration of the system.

2.3 Maintenance Management

Maintenance consists of a series of actions that help to ensure that a tangible asset functions properly, such as maintaining a paper drying press. There are four types of maintenance policy, which are shown in Figure 2.11.

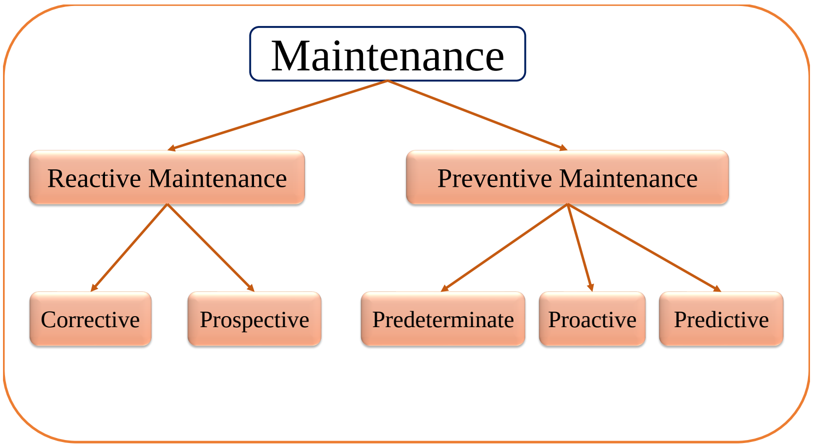


Figure 2.11: Maintenance types [43].

Tsang [41] and Farinha [101] it emphasize that maintenance literally means ”the work of keeping something in good shape” and states that maintenance should be done to prevent equipment or components from failing or to repair normal degradation of equipment when it comes to keeping it in good working order.
Due to the lack of reliability of time-based maintenance methods, industrial processes should use ”online” monitoring not only when the equipment is obsolete, but also throughout the life of the equipment to identify the onset of equipment degradation and failure [102].

Jacobs and Chase [44] refer that the costs shown in Figure 2.12 must be considered. Costs can be divided into three broad categories: acquisition costs, ownership costs, and post-ownership costs.

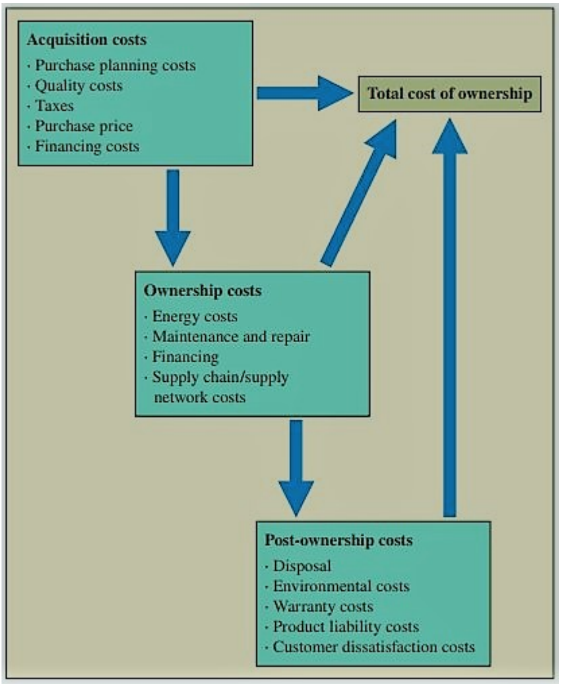


Figure 2.12: Total Cost Ownership [44].

Our study starts from the goal of reducing the costs, namely ownership costs, with the main benefit being the maintenance policy.

Tsang [41] emphasizes the maintenance as a necessary expense that fits within the operating budget and a common item on the list of industry cost reduction programs. Reactive Maintenance - is a maintenance strategy applied when an asset (i.e., a device or component of a system) fails. There are failures that are not very serious, have less impact on the system, and do not cause much damage, so they can be fixed after they occur [103].

Preventive maintenance is a maintenance strategy that is regularly applied to assets to minimize or reduce the likelihood of failure. In this strategy, equipment is inspected regularly and replaced as needed. This applies when the equipment is still in service to reduce the likelihood of failure [104].

Predictive Maintenance - as the name implies, is a maintenance strategy that predicts the likelihood of system or equipment failure. After a failure prediction, equipment can be replaced or repaired, and proper planning can be done before the equipment fails [105].

Proactive Maintenance - is a maintenance strategy that usually targets the root cause of a failure and simply avoids the possibility of failure. It is the exact opposite of reactive maintenance. It is performed before a fault occurs [106].

Maintenance plays a key role in reliability, availability, product quality, risk reduction, greater equipment efficiency and safety [107]. Machines' failure should be predicted with accuracy. Predictive maintenance allows the maintenance recurrence to be as low as possible to prevent unplanned reactive maintenance [108].

The main objective of maintenance planning is no different from the objective of any planning activity, i.e., planning allows decisions to be made early so that decision makers can consider several possible options and make the best decision. A maintenance management plan allows for alignment of Corrective Maintenance (CM) and PM capacity requirements. While scheduling corrective maintenance is difficult at best, increasing PM can reduce the need for CM and thus increase the amount of maintenance work that can be scheduled. CM increases the number of decisions made early enough to consider options, resulting in more economical decisions and lower costs [109]. Kershaw and Robertson [110] refer that predictive maintenance works by regularly monitoring the condition of components rather than replacing them, which means better data, increases plant productivity, and prevents catastrophic failures. Chang et al. [111] for example, explore maintenance optimization and introduces a method that incorporates real-time information about production conditions and machine failures. According to Carnero [112] predictive maintenance can increase safety, quality, and availability in industrial plants.

The graph shown in Figure 2.13 illustrates that continuous investments in preventive maintenance reduce failure costs and, as a consequence, a decrease in the total maintenance cost, in which preventive maintenance costs are added to the failure costs. However, the graph also shows that, from the ideal point of investment in preventive maintenance, more investments bring few benefits to reduce the cost of failures increasing the total cost, which is what the maintenance policy takes into account.

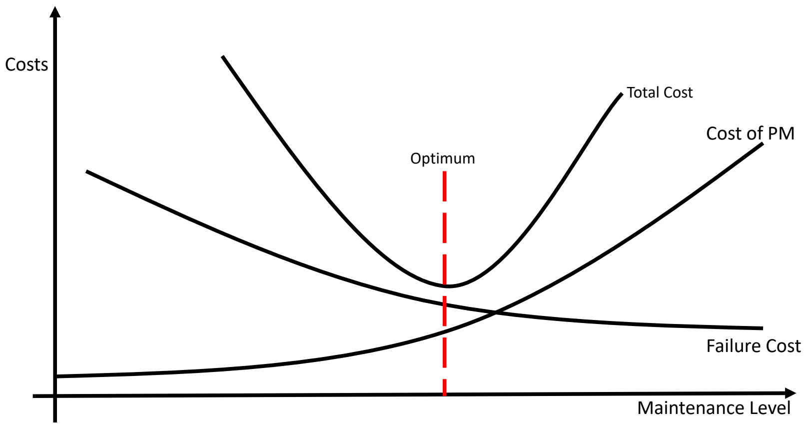


Figure 2.13: Operations strategy model.

Currently, the companies constantly seek maximum production. To this end, actions are taken to reduce the overall costs of the company. Therefore, measures are taken to ensure the availability of production equipment and the good quality of the product [113].
Timely maintenance is a more aggressive strategy that keeps the system in good shape and cost-effective. Some valuable related papers are, for example, [114, 115]. According to Zhou et al. [116] maintenance of options for multicomponent systems considers random failures and disassembly sequences. Bedford et al. [115] specifically, discuss a maintenance model in relation to competing risks.

2.3.1 Predictive Maintenance

To have a prediction with good model performance, is important that the sensor data collected is of good quality. Deep predict models have been successfully used to improve condition monitoring forecasts for industrial equipment. The old-fashioned approach to predictive field maintenance was to manually create a predictive schema for a specific component based on a Boolean combination of some related event codes.
According to Kershaw and Robertson [110], predictive maintenance works by regularly monitoring the condition of components rather than replacing them, which means better data, increases plant productivity, and prevents catastrophic failures. Predictive maintenance can increase safety, quality, and availability in industrial plants [112].
This approach is very empirical, but illustrates an important concept - component failure can be predicted by examining the pattern of data. Predictive maintenance requires a deep understanding of the health and condition of the equipment. This can be done by adding sensors to the device for recording and monitoring signs such as temperature and voltage [117].
In the development of a dedicated computer program, it is possible to highlight possible resolutions for decision-making problems in relation to: manner, scope and schedules of replacements, repairs and regular maintenance of elements of technical objects, mode and schedules of diagnosis and preventive replacement of elements and problems of supplying spare parts to the maintenance system [118].

According to Tsang [119], there are three types of decisions which need to be made in condition-based maintenance:

1. selecting the parameters to be monitored;
2. determining the inspection frequency;
3. establishing the warning limit (the trigger).

In smart industries, predictive maintenance is one of the most used techniques to improve condition monitoring, as it allows one to evaluate the conditions of specific equipment in order to predict problems before failure [120]. For good performance of predictive models, it is important that the sensor data collected are of good quality.

In order to assess the condition of a system, the predictive maintenance approach employs sensors of different kinds. Some examples are temperature, vibration, velocity or noise sensors, which are attached to the main components whose failure would compromise the entire operation of the system. In this sense, predictive maintenance analyzes the history of a system in terms of the measurements collected by the sensors that are distributed among the components, with the objective of extracting a "failure pattern" that can be exploited to plan an optimal maintenance strategy and thus reducing offline periods [121].

When systems start to be very complex or the number of sensor measurements to manage is very large, it can be difficult to estimate a failure. For this reason, in recent years, machine learning techniques are used more and more to predict working conditions of a component. Mathew et al.

[122] propose several approaches to machine learning such as support vector machines (SVMs), decision trees (DTs), Random Forests (RFs), and others that show which technique has the best performance in RUL forecast for turbofan engines.

A major challenge in operations management is related to predicting machine speed, which can be used to dynamically adjust production processes based on different system conditions, optimize production performance and minimize energy consumption [123]. Preventive Maintenance (PM) may have the ability to maintain machines with a high level of reliability [124]. But, the implementation of scheduled maintenance activities can also lead to machine unavailability while PM is being performed [125, 126].

However, the creation of a Predictive Maintenance Program is a strategic decision that, until now, lacks an analysis of the problems related to installation, management and control. According to Shin and Jun [127], when it is a high-value asset, the Operation and Maintenance (O&M) phase requires heavy charges and more effort than in the installation (construction) phase, as these assets have a useful life that any unexpected event of the asset during that period causes catastrophic damage to the industry.

The proposed option-based PM policy can provide flexibility to adjust production output to satisfy the demand requirement [65].

2.3.2 Opportunistic Maintenance

Opportunistic Maintenance (OM) is used when a component and equipment are assumed to fail stochastically, and the failures are independent according to known a probability distribution; in this case, a combination of corrective and preventive maintenance (PM) applies as a failure occurs [128].

OM policy is developed based on combination of age replacement policy and block replacement policy and in practical; OM is applied as the combination of corrective maintenance which is applied when any failure occurred, with preventive maintenance (PM) – a planned and scheduled maintenance approach to prevent failure to happen.

In addition to the classic system-level maintenance strategy, a new maintenance strategy is required to achieve a fast and cost-effective response effectively [129].

Performing preventive maintenance, even when there is no opportunity, it can have a detrimental effect. It can be said that opportunistic maintenance policies are sufficient in the case of systems with a large number of components [130].

In case of CBM, the maintenance activities can begin at an arbitrary point within the opportunistic zone (which is equal to or less than the P-F interval) Figure 2.15. This opportunistic zone is the period during which the degradation started, without leading to a fatal shutdown of the component. Within the opportunistic zone, Planned Maintenance (PM) activities can be carried out against PM costs [131, 132].

[131] defined the opportunistic maintenance zone in age based replacement as a percentage of the PM interval T (Figure 2.14).

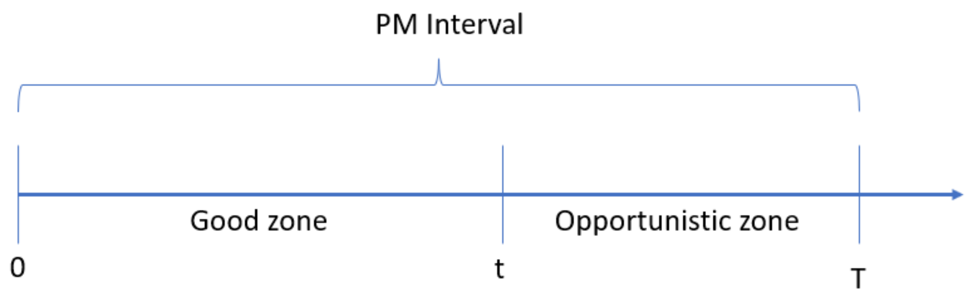


Figure 2.14: Opportunistic maintenance zone in ABR policy [131].

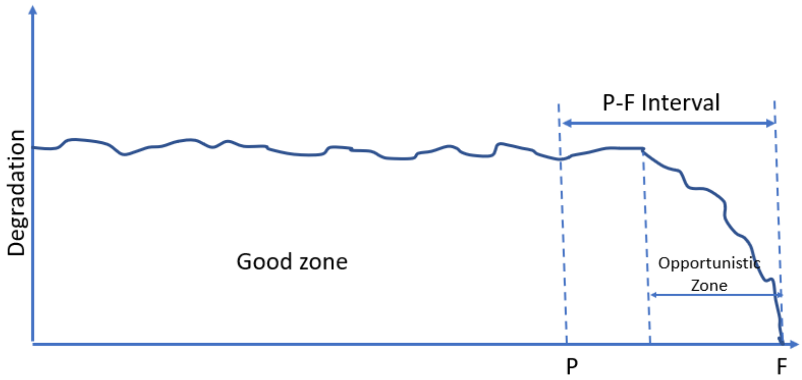


Figure 2.15: Opportunistic maintenance zone in CBM policy [131].

Rao and Bhadury [133] demonstrated that policies with various opportunistic maintenance ages for each increasing failure rate component are better than policies with a single opportunistic maintenance age for each component. In general, the relative advantage of opting for policies with various ages of opportunistic maintenance increases with increasing time/cost of preventive maintenance.

Zhou et al. [134] studied a dynamic programming method in which decisions are based on a combination of OM cost savings and penalty costs, and stated that an ideal maintenance practice is determined by maximizing the savings of cumulative short-term costs. Ding and Tian [135] proposed a method for making opportunistic maintenance decisions by comparing the age of a given component with a limit defined by some percentage of the Mean Time To Failure (MTTF).

Dekker [76] developed a model to determine the ideal age for opportunistic maintenance when the opportunity follows the Homogeneous Process of Poisson (HPP). Letot et al. [136] with the adaptive opportunistic maintenance model, based on the forecast of the conditions of a railroad, demonstrated that the adaptive opportunistic maintenance strategy has a lower cost per unit of time than systematic preventive maintenance. Truong Ba et al. [137], in their results indicate that significant savings can be achieved considering OM. In addition, it is shown that the new consideration of partial opportunities significantly increases the benefit of OM.

Among preventive maintenance control policies, opportunistic maintenance is an effective strategy for reducing the impact of maintenance operations on multistage manufacturing systems [138].

de Jonge and Scarf [139] stated that maintenance optimization consists on the development and analysis of mathematical models that aim to improve or optimize maintenance policies. Substantial developments in the field of maintenance optimization can be observed in [140] review. Rao [130] demonstrates that policies with various opportunistic maintenance years for each increasing failure rate component are better with respect to policies with a single opportunistic maintenance age for each component.

They have developed a systematic method about when to turn off equipment for maintenance is called a maintenance opportunity. The method incorporate real-time information about the production and failure conditions of the machine. The research indicates that the implementation of opportunistic maintenance allows preventive maintenance to be performed significantly more during the scheduled production shifts [114].

Besnard et al. [141] present an opportunistic maintenance optimization model for offshore wind energy systems. Chang et al. [111], for example, explore maintenance optimization and introduce a method that incorporates real-time information about production conditions and machine failures.

According to Hu et al. [142] the opportunistic Predictive Maintenance strategy for global optimization of predictive maintenance costs for the entire complex wind turbine system, considering failure probabilities, repair costs, downtime, and installation cost, it can make maintenance work more economical, with sufficient guarantee of system safety and reliability.

2.4 Artificial Intelligence

Daniyan et al. [143] propose the integration of Artificial Intelligence (AI) systems, since it brings many benefits in the diagnosis of industrial machinery condition problems. They highlight the feasibility of AI combining with a time series model, for fault diagnosis, to optimize the equipment intervention time.

Many artificial intelligence tools help in security in identification to financial fraud. Also, to finding relevant web pages in response to the research of the user, find the best driving route any destination, play chess, translating hundreds of languages and decision-making [144].

The quantity and quality of big data is subject to statistical replication in large-scale scientific experiments and is subject to many cognitive uncertainties. Big data can have the greatest societal impact when combined with artificial intelligence [144].

Artificial intelligence is advancing at breakneck speed, powered by ”deep learning” algorithms that use vast amounts of data to train complex programs called ”neural networks” [145].

According to Mitchell [145], there are some limitations in artificial intelligence systems the technicians dealing with this type of technology know that artificial intelligence presents limitations since it is not able to answer questions so far outside its training capacity. The lack of this understanding makes these programs vulnerable to unexpected bugs and undetectable attacks.

In practice, the learning process is implemented using mathematics, statistics, logic and computer programming. The learning process allows AI models to be trained on data in an iterative process, learning rules to adjust parameters through trial and error. Performance metrics minimize discrepancies between model predictions and experimental data [144].

As a powerful pattern recognition tool, AI has attracted great attention from many researchers and has shown promise in fault identification applications in rotating electrical machines [146].

The area of the artificial intelligence has with subsection Machine learning that can be defined as ”the field of study that enables computers to learn without being explicitly programmed”. It can be said that ”Machine learning algorithms use computational methods to learn information directly from data without using predefined equations as models” [147].

Modern machine learning methods provide excellent performance and are gaining popularity [148]. Can use high-dimensional and multivariate data [149]. One of the most popular tools is Artificial Neural Networks (ANNs), which have been proposed in many industrial applications, soft sensing [150] and predictive control [151]. As shown in this study [152], random forest models are also good predictors.

2.5 Deep Learning

Deep learning is a branch of machine learning based on artificial neural networks. It is essentially based on the essence, i.e. learning multilevel representations and data abstractions. Deep neural architectures have demonstrated excellent performance in both unsupervised and supervised learning-based tasks [153].

Various deep learning algorithms such as Deep Multilayer Perceptron (DMLP), Restricted Boltzmann Machine (RBM), Deep Belief Network (DBN) are employed in recommendation, especially in evaluating prediction Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), etc.

In a case related to the steel industry, Bampoula et al. [154] use neural networks to classify maintenance activities, so that interventions are planned based on the actual condition of the machine rather than ahead of time. Identifying states and the Remaining Useful Life (RUL) at higher resolutions using multiple neural networks can be very difficult because the system can predict fault classification and may fail to detect adjacent states.

According to Yasaka et al. [155] deep learning is used with CNN to achieve high performance in image recognition. The images themselves can be used in the learning process by this technique and no feature extraction is required before the learning process. Other computer vision research includes [156, 157].

Deep learning strategies have been used, with success, in a variety of areas [158]. According to [159] deep neural networks can outperform other methods in voice recognition tasks. A similar approach was used in audio processing [160].

An autoencoder (AE) is an acyclic feed-forward neural network consisting of an input layer, one or more hidden layers connecting the input and output layers, and an output layer [161]. The number of neurons in the input and output layers is equal to the input layer, which minimizes the difference between the input and output. Instead of predicting the output label, the input is recreated from the learned features.

Basically, the purpose of this network is to efficiently learn hidden features so that the output (similar to the input) can be reproduced after the features have been learned. It has two components, as shown in Figure 2.16: an encoder that encodes the input into a hidden/corrupted state, and an decoder that produces an output (similar to the input) after taking the encoded data.

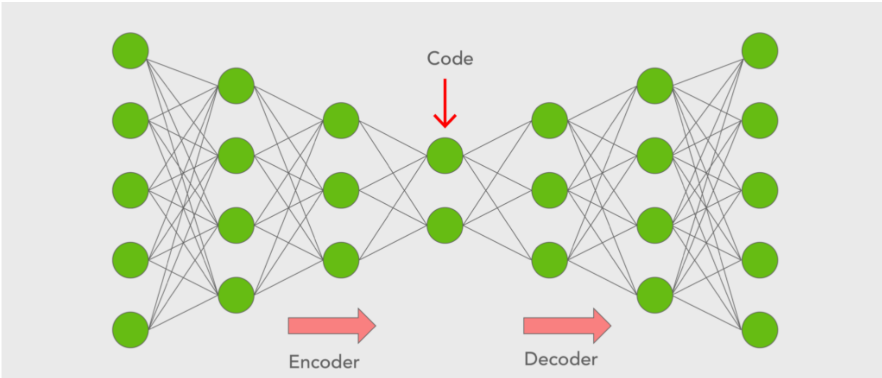


Figure 2.16: Architecture of autoencoders [131].

2.6 Recurrent Neural Network

Hsu et al. [162] demonstrated that neural networks can be a great technology in the support and decision making of large and small companies. There is a trend to use those tools in predictive maintenance systems with the aim of making the prediction systems more intelligent [163].

Experiments were performed with a predictive model based on the Long short-term memory (LSTM) with encoder and decoder architecture. The model consists of two LSTMs, in which the first LSTM has the function of processing an input sequence and generating an encoded state. The encoded state compresses the information in the input stream. The second LSTM, called a decoder, uses the encoded state to produce an output sequence. Those input and output sequences can be of different lengths.
LSTM is very good at predicting in a time series [164, 165]. It could extract patterns from sequential data and store these patterns in internal state variables. Each LSTM cell can retain important information for a longer period when it is used. This information property allows the LSTM to perform well in classifying, processing, or predicting complex dynamic sequences [166].
This technique has already been used to solve problems such as the prediction of vehicle trajectories based on deep learning [167]. Cho et al. [168] has shown great performance for tasks of translating from sequence to sequence. LSTM encoder–decoder models have also been proposed for learning tasks such as automatic translation [168, 169]. There is the application of this model to solve many practical problems, such as the study of the equipment condition, applications in language translations, among others [170, 171, 172].

Many networks showed instability when dealing with exploding or vanishing gradient problems during learning. Those problems happen when the gradient of the error is too large or too small. If it is too large, it overflows and the errors cannot propagate properly through different layers during learning. If it is too small, it vanishes and the network does not learn.

Traditional ANNs are simple and adequate for a wide range of problems. Bangalore et al. have

studied the performance of neural networks for early detection of faults in gearbox bearings, to optimize the maintenance of wind turbines [173].
In a case related to the steel industry, Bampoula et al. [154] used neural networks for classification of maintenance activities, so that interventions are planned according to the actual status of the machine and not in advance. Using multiple neural networks to identify status and RUL at a higher resolution can be very difficult, as the system can predict failure classifications and may not be able to recognize neighboring states. One limitation arises from the need for maintenance records to label datasets and the need for large amounts of data of adequate quality with maintenance events, such as component failures.
Beshr and Zarzoura [174] used neural network models to predict problems of suspended road bridge structures based on global navigation satellite system observations.

2.7 The Paper Process

Non-wood materials were in use for papermaking in China almost 2000 years ago, until developed countries adopted the process of producing pulp and paper from wood sources. This process was invented in Germany by Friedrich Gottlob Keller in 1840 [175].
Nowadays, about 90–91% of the world's pulp and paper production is produced from wood [176]. It involves the extraction of cellulose from either hardwood or softwood fibres. The cellulose obtained is processed into pulp, used in papermaking. The world consumption of paper has grown by about 400% in the last 40 years and continues to grow about 2.1% yearly since 2009 with North America, Europe and Asia for more than 90% of total paper and paperboard consumption [177].

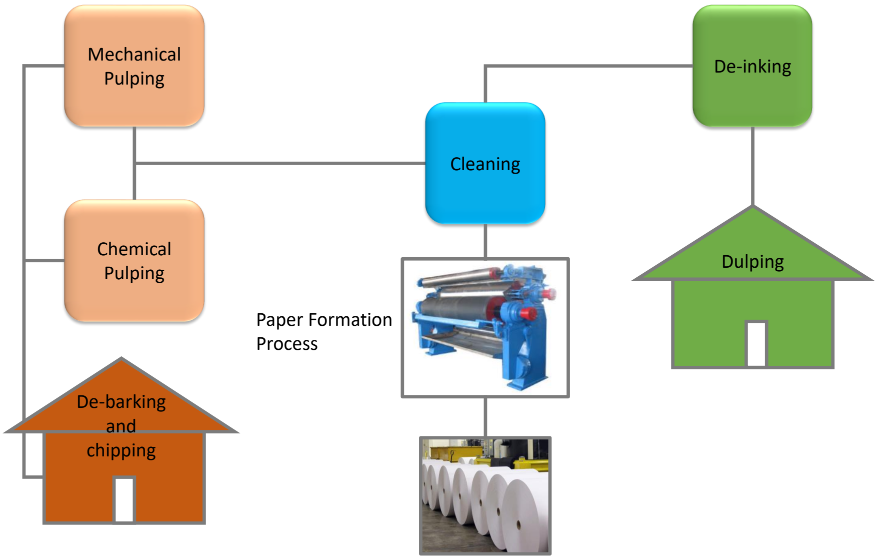


Figure 2.17: Production process of the paper.

Cellulose paper can be divided into four stages: Extraction and preparation of wood; pulping; conversion of wood to brown pulp; conversion of brown pulp to bleached pulp; and the final steps to produce the desired paper. This process is illustrated in Figure 2.17.

Originally, paper was produced by hand as individual sheets until Louis Robert invented the paper machine in France in 1799 [178].
Due to the depletion of wood resources, the use of low-cost raw materials has been introduced to serve as an alternative resource for pulp and paper production [179]. The alternative resources include non-wood fibres, such as agricultural residues and annual plants (plants that germinate, flower, set seed, and die all in one season), considered as valid alternative sources of cellulose for pulp and paper production [180]. Properties that make them suitable include high yielding ability, high pulping quality, good adaptation to prevailing climatic conditions and low-cost [181].

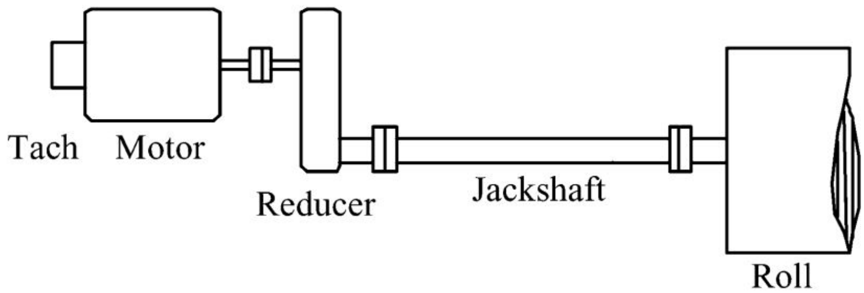


Figure 2.18: Schematic of a paper pulp drying press [182].

A paper machine consists of various mechanical sections, each driven by a motor or an arrangement of a main motor and one or more auxiliary drives and, usually controlled by speed or torque. Typical sections are: Fourdrinier, Press, Dryer, Calender and Reel [182].
Figure 2.18 shows some of the constituent components of the pulp press. A more detailed study of the nature of paper machines can be found in studies [183, 184, 185], in which they show the losses and damage that can occur in these machines. The dewatering and drying of the paper sheet as it passes through the paper machine is very complex and is influenced by many factors [185]. Therefore, it is necessary to monitor all variables underlying the papermaking process in this section.

2.8 Dataset

Organizations have a great concern with their future, which leads many to acquire technologies to make them more flexible and competitive with their competitors. With the acquisition of sensors it is possible to have information about each constituent equipment of the machine. There is a wide range of sensors, not forgetting the IoT technology that allows the digital connection of objects. Some of these technologies are present in Figure 2.19.
For data processing it is necessary to take into consideration the problematic existences which correspond to three analysis methods: univariate analysis which consists in describing a population by examining one variable at a time; bivariate analysis, which aims to study the existing relationships between two variables for the purposes of explanation and/or prediction; and multivariate analysis, which includes the methods analysis the relationships of multiple dependent variables and/or

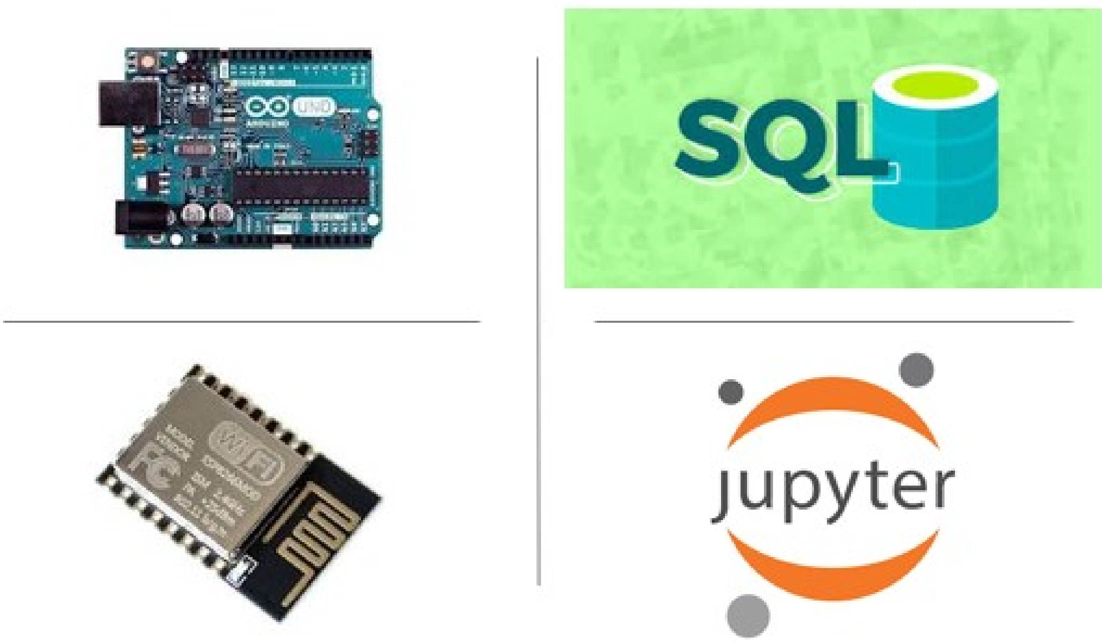


Figure 2.19: Support technologies.

multiple independent variables, whether or not cause/effect relationships are established between these two groups.

For the present thesis ten datasets were used, of these two are of industrial origin and the remaining were obtained from the following sources:

- EURONEXT;
- ECONOMIC RESEARCH;
- MACROTRENDS;
- WORLD STEEL ASSOCIATION.

The data from the pulp paper company has the characteristics described below.

The first dataset contains data samples from 1 February, 2018, to December, 2020, for a total of 33313 samples per hour.

The second dataset contains data samples from 1 January, 2018, to 20 December, 2021, for a total of 33313 samples five minute.In this second set of data is included the variable of paper pulp production in the presses.

The data set has the variables i) Current Intensity: current absorbed by the press motor, in Ampere; ii) Hydraulic Unit Oil Level (in percentage); iii) Torque of the motor (in N.m); iv) VAT Pressure: Pressure inside the cuba (in Kpa); v) Rotation Velocity: velocity of rotation of the press' rolls, in rotations per minute; vi) Temperature at Hydraulic Unit, in degree Celsius, and vii) Paper pulp production .

Figure 2.20 shows the plot of the raw data of the press. As the graph shows, there are zones of typical operation and spikes of discrepant data. The data are not homogeneous. There are many discrepant samples in the extreme quantiles and the distribution of data is not linear.

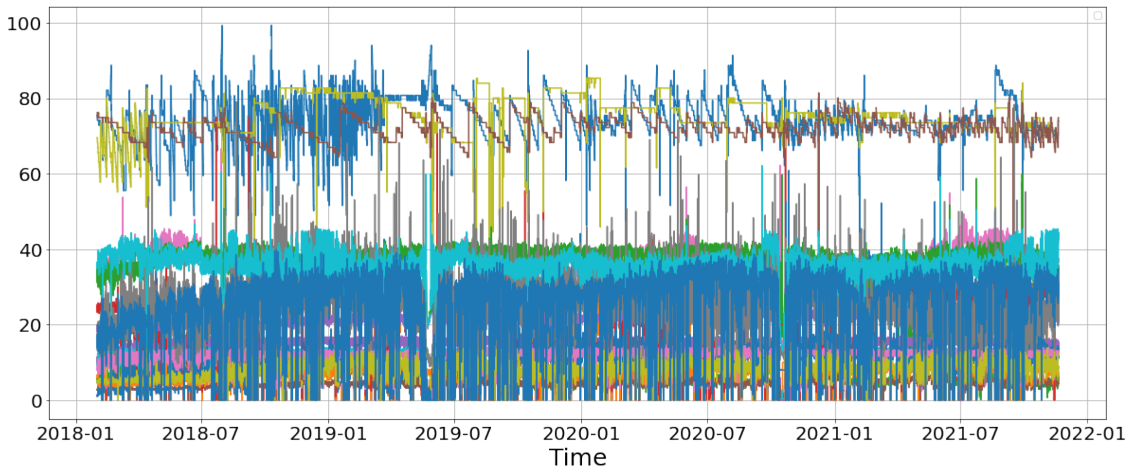


Figure 2.20: Plot of the sensor variables before applying data cleaning treatment.

Data quality is essential for developing effective modeling and planning. Data with discrepant values, as those shown in the charts, can pose difficulties to machine learning models. Therefore, data must be processed and structured prior to analysis.

There are several treatment methods designed for this purpose, but a careful selection is needed so that information is not impaired. In the present work, the approach followed was the quantile method [18]. Table 2.1 describes in all variables for the press one, press three and press four.

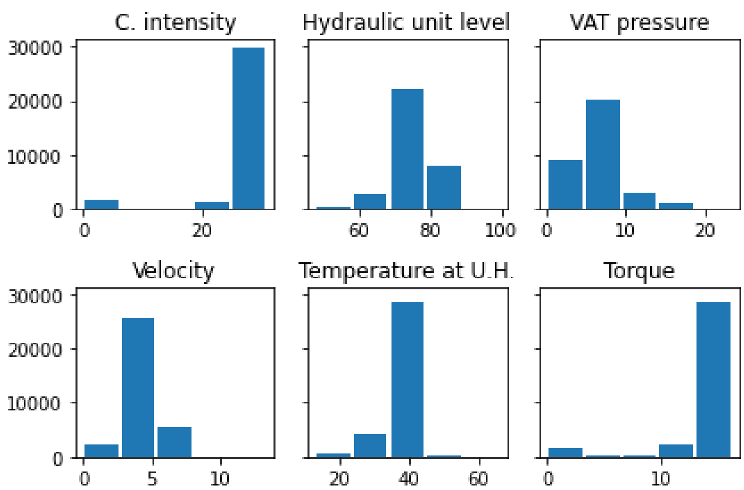


Figure 2.21: Histogram of variables showing the number of samples per quartile.

In Figure 2.21 the histogram of data of the press one is present It shows the dispersion of frequencies data. It is possible to verify that there is a significant variation among them.

Figure 2.22 shows the amplitude of each sample concerning the lower and upper bounds for each variable. As the figure shows, the distribution of data is skewed for all variables.

Table 2.1: Statistical parameters of the dataset variables for the three presses, before processing: $C_intensity$, hydraulic unit oil level, torque, VAT pressure, velocity, and temperature. .

	Count	**Mean**	**std**	**min**	Q_1**—25%**	Q_2**—50%**	Q_3**—75%**	**max**
Hydraulic_unit_level1	399 492	74.45	5.81	44.95	71.47	74.14	78.12	99.34
Production1	398 746	26.50	7.75	0	25.03	29.03	31.26	37.14
VAT_pressure1	399 143	6.71	3.03	0	4.71	5.83	7.97	75.94
C_intensity1	395 991	26.36	6.66	0	26.53	28.27	29.04	75.00
Torque1	396 627	14.01	3.45	0	14.43	15.08	15.47	18.28
Velocity1	396 960	4.33	1.45	0	3.70	4.64	5.15	13.45
Temperature_at_U.H.1	399 491	36.97	3.69	13.08	35.52	37.17	38.80	66.58
C_intensity3	399 480	26.04	7.44	0	26.30	28.63	29.83	50.00
Hydraulic_unit_level3	399 293	75.69	4.88	1.23	73.50	76.13	78.76	85.34
Torque3	399 367	11.65	2.93	0.06	12.02	12.55	12.85	16.75
VAT_pressure3	394 382	9.67	4.04	0	7.12	9.85	12.30	60.87
Velocity3	397 181	6.74	2.21	0	5.81	7.29	8.13	13.94
Temperature_at_U.H.3	399 491	37.38	3.26	14.77	36.29	38.17	39.35	60.24
Production3	368 874	24.48	6.29	0	22.73	26.45	28.39	34.42
C_intensity4	395 611	20.31	4.96	0.02	20.06	21.56	22.42	41.57
Hydraulic_unit_level4	399 492	72.16	2.90	43.33	70.91	72.23	73.56	81.44
Torque4	383 119	11.83	1.85	0	11.82	12.27	12.62	17.49
VAT_pressure4	399 463	24.18	5.79	0.01	21.30	25.00	28.10	80.91
Velocity4	396 221	7.95	2.66	0	6.67	8.63	9.71	14.40
Temperature_at_U.H.4	399 491	35.77	3.57	12.93	34.31	35.57	37.41	62.16
Production4	399 194	26.90	7.97	0.00	25.31	29.53	31.67	39.85

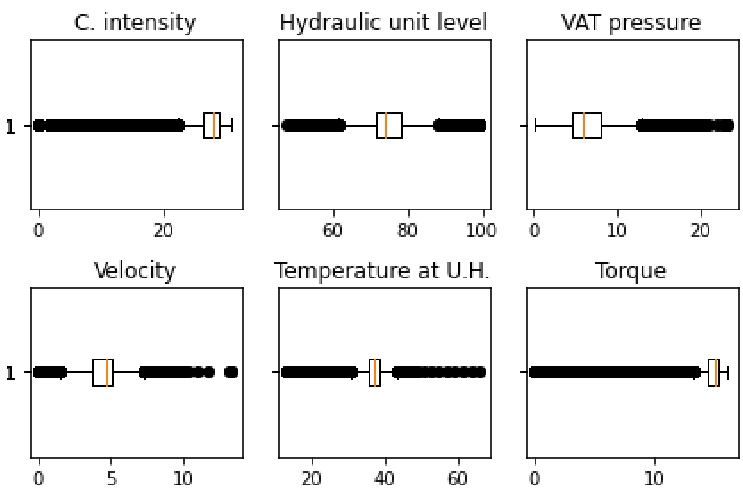


Figure 2.22: Distribution of data points of all the sensors, with Low and High extremes.

In Figure 2.23 it is verified individually each variable autocorrelations that decay relatively quickly. In this there was the need to resort to data processing methodologies in order to eliminate discrepant data and also to improve the autocorrelation of the variables.

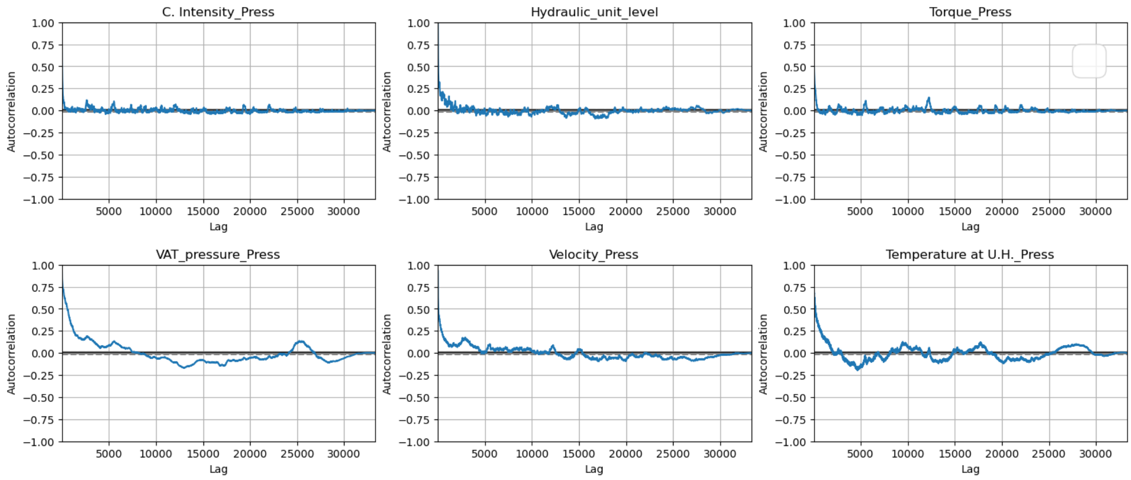


Figure 2.23: Autocorrelation between samples of all variables, calculated for 200 days.

2.9 Data Pre-processing

2.10 Conclusion

Since operations management is a practice that involves planning, execution and monitoring of actions within a company, one of the great challenges is to keep the production flow at a high level so that there are no interruptions and these are converted into losses.

Many of these failures may present certain patterns that make it impossible to explore them in order to perform a prediction based on these patterns. The predictive maintenance policy, supported by the mathematical tools mentioned has been widely explored by researchers to solve problems related to various unwanted during the manufacturing process.

Being that a machine a set of components, failure of a Component can compromise the entire operation, for this, if there is possibility to monitor these components can have an advantage over what may happen.

In order to be able to make a forecast, it is first necessary to have data. This data is very important as it contains information about the machine's operation and can have different sampling periods.

Since a machine is made up of several components it is important to have these concepts since they allow us to deal not only individually with the phenomena that may exist in the variables but also allows us to understand them through the relationships that may exist between the variables. Using the documentation available in web of science, scopus and google academic databases, we can validate a global approach for the maintenance of results in production, and we can validate the use and efficiency of several tools in solving related problems.

In the maintenance field, some research is applied to solutions that can detect and diagnose failures based on this method, as well as the efficiency of neural networks and principal component analysis.

The cumulative costs associated with plant failures are significant. For this reason, maintenance methods have evolved to deal with these dilemmas, such as predictive maintenance (PM), which consistently demonstrates the ability to maintain the integrity of the company by generating information about the condition of the equipment; this data makes maintenance effective.

The state of the art proved to be sound, leading to a scientific article published in an indexed journal with the intention of sharing these conclusions with other researchers willing to address this type of problem (Appendix A).

Chapter 3

Methodology

This chapter presents the methodologies applied to process treat data before the forecast algorithms are applied. The data set and its respective variables present behaviours that must be taken into account, and these behaviours can be seen in the representation of data in time series. The chapter demonstrates the behaviour of data before and after the application of a data processing method and beyond, as well as the autocorrelation of variables and the correlation between them. Finally, the chapter presents the forecasting tools from the most classical to the most complex that will be considered for the forecasting tests.

3.1 Processing Data Method

3.1.1 Interquartile Range (IQR) Method

The method of eliminating discrepant values is based on the idea that extreme values are most probably data reading failures. They often happen due to sensor failures, communication interference or other type of problems during data acquisition. As a result, the dataset sometimes contains invalid samples such as readings outside of the expected sensor ranges, or zero when the machine was stopped. Those samples can be eliminated, so that they do not negatively affect the machine learning process.

In the present work, limits were calculated for each variable and the samples out of the allowed range were replaced by the average. The limits were calculated using the following equations:

$$Q_{\frac{1}{4}} = \frac{1}{4}(n+1) \quad (3.1)$$

$$Q_{\frac{3}{4}} = \frac{3}{4}(n+1) \quad (3.2)$$

$$IQR = Q_{\frac{1}{4}} - Q_{\frac{3}{4}} \quad (3.3)$$

$$Down_{limit} = Q_{\frac{1}{4}} - K \times IQR \quad (3.4)$$

$$Up_{limit} = Q_{\frac{3}{4}} + K \times IQR \quad (3.5)$$

$Down_{limit}$ is the lower limit accepted for the variable, calculated by subtracting the constant k multiplied by IQR to $Q_{\frac{1}{4}}$. Up_{limit} is the upper limit accepted for the variable, calculated by adding the constant k multiplied by IQR to $Q_{\frac{3}{4}}$, where k is the constant of variation of the limits. The limits are calculated for each variable. Sample data points that contain values that are out of the interval $[Down_{limit}, Up_{limit}]$ are replaced by the average.

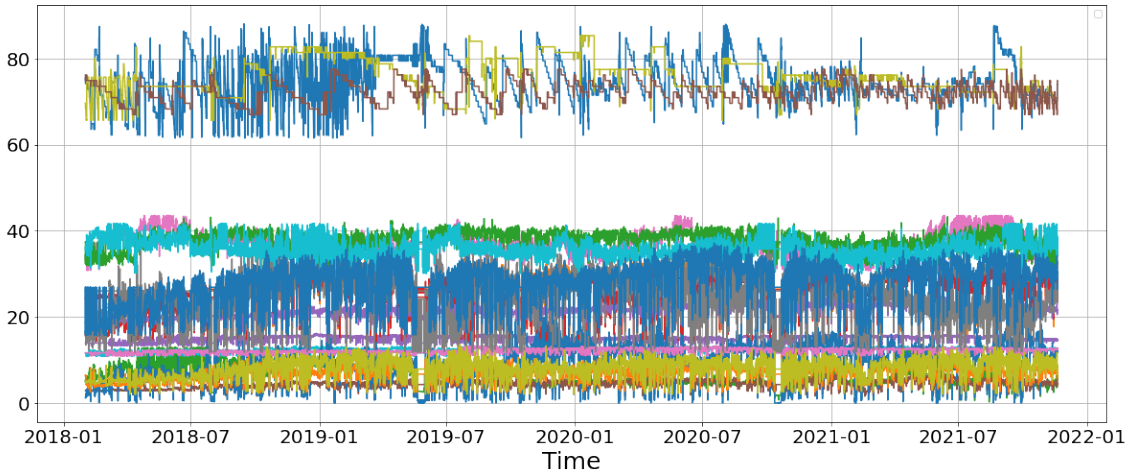


Figure 3.1: Plot of the dataset variables without extreme values.

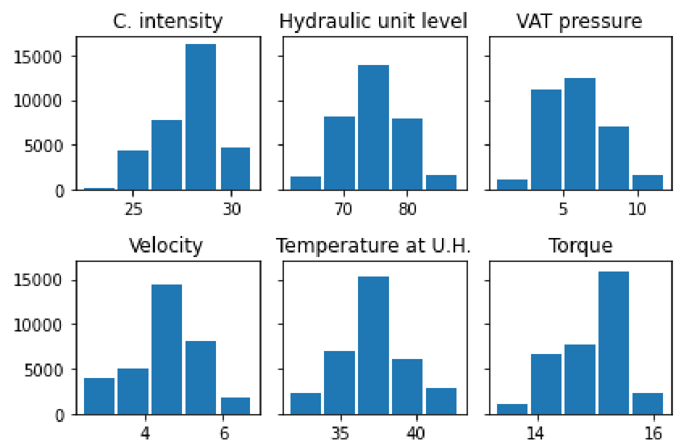


Figure 3.2: Histogram of variables after removing discrepant data. The variables are Current intensity, Hydraulic unit level, VAT pressure, Rotation velocity, Temperature in the Hydraulic Unit, and Torque.

Use of the quantile method removes extreme values, which are often due to sensor reading errors, stops, or other abnormal situations. After those samples are removed, it is possible to see more normal data distributions, such as those shown in Figure 3.1 and Figure 3.2.

Removing the extremes, it can be seen in Figure 3.3 that the variables show more balanced distribution. Also it is possible to see in Figure 3.4 that there is a significant increase in lag.

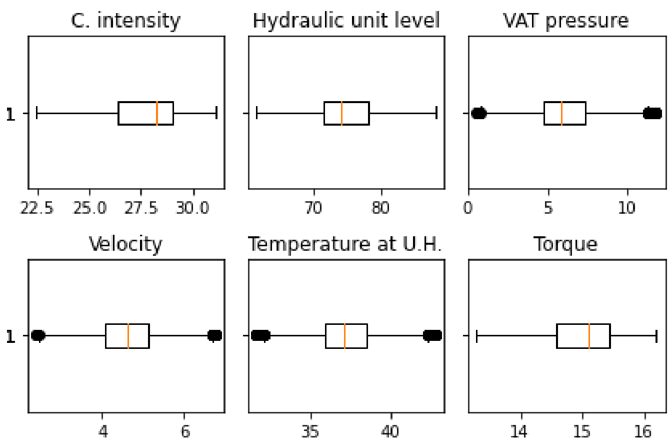


Figure 3.3: Distribution of data points of all the sensors, without Low and High extremes.

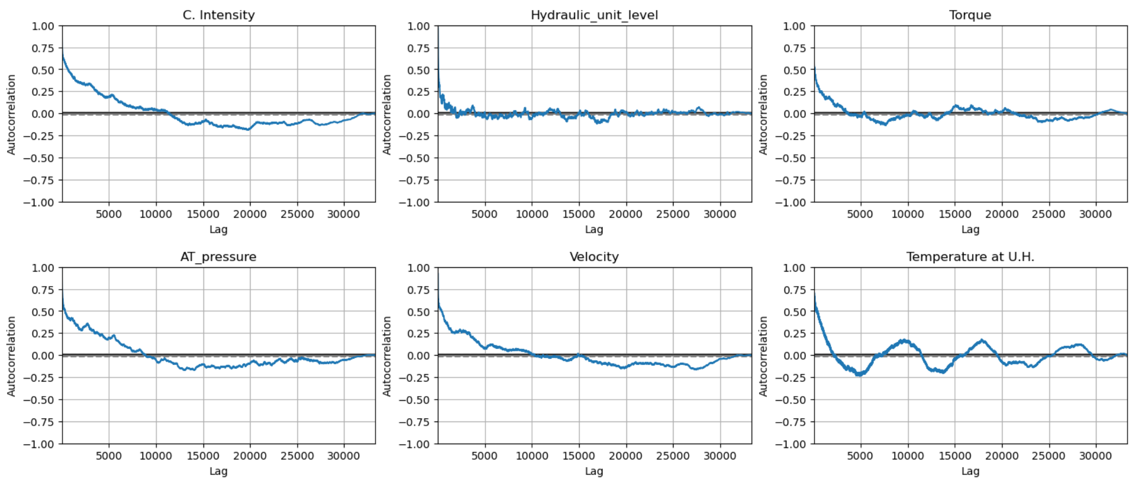


Figure 3.4: Autocorrelation between samples of all variables, calculated for 200 days.

3.1.2 Data Filtering using LOWESS

LOWESS/LOESS (locally weighted/estimated scatterplot smoothing) is a non-parametric regression technique developed by Cleveland [186]. Robust locally weighted regression is a method for smoothing variables, $(x_i, y_i), i = 1, \cdots, n$, in which the fitted value at z_k is the value of a polynomial, fit to the data using weighted least squares, where the weight for (x_i, y_i) is large if x_i is close to x_k and small if it is not. The number of samples (n) used for each local approximation (z_k) is a parameter of the model. The degree of the polynomial function is also a parameter of the model. Often the polynomial degree is 1, which means a linear regression is performed.

The samples were registered with sampling period of 1 min for press number 2 and 5 min for press number 4. For most of the experiments the dataset was downsampled, in order to reduce processing time. The downsampling rate varied, although most of the time the 12 or 60 samples of each hour are averaged, which is equivalent to using a sampling period of 1 hour.

Using this approach, Figure 3.5 shows that although there is no elimination of extremes, there is a smoothing in them. Figure 3.6 shows the respective histogram, Figure 3.7 shows distribution of sensor data points and the Figure 3.8 shows the variable autocorrelation.

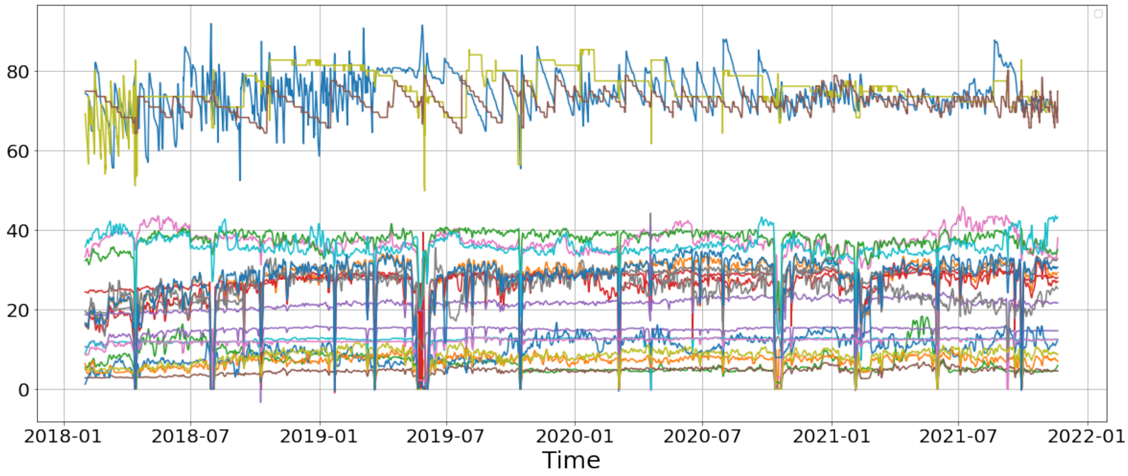


Figure 3.5: Plot of the dataset variables without extreme values: Current intensity, Hydraulic unit level, VAT pressure, Rotation velocity, Temperature in the Hydraulic Unit, and Torque.

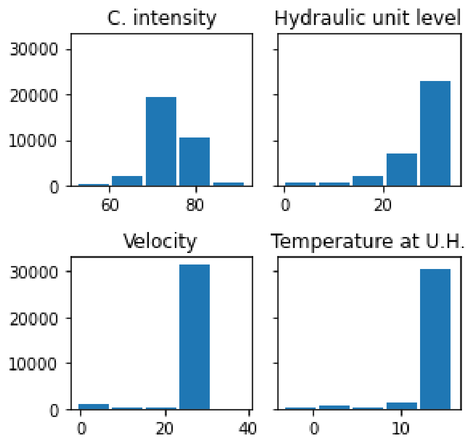


Figure 3.6: Histogram of variables after removing discrepant data. The variables are Current intensity, Hydraulic unit level, VAT pressure, Rotation velocity, Temperature in the Hydraulic Unit, and Torque.

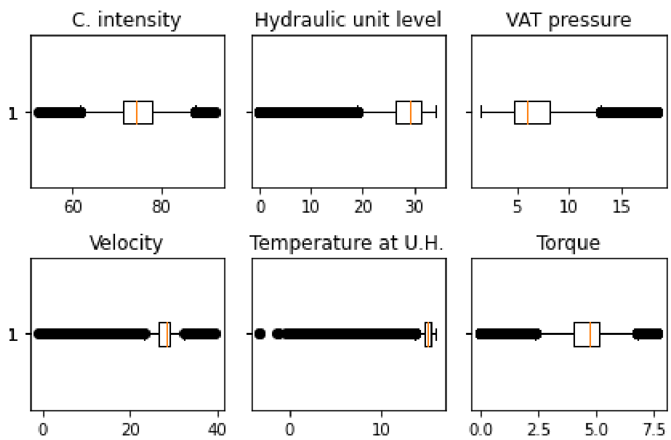


Figure 3.7: Distribution of data points of all the sensors, with some Low and High extremes.

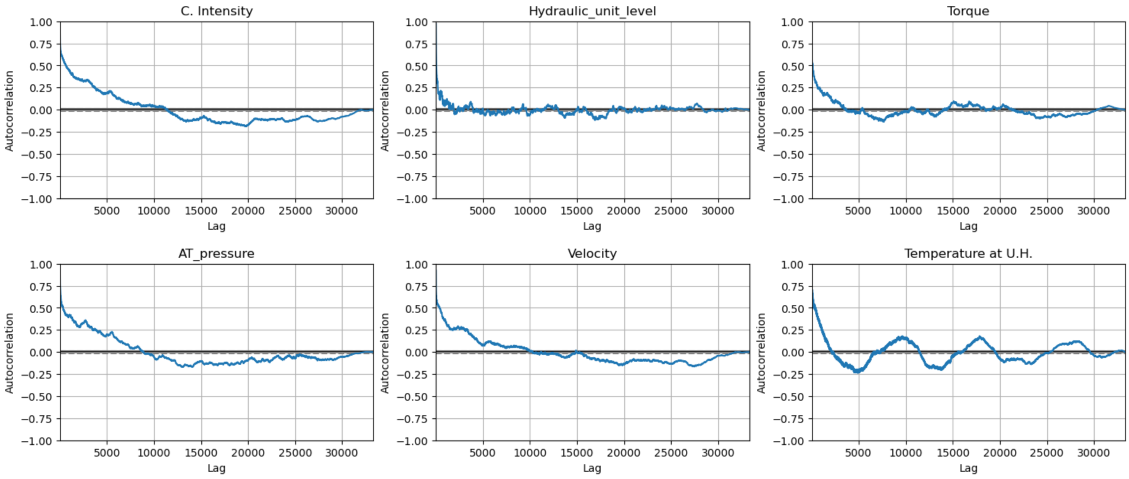


Figure 3.8: Autocorrelation between samples of all variables, calculated for 200 days.

For the tests of the short- and long-term forecasting methods, we used resample, a function from the Pandas library. With this function, we were able to run the prediction tests with an Asus i5 computer, 6GB Ram, 500GB rum storage and a VIDIA GeForce GT 720M graphics card.

3.1.3 Correlation of Variables

Correlation is a statistical measure of the relationship between two variables. This measure is best for variables that demonstrate a linear relationship between them. The fit to the data can be represented visually in a scatter plot. Using scatter plots, we can often assess the relationship between variables and determine if they are related [187].

Correlation coefficient formulas are used to find out how strong a relationship between data is. The formulas return values ranging from -1 to 1, where:

- 1 indicates a strong positive relationship.
- -1 indicates a strong negative relationship.
- A score of zero indicates no relationship.

In Figure 3.9 we can see that there is a high correlation on the values of pulp produced, in relation to the current and the torque. These correlation values reach 0.9, being the maximum value of the correlation 1. One can validate the feasibility of the predictions of these variables together.

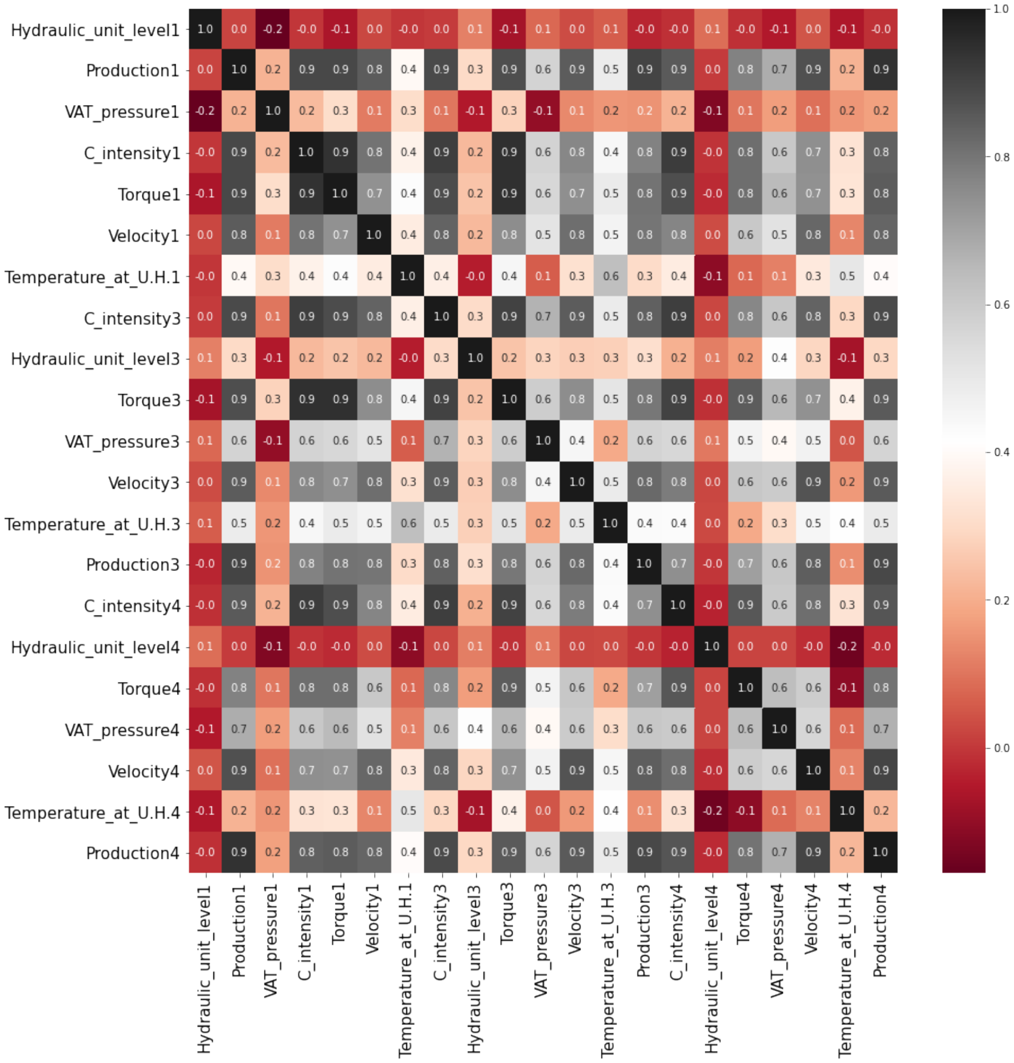


Figure 3.9: Correlation between all variables of the paper pulp presses.

3.2 Forecasting with time series methods

3.2.1 Forecasting for AR Model

Since the number of the samples is very high, there was a need to down sample the data-set, from a period of minutes to a period of days, in order to have a forecast in days. That was done by averaging the samples of each day using the python pandas function *df.resample ('D'), Mean ()*.

As a first approach to predict future behaviour, an autoregressive model was applied. Autoregressive models are adequate to model variables that mostly on their previous behavior and a stochastic value, thus satisfying the following equation:

The notation AR(p) refers to the autoregressive model of order p. The AR(p) model is written as

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \quad (3.6)$$

Where means X measured in time period t, p is the order,$\varphi_1, \cdots, \varphi_p$ are real parameters and ε_t is a white noise process independent and identically distributed.

3.2.2 Forecasting for MA

MA(q) refers to the moving average model of order q:

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \quad (3.7)$$

where $\theta_1, \cdots, \theta_q$ are the parameters of the model, μ is the expectation of X_t (often assumed to equal 0), and the $\varepsilon_{t-1}, \cdots$ are white noise error [188].

3.2.3 Forecasting for ARIMA Model

The notation ARMA(p, q) refers to the model with p autoregressive terms and q moving-average terms. This model contains the AR(p) and MA(q) models [189].

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \quad (3.8)$$

3.2.4 Forecasting for SARIMA Model

Some time series present a seasonal periodic component. A seasonal autoregressive model is characterized by the existence of a significant correlation between observations spaced by a multiple time interval Khandelwal et al. [190].

Using the data for the press 4, Figure 3.10 presents the data with a sampling rate per week it is possible to verify that there is has seasonality of great relevance, but it is possible to verify that

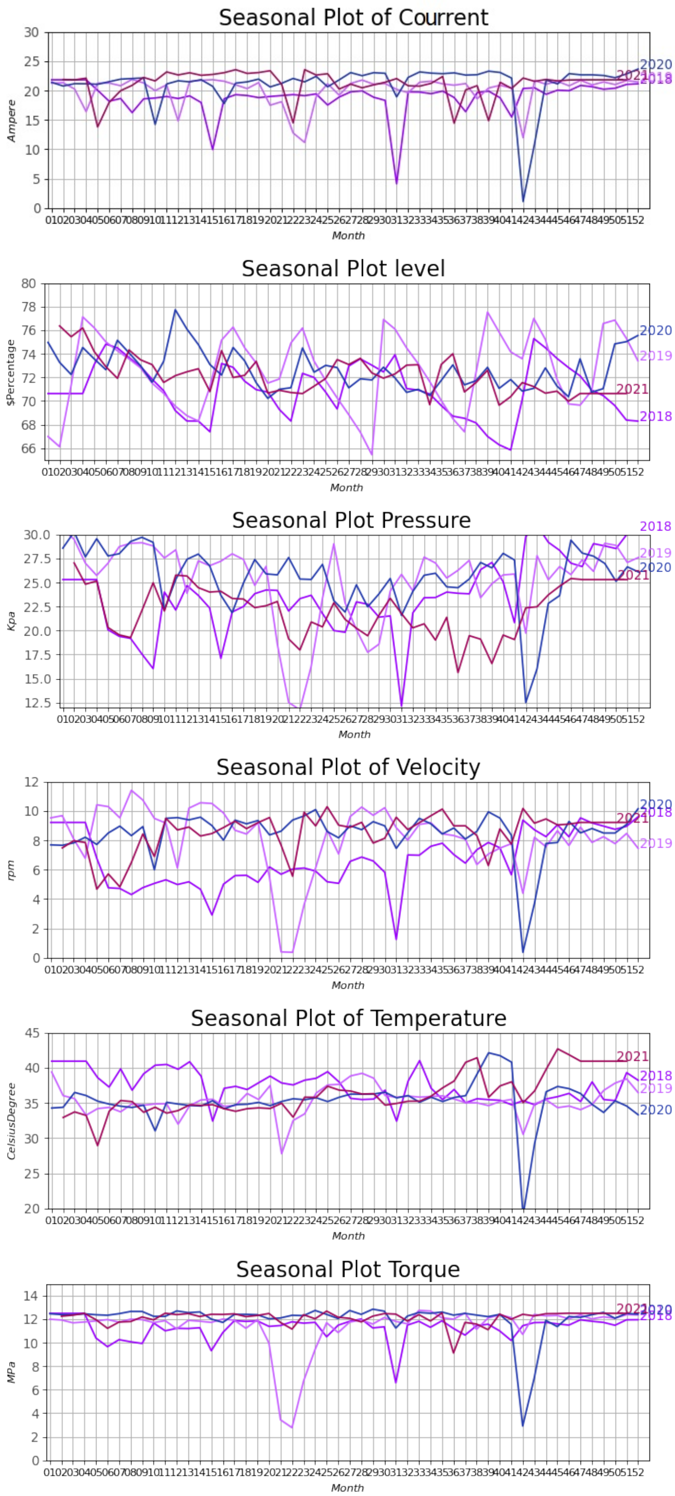


Figure 3.10: Data sample for week press 4.

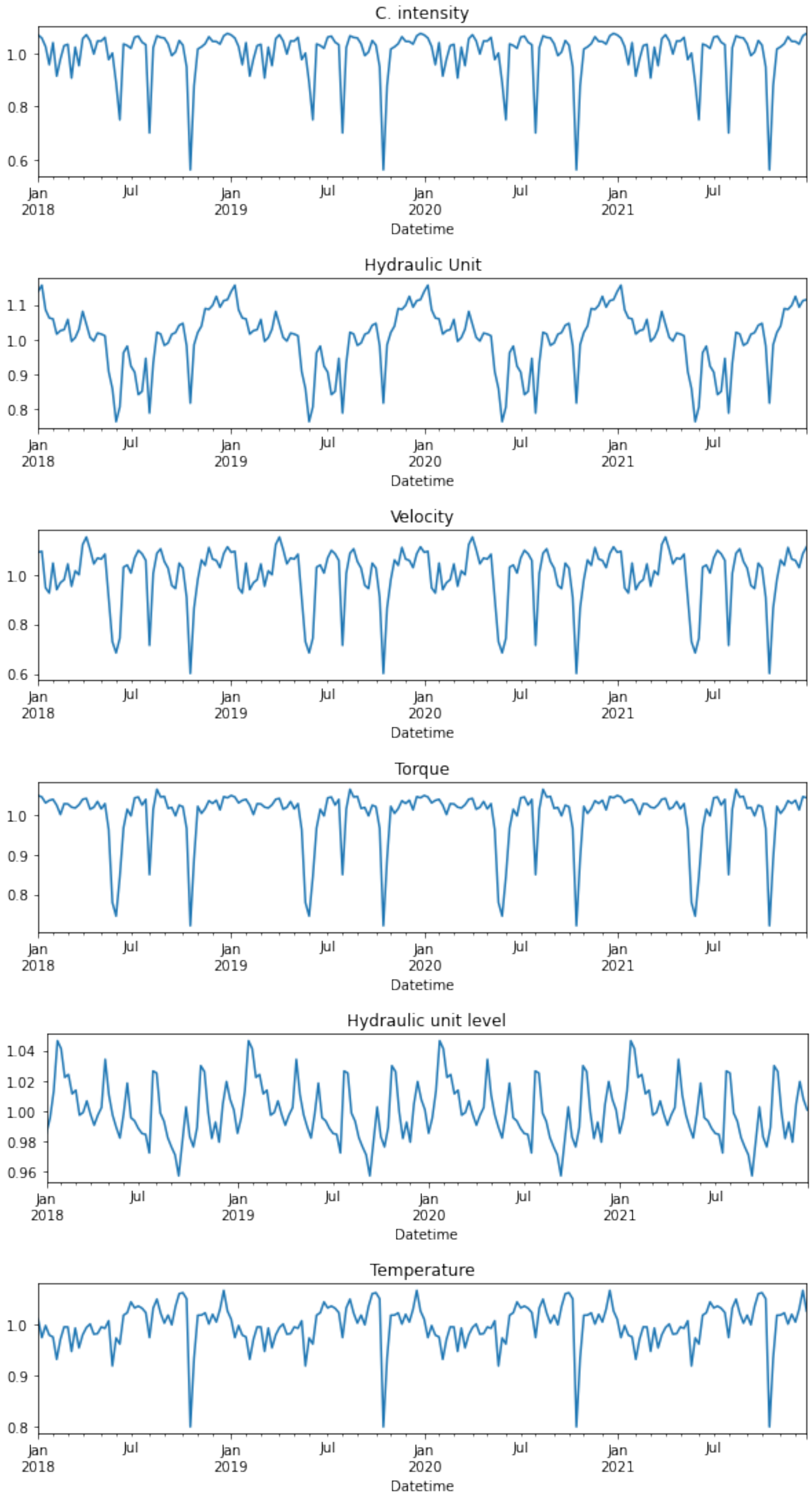


Figure 3.11: Seasonality on press 4 in a week period.

there are variations that present a great relevance. this concluson is also drawn through Figure 3.11.

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is a general case of the models proposed by Yip et al. [191], for the adjustment of stationary time series. However, when there is a seasonal component in the data, the model class is called SARIMA (p, d, q) (P, D, Q), given by:

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t + \sum_{i=1}^{P} \Phi_i X_{t-si} + \sum_{i=1}^{Q} \Theta_i \varepsilon_{t-si}. \quad (3.9)$$

Where (p, d, q) refer to the model orders of the seasonal part: p is trend autoregression order, d is trend difference order and q is trend moving average order. (P, D, Q) is the same but with the Seasonal component. The parameters $\Phi_1 \cdots \Phi_P$, are the parameters referring to the seasonal autoregressive part and $\Theta_1 \cdots \Theta_Q$, are the parameters of moving averages, and i is an error that cannot be estimated from the model and D indicates the number of seasonal differences made in the series to make it stationary. The calculation of the parameters of the models that best fits, was made using the most frequent Akaike Information Criterion (AIC), which is defined by:

$$AIC = 2log(L.K) + 2(K) \quad (3.10)$$

where $L.K$ is the maximized log-likelihood and K is the number of parameters in the model.

3.3 Forecasting with deep learning

3.3.1 LSTM With Encoder and Decoder

The LSTM is a deep learning recurrent neural network architecture that is a variation of traditional recurrent neural networks (RNNs). It was introduced by Hochreiter and Schmidhuber in 1997. The most popular version is a modification refined by many works in the literature [192, 193], which is called vanilla LSTM (hereinafter referred to as LSTM).

The LSTM is excellent at handling time series data only with its network parameters. For example, weights and polarization are adjusted or optimized [194]. The primary modification of the LSTM when compared to the RNN architecture is the structure of the hidden layer [195]. The LSTM model is a powerful type of RNN capable of learning long-term dependencies [196].

Kong et al. demonstrate some relevant conclusions such as (1) LSTM has a good predictive capacity; (2) their use can significantly improve the profit of service providers, so there is an opportunity when it comes to exploring the forecast in real time [197].

Ayyaz and Alpay apply Long Short-Term Memory (LSTM) neural network approaches to predict real production data, obtaining satisfactory results, superior to conventional models [198]. In their study to improve maintenance planning to minimize unexpected stops, they apply a new method that consists of the combined use of decomposition in empirical mode of ensemble and long-term memory. Their results showed a performance superior to other state of the art models. Essien and Giannetti [199] use a deep convolutional LSTM encoder-decoder architecture model on real data, obtained from a metal packaging factory. They show that it is possible to perform combinations of LSTM with other networks to significantly improve the results. LSTM neural networks achieved the best performance in a range of computational sequence labeling tasks, including speech recognition and machine translation [200]. There are a variety of engineering problems that can be solved with predictive neural models. The architecture of an LSTM network includes the number of hidden layers and the number of delay units, i.e., the number of previous data points considered for training and testing. Currently, there is no general rule for selecting the number of delays and hidden layers [201]. A deep LSTM can be built by stacking multiple LSTM layers, which generally works better than a single layer. Deep LSTM networks have been used to solve many real-world sequence modelling problems [202].

Figure 3.12 shows the internal architecture of an LSTM unit cell. According to [203], the internal calculation formulae of the LSTM unit are defined as follows:

$$i_t = \sigma(x_t U^i + h_{t-1} W^i + b_i) \quad (3.11)$$

$$f_t = \sigma(x_t U^f + h_{t-1} W^f + b_f) \quad (3.12)$$

$$o_t = \sigma(x_t U^o + h_{t-1} W^o + b_o) \quad (3.13)$$

$$a_t = tan(x_t U^C + h_{t-1} W^C + b_C) \quad (3.14)$$

where U^i, U^f, U^o and $U^C$are the weight matrices for mapping the current input layer on three gates and the state of the current input cell.

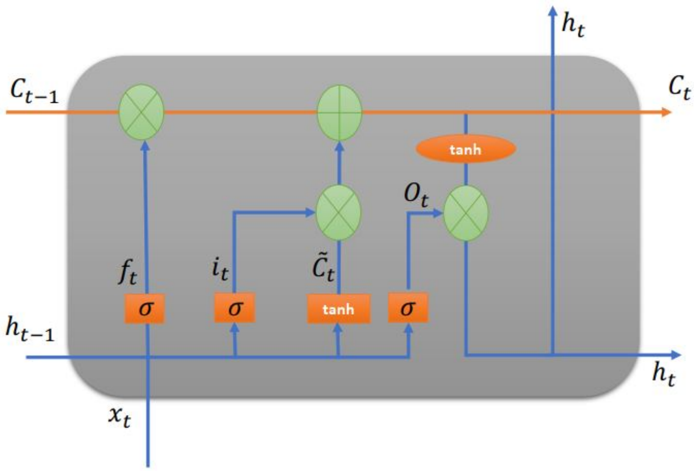


Figure 3.12: Detailed layout of a long short-term memory unit [204].

Matrices W_q and U_q contain the weights of the input and recurrent connections, where the index can be the input gate i, output gate o, the forgetting gate f or the memory cell c, depending on the activation being calculated. $c_t \in \mathbb{R}^h$ is not just a cell of an LSTM unit, but contains h cells of the LSTM units, while i_t, o_t and f_t represent the activations of, respectively, the input, output and forget gates, at time step t, where:

- $x_t \in \mathbb{R}^d$: input vector to the LSTM unit;
- $f_t \in (0,1)^h$ forget gate's activation vector;
- $i_t \in (0,1)^h$ input/update gate's activation vector;
- $o_t \in (0,1)^h$ output gate's activation vector;
- $h_t \in (-1,1)^h$ hidden state vector, also known as the output vector of the LSTM unit;
- $\tilde{c_t} \in (-1,1)^h$ cell input activation vector;
- $c_t \in \mathbb{R}^d$: cell state vector.

$W \in \mathbb{R}^{h\times d}$, $U \in \mathbb{R}^{h\times h}$ and $b \in \mathbb{R}^h$ are weight matrices and bias vector parameters, which need to be learned during training. The indices d and h refer to the number of input features and number of hidden units.
W^i, W^f, W^o and W^C are the weight matrices for mapping the previous output layer on three gates and the current state of the input cell. b_f, b_i, b_o, and b_c are polarization vectors for calculating the state of the gate and the input cell. σ is the gate activation function, which is normally a sigmoid function. tan is the hyperbolic tangent function which is the activation function for the current state of the input cell.

3.3.2 Gated Recurrent Unit With Encoder and Decoder

It shows that the Gated recurrent unit (GRU) has an updated gate and a reset gate similar to forget and input gates on the LSTM unit. The refresh gate defines how much old memory to keep, and the reset gate defines how to combine the new entry with the old memory. The main difference is that the GRU fully exposes its memory content using just integration (but with an adaptive time constant controlled by the update gate).
The GRU was introduced by Cho et al. [205]. Although it was inspired by the LSTM unit, it is considered simpler to calculate and implement. It retains the LSTM immunity to the vanishing gradient problem. Its internal structure is simpler and, therefore, it is also easier to train, as less calculation is required to upgrade the internal states. The update gate controls the extent to which the state information from the previous moment is retained in the current state, while the reset gate determines whether the current state should be combined with the previous information [205].
Figure 3.13 shows the internal architecture of a GRU unit cell. These are the mathematical functions used to control the locking mechanism in the GRU cell:

$$z_t = \sigma(x_t W^z + h_{t-1} U^z + b_z) \quad (3.15)$$

$$r_t = \sigma(x_t W^r + h_{t-1} U^r + b_r) \quad (3.16)$$

$$\tilde{h}_t = tan(r_t \times h_{t-1} U + x_t W + b) \quad (3.17)$$

$$h_t = (1 - z_t) \times \tilde{h}_t + z_t \times h_{t-1} \quad (3.18)$$

where W^z, W^r, W denote the weight matrices for the corresponding connected input vector. U^z, U^r, U represent the weight matrices of the previous time step, and b_r, b_z and b are bias. The σ denotes the logistic sigmoid function, r_t denotes the reset gate, z_t denotes the update gate, and $\tilde{h}_t$ denotes the candidate hidden layer [206].

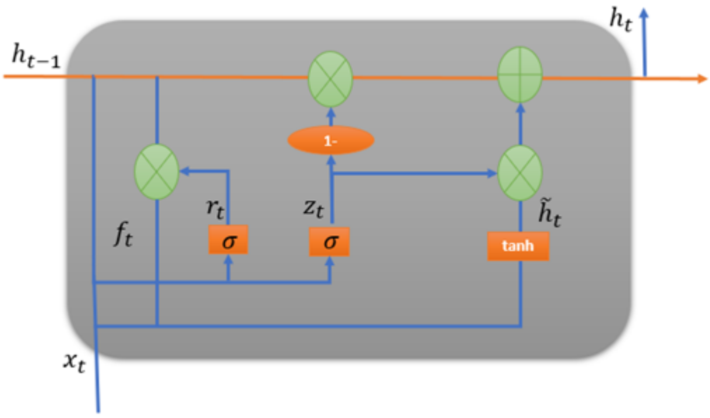


Figure 3.13: The cell structure of a gated recurrent unit.

It shows that the GRU has an updated gate and a reset gate similar to forget and input gates on the LSTM unit. The refresh gate defines how much old memory to keep, and the reset gate defines how to combine the new entry with the old memory. The main difference is that the GRU fully exposes its memory content using just integration (but with an adaptive time constant controlled by the update gate).

3.4 Model Evaluation

In the present experiments, LSTM and GRU neural network models are compared. To evaluate the model prediction performance, the models used were root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), and mean absolute error (*MAE*). They are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y})^2} \quad (3.19)$$

where Y_t is the actual data value and $\hat{Y}$ is the forecast obtained from the model value. The prediction error is calculated as the difference between Y and $\hat{Y}$, i.e., the difference between the output desired and the output obtained. n is the number of samples used in the test set.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|Y_t - \hat{Y}_t| \quad (3.20)$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|Y_t - \hat{Y}_t|}{|Y_t|} \quad (3.21)$$

3.5 Conclusion

The application of a classical predictive model and a machine learning model requires sensitivity to the data structure, i.e. the data must be treated as a model input, and many data may have non-numeric formats, duplicate values, null values and discrepant values.

The chapter shows that data processing is essential for checking the behaviour of variables over time, such as trend and seasonality. The trend and seasonality did not provide enough evidence to be considered in the study.

With the processing of the data, there was an increase in the autocorrelation of the variables, which gave results of considerable accuracy with classical forecasting models, although there were limitations. With the proposal of the recurrent neural network model, these limitations were overcome.

The results of the method research and testing were published at the INCOME-VI AND TEPEN 2021 conference in China (Appendix B), and many of the methods presented that support the development of the following chapters.

Chapter 4

Tests and Results with Maintenance Data.

In this chapter the different results from different methodologies of data prediction presented in the previous chapter are used for the tests based on the data of press two. The tests aim to present the best parameters and hyperparameters in order to obtain better prediction results, and the methods of data treatment were also considered to achieve better results.

4.1 Test with Autorregressive and SARIMA model

As can be seen from the last section, production management brings many challenges, one of these is short and long term planning. There are tools that can help to support these problems such as predictive tools. It should be emphasized the extreme importance that maintenance policies, namely the predictive maintenance policy, have in avoiding failures that can bring sudden stops.

After analyzing and processing the data, the prediction test was performed using the autoregressive model in order to validate this model's performance against the others presented in the previous chapter. A sliding window of 10 days was applied to the model, corresponding to 1440×10=14400 data samples and a forecast window of the next 30 days that corresponds to $1440 \times 30 = 43200$.

In order to perform the test it was necessary to reduce the sampling rate since the computer used was not robust enough to make a prediction with a very high sampling rate (one sample per minute), thus giving a capacity error. Reducing the sampling rate to hourly, it was noted that although it was possible to perform the prediction, reach a point where it stabilized as shown in Figure 4.1.

With the reduction of the sampling rate using again the AR forecasting model, the results are stable for a 10-day forward forecast with a 30-day backward sliding window as shown in Figure 4.2.

Figure 4.3 shows that the ARIMA method gives stable predictions, when using 10 days sliding window.

Figure 4.4 shows SARIMA model forecasts with a 10-day delay window and a 30-day forward forecast.

Table 4.1 shows the results of the forecasting errors in the period referring to the two models, AR and SARIMA. The table shows the Mean Average Error (MAE), the Mean Squared Error (MSE) and the Mean Average Percent Error (MAPE).

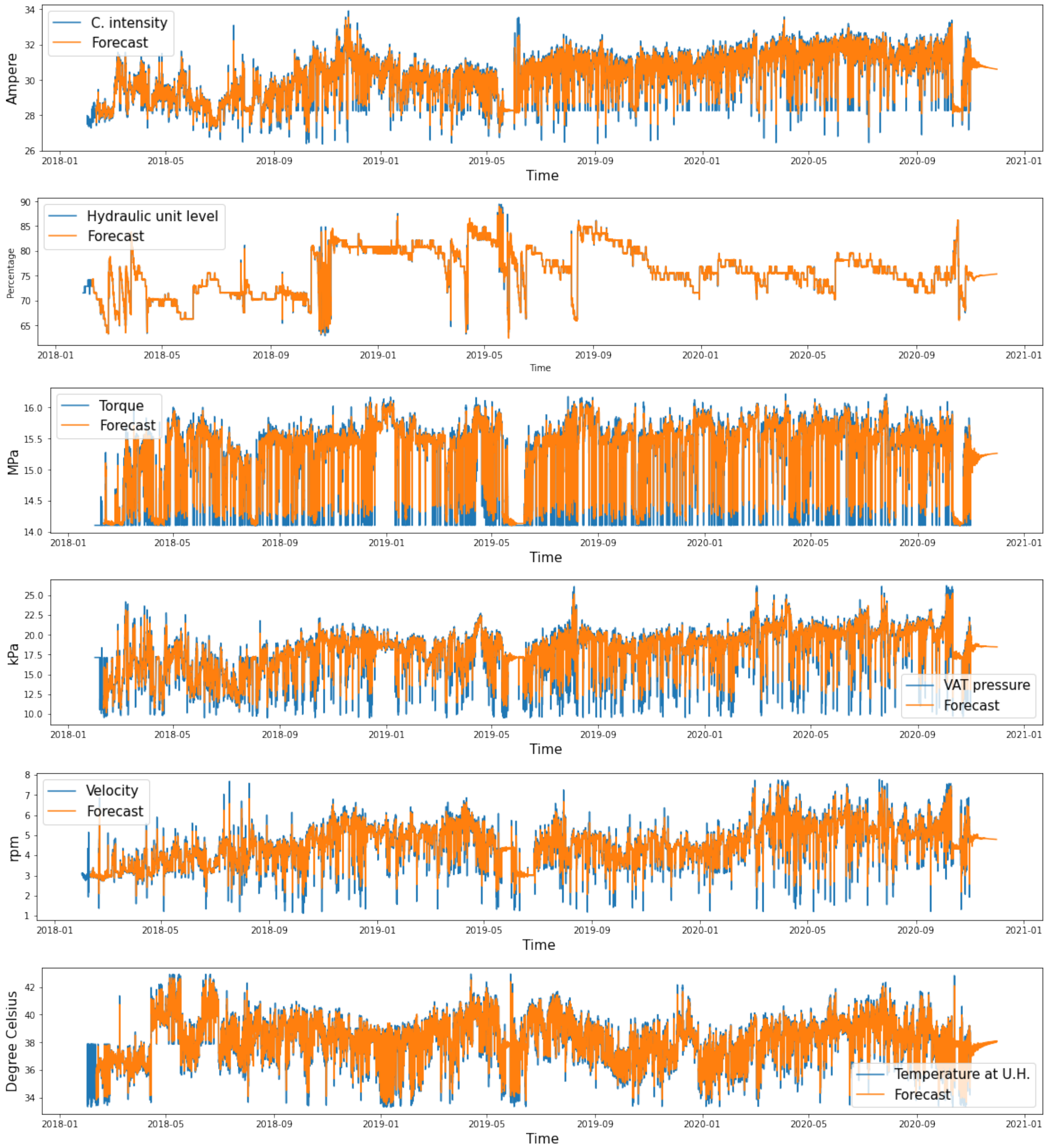


Figure 4.1: Prediction of the six variables with sample rate per hour, using Autorregressive method, with 10 days lag and 30 days forward.

Table 4.1: Forecasting errors of the classical models.

		C. intensity	Hydraulic unit level	Torque	VAT pressure	Velocity	Temp. at U.H.
	RMSE	1.76	2.64	0.77	2.03	1.30	1.53
AR	**MAPE**	4.63	2.89	4.64	10.89	22.79	3.50
	MAE	1.43	2.05	0.65	1.78	1.12	1.26
	RMSE	1.45	3.00	0.78	2.90	0.81	1.50
ARIMA	**MAPE**	4.50	3.78	4.86	15.39	13.36	3.42
	MAE	1.35	2.79	0.68	2.39	0.65	1.23
	RMSE	2.71	2.92	0.94	2.20	2.25	1.23
SARIMA	**MAPE**	6.99	3.69	5.37	11.27	43.91	2.95
	MAE	2.16	2.71	0.78	1.90	2.09	1.08

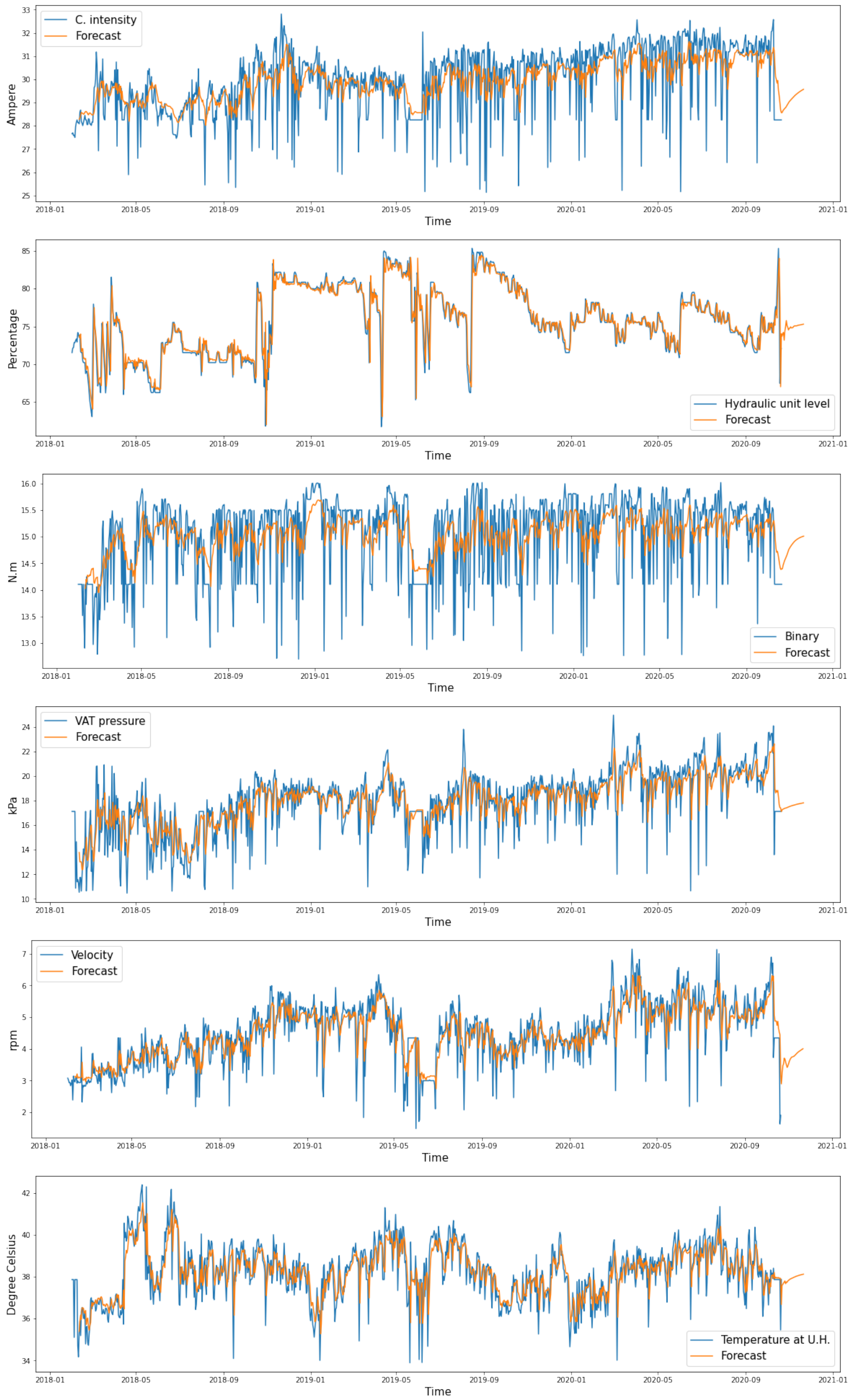


Figure 4.2: Prediction of the six variables with sample rate per day, using a Autorregressive method, with 30 days lag and 10 days window.

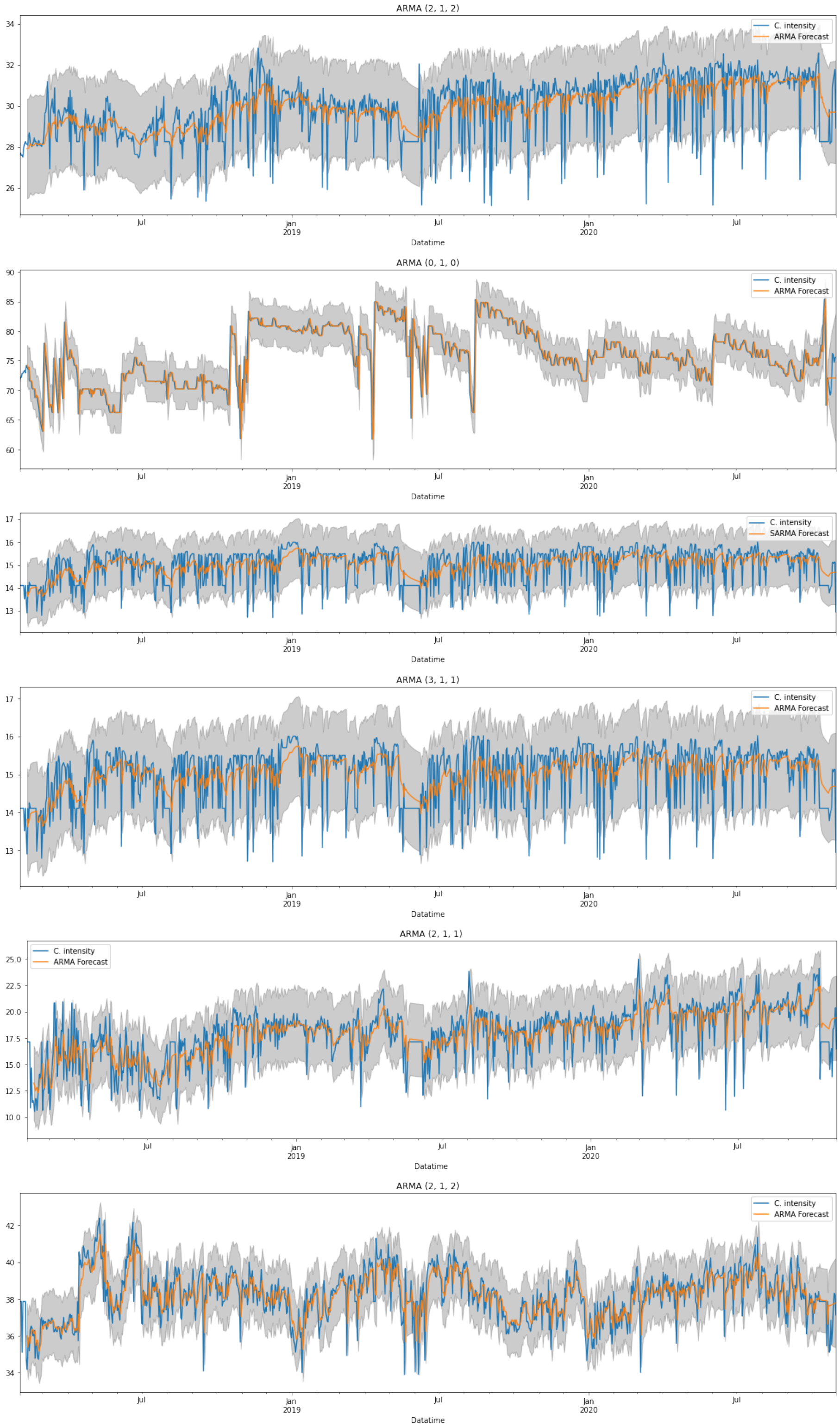


Figure 4.3: Prediction of variables using the ARIMA method, with 10 days lag and 30 days forward.

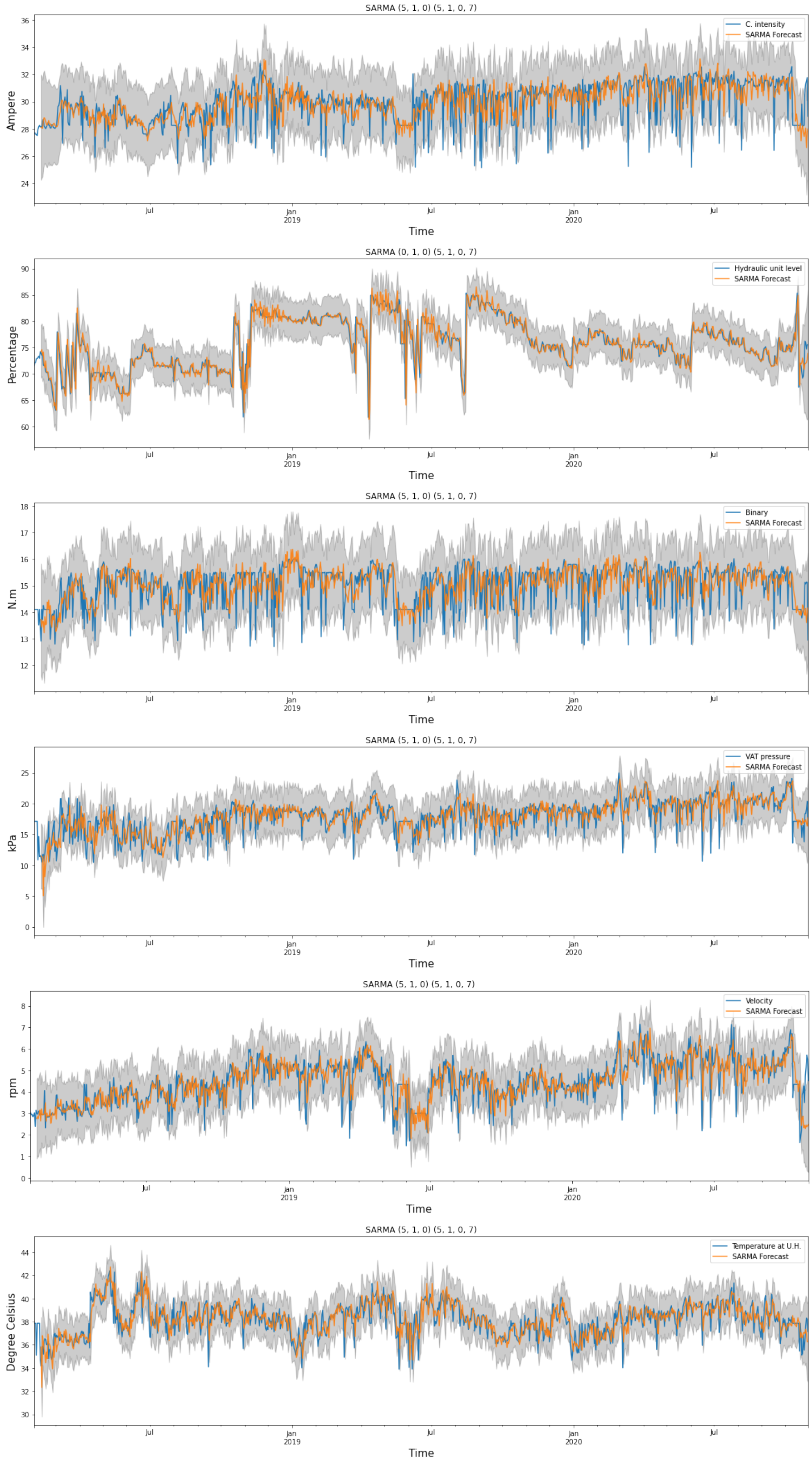


Figure 4.4: Prediction of variables using the SARIMA method, with 10 days lag and 30 days forward.

4.2 Test with Recorrent Neural Network LSTM model

To perform the prediction test, the deep network model was proposed, given that it involves a large amount of data to forecast 30 days ahead. The SARIMA model for a 10-day forecast had as result a stabilization, which is not intended.

The architecture for the model is encoder and decoder, as shown in Figure 4.5, with a hidden LSTM layer in the middle and a dense layer at the output. Prior processing of data is essential for training and test validation of a recurrent neural network and prevents undesirable phenomena during this process.

The prediction model has six variables in input and six variables in output. The goal is to predict the values of these variables with the highest confidence for additional predictive maintenance benefits. The models were implemented in Python using the TensorFlow library and Keras.

TensorFlow is an open-source library of comprehensive machine learning a variety of tasks. It allows the creation and training of neural networks to detect and decipher patterns and correlations, analogous to human learning [207]. Keras is an open source neural network library in Python language. Keras can acts as an interface for the TensorFlow library [208].

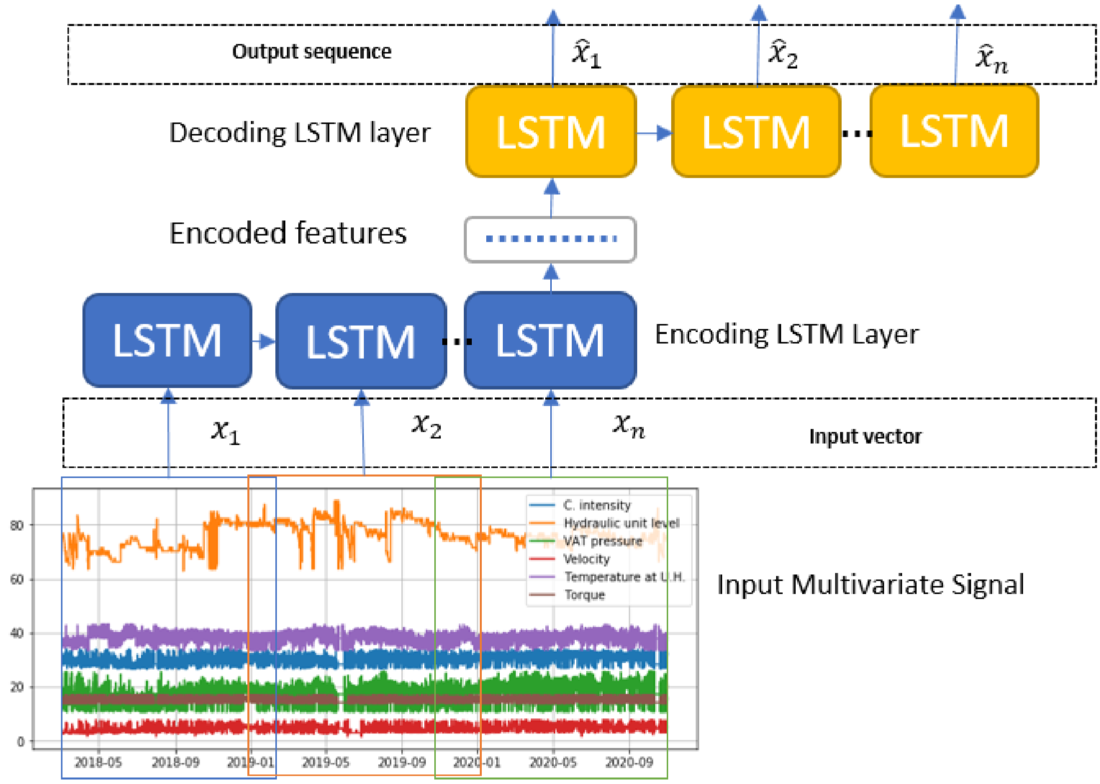


Figure 4.5: Model summary of one of the LSTM networks used. The model receives a window of n samples of each variable and predicts the value of those variables as predicted 30 days ahead.

To train and test the models, the dataset was divided into train and test subsets. Validation was performed using the test set, but those samples were not incorporated into the training set. The training set contained 85% of the samples and the test set the remaining 15% of samples. These values are adequate for convergence during learning. As an example, Figure 4.6 shows a learning curve for a model with 70 units in the middle layer and a window of 30 lag samples. The Figure shows that learning converges and takes less than 10 epochs.

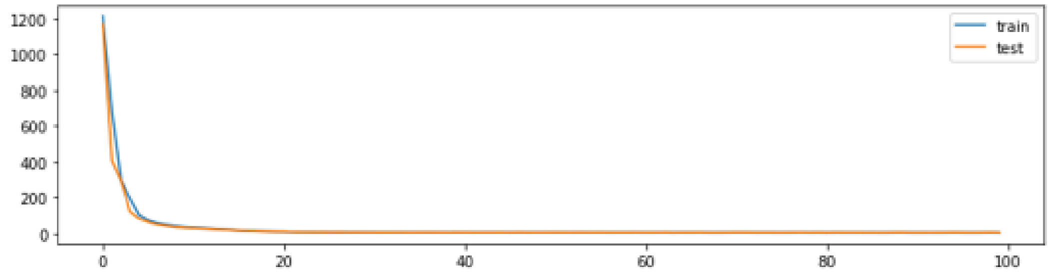


Figure 4.6: Example of learning curve, showing loss during training of an LSTM model.

Without the treatment of the data, the first test was made using the sampling rate per hour and it was verified the impossibility of performing the prediction due to the high number of oscillations the variables presented, as shown in Figure 4.7, where orange line is the train values and green line is a test values.

4.2.1 Test to Determine Historical Window Size and Number of LSTM Units Using One Sample per Day

The first experiments performed aimed to determine the best window size to use. The smaller the window, the smaller and faster the model that can be used. However, if the window is too small, it may be insufficient to make accurate predictions.

The original dataset had 1,445,760 data points, which is very large and would require a lot of memory and time to train and test. The experiments were performed after downsampling the data, so that there is only one sample per day. Each sample is the average of 1004 original samples. The downsampled dataset is, therefore, less than one thousand of the original dataset.

The results are measured in the test set. Figure 4.8 shows the MAPE and MAE measured for each variable. It also shows the global RMSE measured globally for the train and test sets.

As Figure 4.8 shows, models with windows of 40 and 50 samples allow better learning and produce smaller prediction errors.

Additional experiments to determine the best size for the number of cells in the hidden layer were performed.

For those experiments, a window of 40 historical samples was used, relying on the results of the previous experiments.

Figure 4.8 shows the results obtained for experiments with a window of 40 days and different numbers of hidden cells. As the results show, the model with the best performance is the one with 50 hidden cells.

After the results of the first experiments with one sample per day, additional experiments were conducted to determine if there was any considerable loss in down sampling from one sample per minute to one sample per day. A first experiment, which consisted of halving a down sampling period from 24 hour to just 12 hour, was performed. Therefore, the dataset doubled in size, since it contained two samples per day instead of just one.

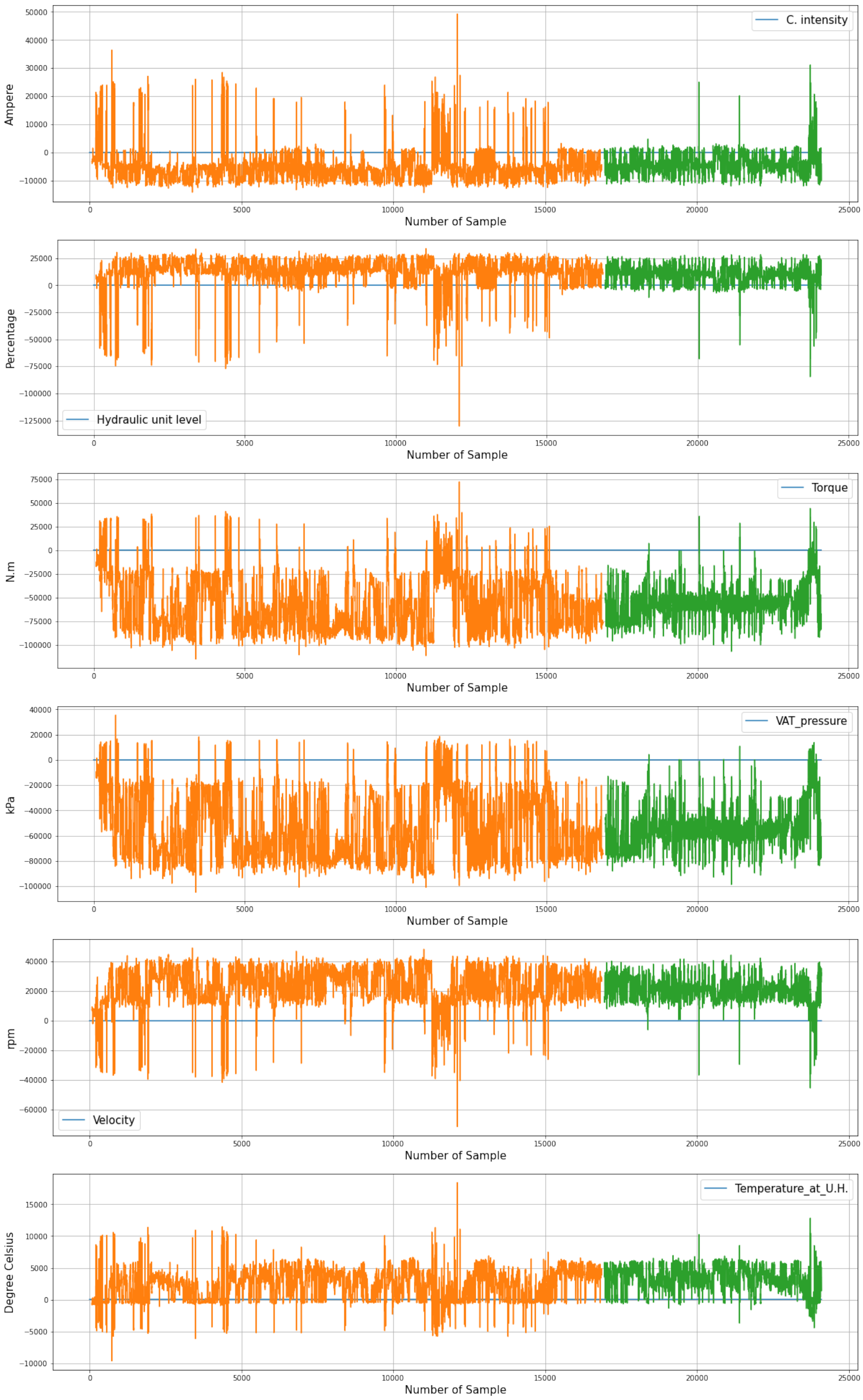


Figure 4.7: Prediction of the current intensity variable without data processing.

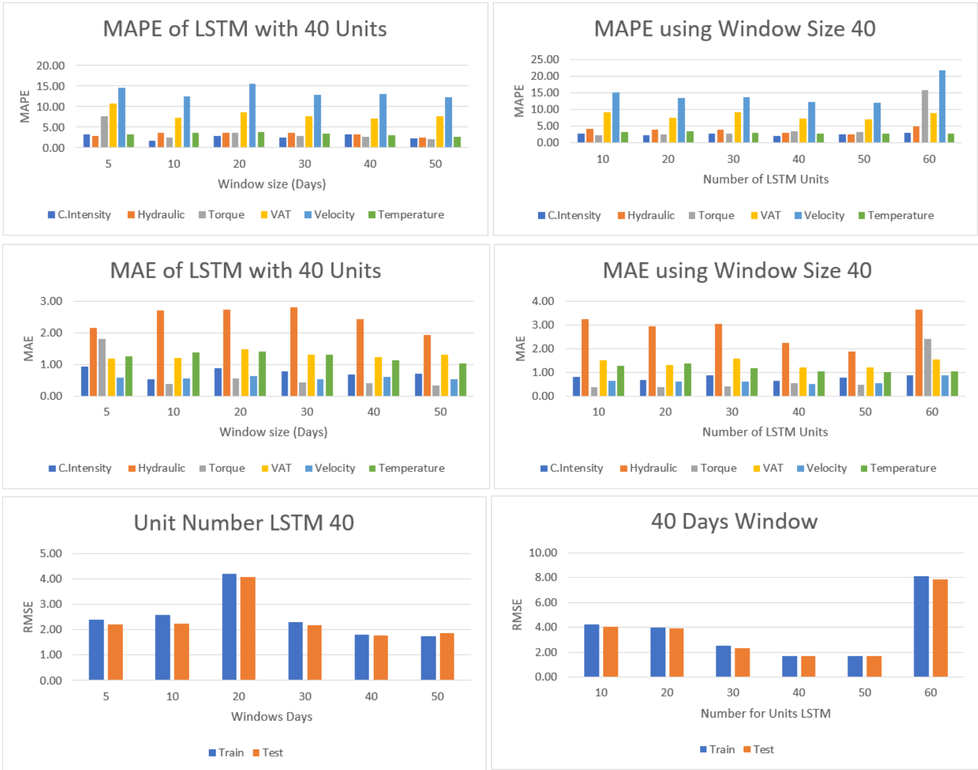


Figure 4.8: Results obtained with a different number of LSTM cells in the hidden layer, as well as different sliding window sizes, to predict values 30 days in advance with downsampling to one sample per day.

4.2.2 Test to determine Historical Window Size and Number of Unit LSTMs Using Two Samples per Day

According to the results shown in Figure 4.9, it concluded that a window of 10 days (20 samples) shows the best performance. This shows that the model can exhibit approximately the same performance with even fewer input samples when compared to the models above. The models used for those experiments had 20 cells in the hidden layer.

Once the impact of the window size was determined, additional experiments were performed to gain a better insight into the impact of using more or less cells in the hidden layer. Figure 4.9 shows results of using different numbers of cells.

Figure 4.10 shows a graphical representation of the results obtained using the model with 40 cells in the hidden layer and a 10-day sample window. As the graph shows, predictions usually follow the actual signal most of the time. However, there are still some areas where the actual signal differs from the predicted value by a small percentage, namely speed and temperature. Most of the differences may be due to behaviors that are difficult to capture due to the small dimensions of the dataset. As more data are collected, neural models may be able to capture more patterns and make more accurate predictions.

In addition to the graphs shown in Figure 4.10, in Tables 4.2, and Table 4.3, the magnitudes of the RMSE errors in the training set and test set are also presented. They were measured in the model validation dataset.

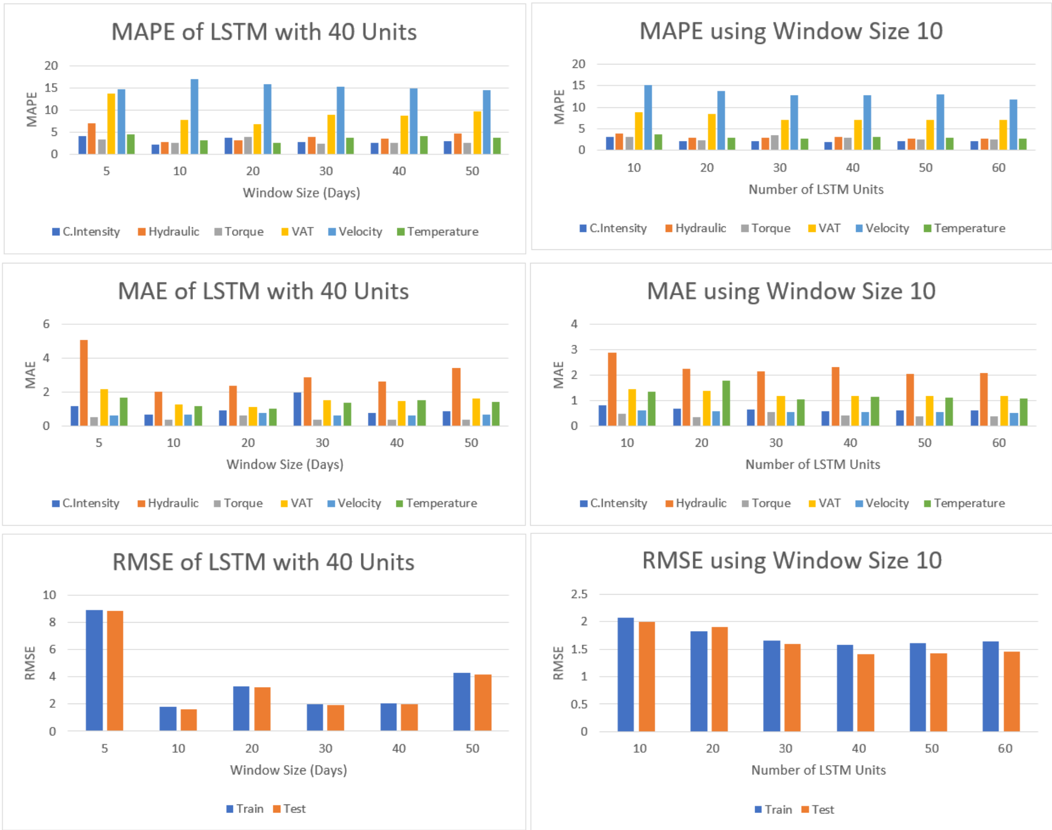


Figure 4.9: Results obtained with a different number of cells in the hidden layer, also using different window samples to predict values 30 days in advance with resampling for the two samples for a day.

After the test with the LSTM neural network, it was verified that the higher the sampling rate the more difficult is to learn the LSTM neural network. In summary, the model did not show ability to perform good predictions as is shown in Table 4.4.

Through the model learning curve it was validated the instability in the training and validation process as shown in Figure 4.11. In order to circumvent this situation it was proposed to perform the comparison between two models whose architecture is the same, but the neuronal units are different.

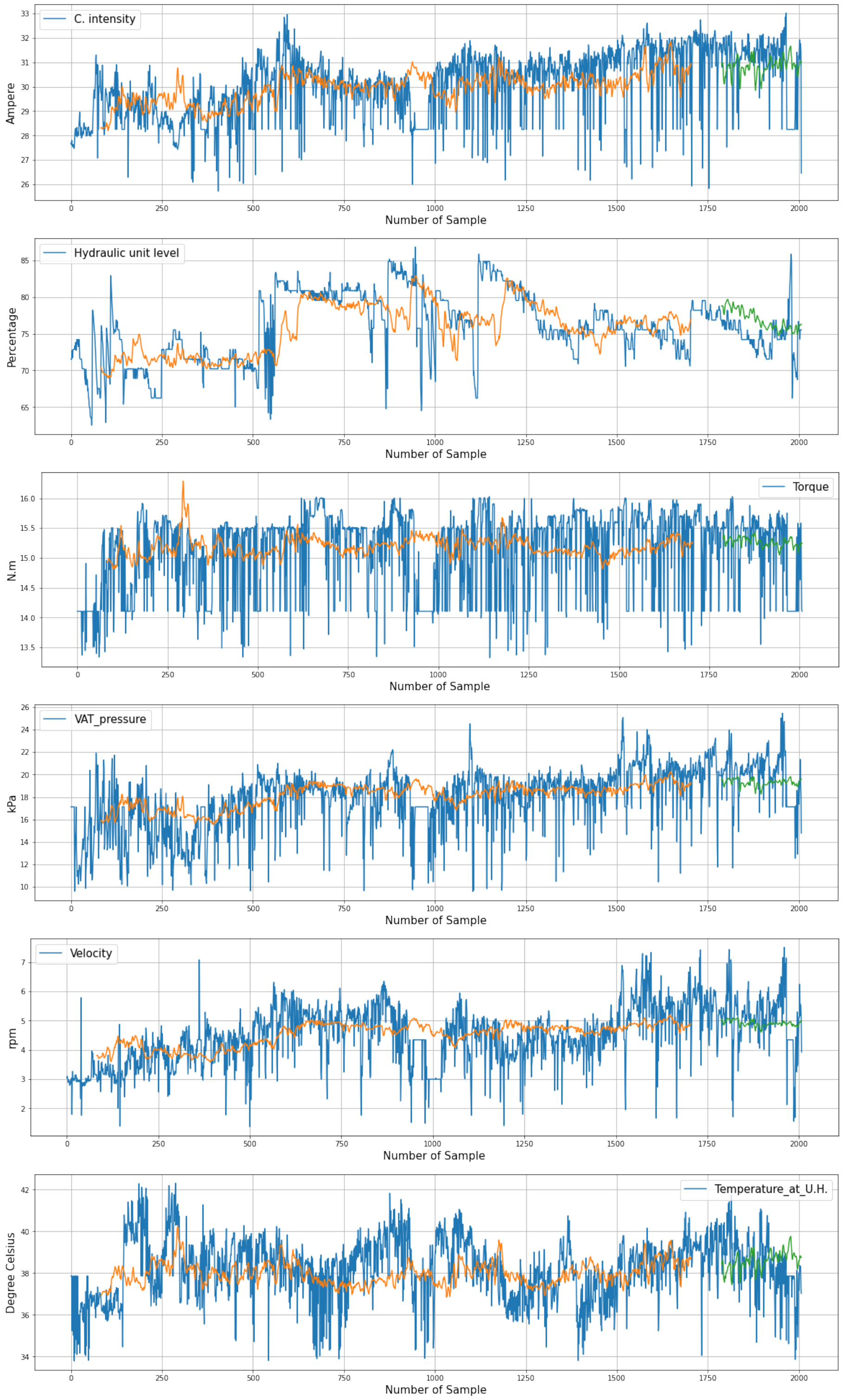


Figure 4.10: Variable forecast with a window of samples of 10 days, sampling rate two samples per day, and a network model with 40 units in the hidden layer. The blue lines show the actual value. The orange lines show the predictions during the training set and the gray lines show the predictions in the test set.

Table 4.2: The magnitude of RMSE errors in the test and training set, using one sample per day.

Window Size (Days)	Train	Test	Units	Train	Test
5	2.39	2.20	10	4.23	4.07
10	2.57	2.24	20	3.99	3.93
20	4.21	4.09	30	2.52	2.35
30	2.31	2.19	40	1.68	1.70
40	1.81	1.77	50	1.66	1.70
50	1.74	1.86	60	8.14	7.85

Table 4.3: The magnitude of RMSE errors in the test and training set, using two samples per day.

Window Size (Days)	Train	Test	Units	Train	Test
5	8.91	8.87	10	2.07	1.99
10	1.80	1.61	20	1.82	1.91
20	3.29	3.23	30	1.65	1.59
30	1.98	1.94	40	1.58	1.41
40	2.07	1.98	50	1.61	1.42
50	4.32	4.16	60	1.64	1.46

4.3 Test with Recorrent Neural Network GRU model

Deep learning networks are very sensitive to hyperparameters. If the hyperparameters are not set correctly, the prediction output will produce high frequency oscillations [209]. Important hyperparameters for GRU network models are the number of hidden units in the recurrent layers, the dropout value, and the learning rate value.

This test was performed to confirm the ability of the models to learn, and then to determine the optimal hyperparameters of the LSTM and GRU.

Individually, these hyperparameters can significantly influence the performance of the LSTM or GRU neural models. Studies, such as [210, 211], demonstrate how important the adjustment of hyperparameters is, as it optimizes the learning process and can present good results against more complex neural network structures.

To begin the test considered the same case to determine the best size for the sliding window, experiments were performed, resampling to just one sample per day, which gave a total of 1004 samples, 70% of which were used for train and 30% for test.

4.3.1 Testing the Convergence of the Learning Process

Figure 4.12 shows the learning curve for a GRU model with 40 units in the hidden layer and 12 samples in the window. The graph shows the losses measured in the train and test sets. Although the learning curve shows that the model learns very fast in less than 10 epochs, in the following experiments the number of epochs is limited to 15.

4.3.2 Model Performance with Different Window Sizes

The first experiments performed were aimed at finding the best windows for the LSTM model and the GRU model. Therefore, the dataset has a total of 1004 samples. The model used has 40 units

Table 4.4: MAE, MAPE using the LSTM neural model with the relu activation function.

Resampling (hour)	C. intensity	Hydraulic unit level	Torque	VAT pressure	Velocity	Temp. at U.H.
12	2.42	2.92	3.72	10.36	17.19	2.30
6	357.00	134.54	210.17	108.43	660.79	222.86
3	0.00	0.00	0.00	0.00	0.00	0.00
1	0.00	0.00	0.00	0.00	0.00	0.00
Resampling (hour)	**C. intensity**	**Hydraulic unit level**	**Torque**	**VAT pressure**	**Velocity**	**Temp. at U.H.**
12	0.71	2.22	0.55	1.57	0.64	0.88
6	105.86	102.49	31.66	18.19	25.16	84.91
3	0.00	0.00	0.00	0.00	0.00	0.00
1	0.00	0.00	0.00	0.00	0.00	0.00

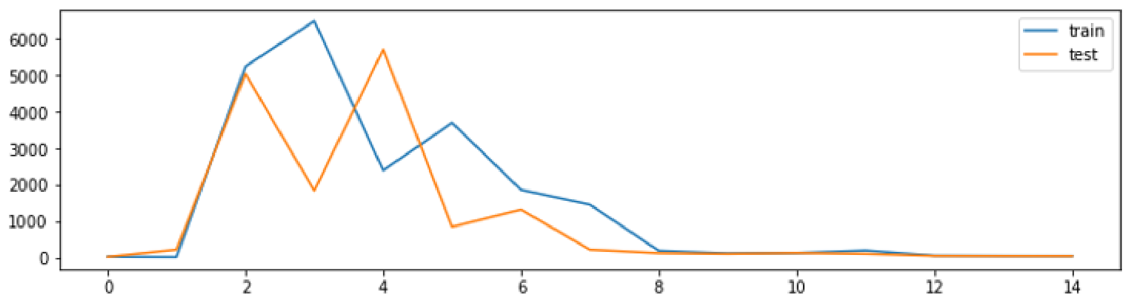


Figure 4.11: Example of learning curve, LSTM model.

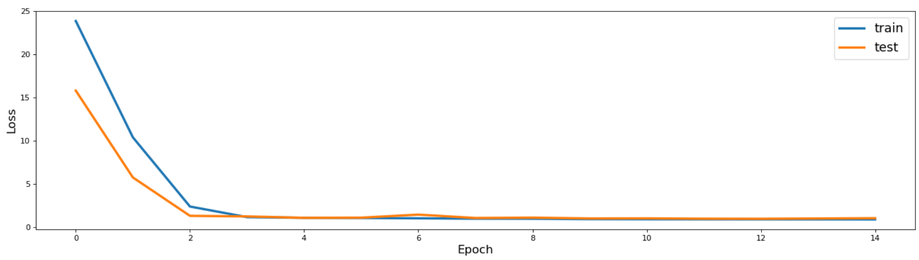


Figure 4.12: Learning curve of a GRU model.

in the hidden layer.

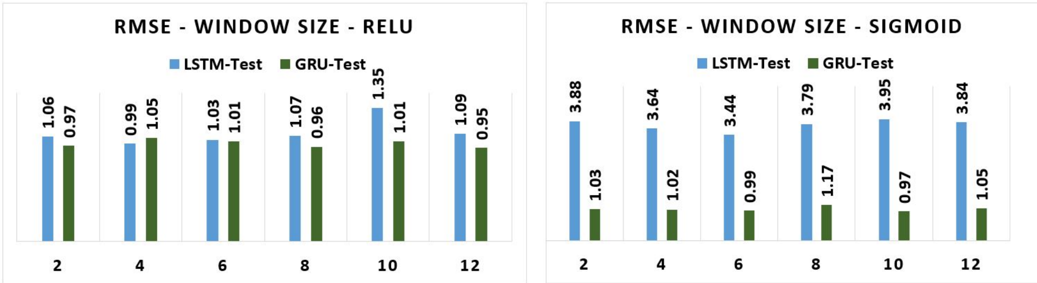
Figure 4.13 shows the results of two models with different window sizes and two different activation functions in the output layer. RMSE is the mean of all variables.

As shown, GRU always outperforms LSTM regardless of window size or activation function used. The window size has little effect on the performance of the model, with a small difference between 2 days and 12 days. On the other hand, the results are better when ReLU is used in the output layer.

The difference in performance between GRU and LSTM is larger when using the Sigmoid function than when using the ReLU function.

Figure 4.14 shows the MAPE and MAE associated with the 30 day forecast, for past windows of 2 to 12 days. The charts demonstrate that the LSTM architecture that uses a ReLU activation function in the output layer has lower errors. Using the sigmoid function, the LSTM errors are much larger.

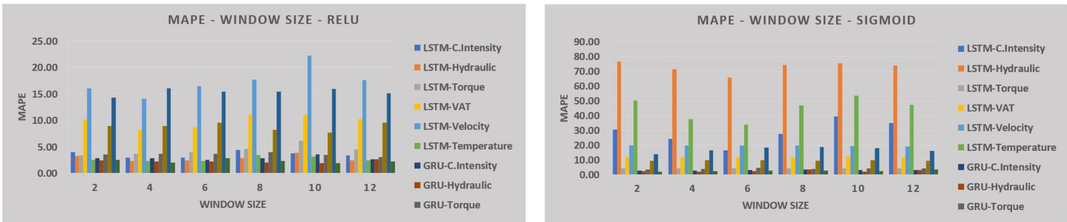
The GRU, however, in general performs better than the LSTM for all variables and activation



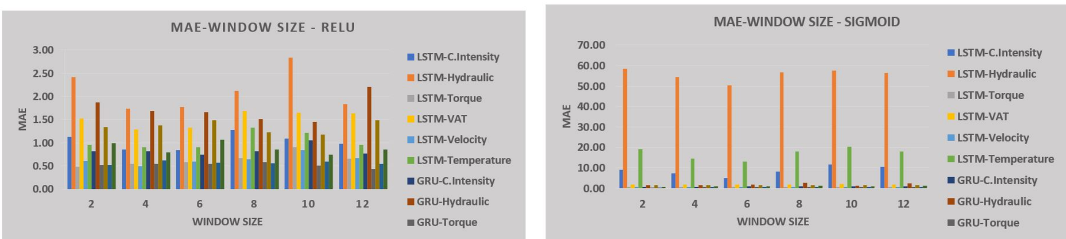
(a) Results with ReLU activation (b) Results with sigmoid activation

Figure 4.13: RMSE values for LSTM and GRU models, with different window sizes and activation functions for the output layer.

functions. The prediction error results are much more stable for the GRU than they are for the LSTM. Table 4.5 shows exact error values for the best window sizes for the LSTM model and GRU model.



(a) MAPE using ReLU activation (b) MAPE using sigmoid activation



(c) MAE using ReLU activation (d) MAE using sigmoid activation

Figure 4.14: MAPE and MAE errors, for each variable, using ReLU and sigmoid activation functions, for window sizes of 2, 4, 6, 8, 10, and 12 days, using one sample per day. Exact values are shown in Tables 4.5 for the best window sizes.

4.3.3 Experiments to Determine Model Performance with Different Resample Rates

In the present experiments, the dataset contains a large number of samples, so only undersampling techniques are necessary in order to reduce the number of data points. The same method was used to average a number of samples, depending on the size of the dataset desired.

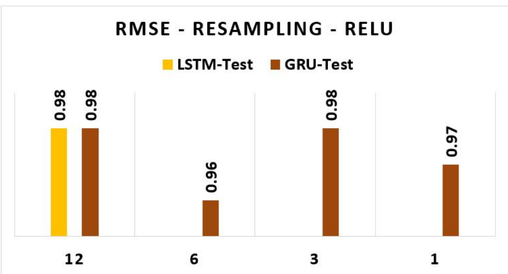
Experiments were performed undersampling to obtain one sample per 12 h (two per day), one per six hours (four samples per day), one per each three hours, and, finally one sample per hour. So, the dataset size was greatly reduced.

Table 4.5: Summary of the best prediction errors obtained with the LSTM models and GRU. Window is the historical window size in days. AF is the output activation function.

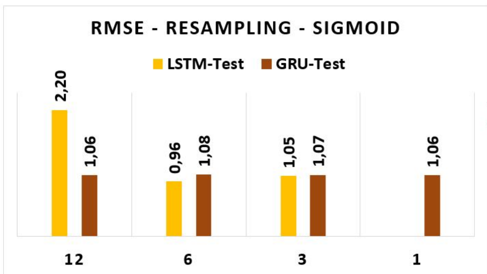
				MAPE			
	Window-AF	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
LSTM	4-ReLU	2.95	2.32	3.68	8.28	14.06	2.38
	6-Sigmoid	16.48	65.98	4.24	12.09	19.70	34.00
				MAE			
	Window-AF	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
	4-ReLU	0.86	1.74	0.54	1.29	0.50	0.91
	6-Sigmoid	4.91	50.34	0.61	1.83	0.72	13.02
				MAPE			
	Window-AF	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
GRU	12-ReLU	3.63	1.95	3.53	7.74	15.99	1.92
	10-Sigmoid	2.57	2.21	3.74	9.53	15.41	2.82
				MAE			
	Window-AF	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
	12-ReLU	0.77	2.20	0.44	1.49	0.55	0.86
	10-Sigmoid	0.93	1.32	0.61	1.42	0.64	0.91

The window sizes were the best of the previous experiments: a window size of 4 days for the LSTM and 12 days for the GRU, with the ReLU. A window size of 6 days for the LSTM and 10 days for the GRU, with the sigmoid.

Figure 4.15 shows the average RMSE errors for both models. As the results show, sometimes the LSTM overperformed the GRU, namely when using the sigmoid function with periods of six and three hours. However, the difference was not statistically significant. On the other hand, the GRU was able to learn in all the situations and the RMSE error was always approximately 1. So, the GRU is robust and accepts larger periods with minimal impact on the performance, while the LSTM model is much more unstable.



(a) Using ReLU function at the output layer



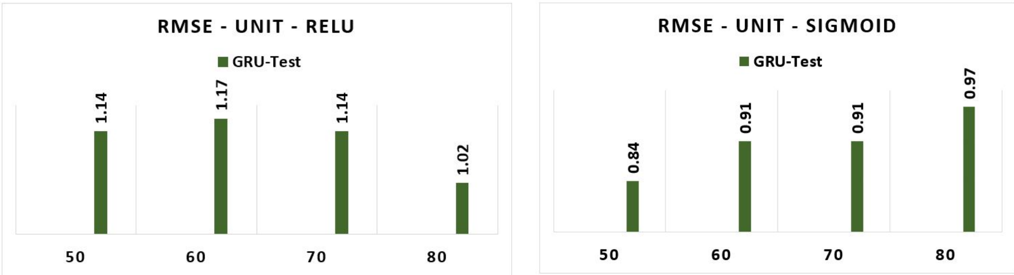
(b) Using sigmoid at the output layer

Figure 4.15: RMSE value for LSTM and GRU model with ReLU and sigmoid at the output layer, for different undersampling rates: using one data point per 12 h, one per six hours, one per 3 h, and one per hour.

4.3.4 Experiments with Different Layer Sizes

An additional experiment was performed to compare the performance of the models with different numbers of units in the hidden layer.

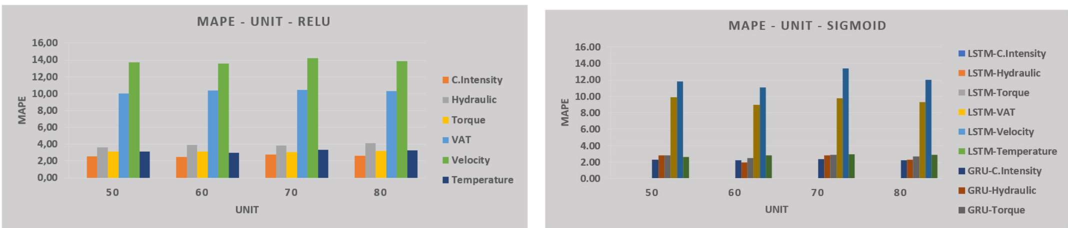
Using the GRU model, it is possible to learn with a larger number of samples, and with different variations of the model units, as shown in Figures 4.16 and 4.17. The LSTM was unable to learn with the resampling rate period of 1 h; therefore, results are missing. The window used in the experiments was 10 days for the sigmoid and 12 days for the ReLU, which were the optimal windows for the GRU using the ReLU and sigmoid functions, respectively.



(a) Using ReLU function at the output layer

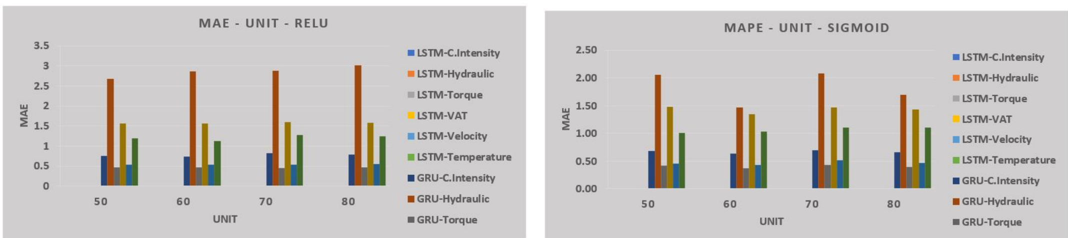
(b) Using Sigmoid at the output layer

Figure 4.16: RMSE errors measured, with different numbers of cells in the hidden layer.



(a) MAPE using ReLU function

(b) MAPE using sigmoid function



(c) MAE using ReLU function

(d) MAE using sigmoid function

Figure 4.17: MAPE and MAE obtained with different numbers of units in the hidden layer, measured when predicting future values 30 days in advance, with a resampling period of one hour. The LSTM was not able to learn, so the results are just for the GRU.

As the charts show, the GRU, using the sigmoid activation function, achieves the lowest RMSE error with 50 units in the hidden layer. Experiments described in Section 4.3.3 showed that the GRU with the same parameters, with 40 units in the hidden layer, had an RMSE error of 1.06. Table 4.6 shows the best results for the GRU model, after the tests with different numbers of cells in the hidden layer.

Table 4.6: Summary of the best results obtained with different numbers of units in the hidden layer.

			MAPE			
Unit	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
80-ReLU	2.66	4.09	3.19	10.31	13.83	3.29
50-Sigmoid	2.30	2.80	2.85	9.87	11.80	2.66
			MAE			
Unit	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
80-ReLU	0.78	3.02	0.47	1.58	0.55	1.25
50-Sigmoid	0.68	2.05	0.42	1.48	0.46	1.01

4.3.5 Comparing Many-to-Many and Many-to-One Architectures

An additional experiment was performed, in order to determine if the models are better trained to predict all the variables at the same time (one model, six outputs, many-to-many variables) or trained to predict just one variable (six models with six variables in input, with only one in output, i.e., many-to-one variable).

This experiment was just performed for the GRU, which presented the best results in the previous experiments.

According to the graphs presented in Figure 4.18, it is clear that architecture 'many-to-many' presents slightly better results. Therefore, there is no advantage in training one model to predict each variable.

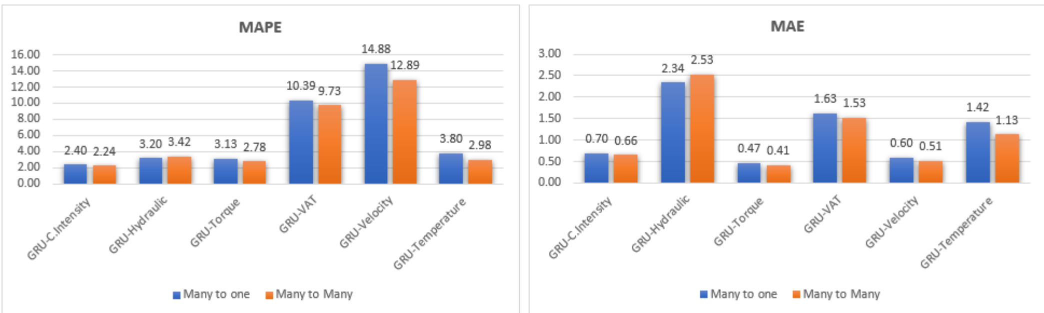


Figure 4.18: Comparison of the performance of the GRU models, trained to predict many-to-many and many-to-one variables.

4.3.6 Tests with Different Activation Functions in the Hidden Layer

An additional step was to test combinations of different activation functions, for the hidden and output layers of the GRU. The activation functions tested were sigmoid, hyperbolic tangent (tanh), and ReLU. Figure 4.19 shows a chart with the average RMSE of the models. Globally, ReLU in the hidden layer and tanh for the output are the best models, even though ReLU–sigmoid and ReLU–ReLU are closely behind.

Table 4.7 shows the RMSE error for the different combinations of activation functions, for each variable. As the table shows, different variables may benefit from different functions, although, in general, a first layer of ReLU and a second layer of ReLU, sigmoid, or tanh are good choices.

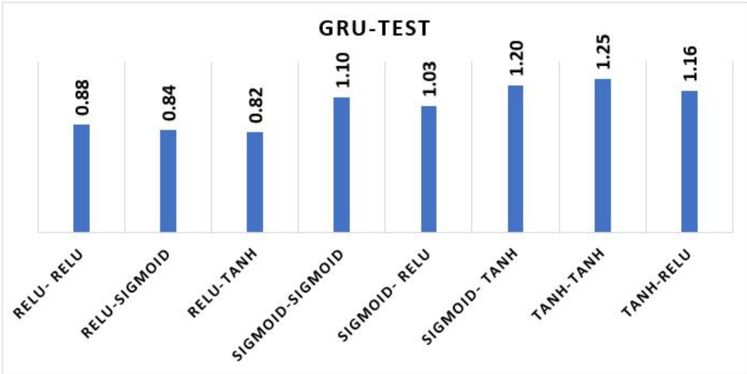


Figure 4.19: Average RMSE values, different types of activation functions.

The values shown in Table 4.7 are calculated for the raw output predicted. However, the raw output values have some sharp variations, which are undesirable for a predictive system. Therefore, the values were filtered and smoothed using a median filter. Figure 4.20 shows plots of selected results, where the signals and predictions were filtered with a rolling median filter, with a rolling window of 48 h. Table 4.8 shows the MSE errors calculated after smoothing. As the table shows, after smoothing, the prediction errors decrease.

Table 4.7: Average RMSE obtained for the six variables, with different activation functions, calculated after the values were smoothed with a median filter.

	RMSE					
Function	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
ReLU–ReLU	0.96	3.48	0.42	1.90	0.84	1.90
ReLU–Sigmoid	0.93	1.72	0.53	1.60	0.79	1.19
ReLU–Tanh	0.83	2.47	0.48	1.70	0.76	1.25
Sigmoid–Sigmoid	0.98	6.40	0.45	2.14	0.89	1.35
Sigmoid–ReLU	1.22	4.87	0.43	1.86	0.74	1.31
Sigmoid–Tanh	1.19	7.38	0.45	2.03	0.78	1.35
Tanh–Tanh	1.36	7.84	0.44	2.24	0.91	1.35
Tanh–ReLU	0.86	7.3	0.42	1.91	0.76	1.41

Table 4.8: Average RMSE obtained for the six variables after the average clean method, with different activation functions using the GRU model.

	RMSE					
Function	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
ReLU–ReLU	0.71	3.33	0.28	1.36	0.66	0.80
ReLU–Sigmoid	0.61	1.58	0.39	1.08	0.61	0.78
ReLU–Tanh	0.54	2.33	0.35	1.13	0.54	0.82
Sigmoid–Sigmoid	0.73	6.36	0.30	1.70	0.68	0.94
Sigmoid–ReLU	1.03	4.80	0.28	1.32	0.50	0.89
Sigmoid–Tanh	0.98	7.35	0.29	1.53	0.54	0.94
Tanh–Tanh	1.18	7.81	0.29	1.80	0.70	0.96

Figure 4.20 shows examples of plots of different prediction lines in part of the test set. As the results show, in some cases the ReLU–tanh combination is the best, while in other cases, the ReLU–sigmoid offers better performance. The ReLU–tanh combination is better, in general, but in the

case of temperature, the sigmoid output shows the best performance.

4.4 Improving data pre-processing for GRU model

For model validation the data were divided into two subsets. The training subset uses the first 80% of the total data and the test subset contains the remainder 20% of the data samples.

The purpose of the experiments is to find the best data preprocessing methods, neural model architectures and hyperparameters that produce the best results predicting future behaviour of the paper pulp presses. The tests were performed using a GRU neural network with data encoder and decoder architecture, which was the architecture that showed best results in previous work [212].

The experiments aim at testing different pre-processing methods. Elimination of discrepant values is Method 1. Data smoothing using the LOWESS filter is Method 2. The combination of both, first the elimination of discrepant data, then smoothing, is called Method (1, 2). The architecture of the neural network was the same for all the experiments, and it is the same that showed best results in previous work. Nonetheless, experiments were still performed with a smaller and faster GRU, with just 50 units, and a larger and slower network, with 500 units.

For press number 2, LOWESS method presented better results using a window of 5 days. The window size was halved because the number of data samples available from press 2 was too small for using larger windows. The dataset for press 4 contains 34800 hours of data, while the dataset for press 2 contains just 24096 hours of data.

Figure 4.21 shows the RMSE values of predictions for press 2, with the smaller and the larger GRU neural networks, with and without LOWESS filtering. As the figure shows, the prediction errors are much smaller when data are filtered. The difference is even more notorious in the larger network. For the same press and the same architecture, increasing the GRU units of the neural network to 500, it is verified that the combination of the methods leads to the same result, but with much smaller errors. The hydraulic variable in particular shows a larger error for both network structures.

For data originary from press number 4, the LOWESS filter presented better results using a window of 36 days. From the RMSE diagram in Figure 4.22, it can be seen that the results for press 4 also show much lower errors when the LOWESS filter is applied. The smaller model, with 50 GRU neural units, shows errors slightly larger than the larger model. For the same press using 500 GRU neural units, the RMSE errors are smaller, as demonstrated by the smaller area of the chart polygons.

Applying the two methods to press 2 data, it can be seen that while the errors in Table 4.9 are small, the important information are omitted from the graph in Figure 4.23, which is not good for possible press failure analysis.

Figure 4.24 shows the result of predicting the model with the better method of data processing for the press 4, which in this case falls on the intersection of the two methods. From the Table 4.10 it can be seen that the error is smaller.

Table 4.9: Prediction error results for 30 days advance forecast, using the two data preprocessing methods, removal of discrepant data and smoothing, for the 500 unit GRU and LOWESS with 5 days window, for press 2.

Prediction errors for press 2						
	C. Intensity	Hydraulic	Torque	VAT	Velocity	Temperature
MAPE	0.62	1.85	2.24	3.91	10.27	0.96
MAE	0.2	1.39	0.35	0.82	0.57	0.38
RMSE	0.23	1.55	0.37	0.95	0.6	0.5

Table 4.10: Prediction error results for 30 days advance forecast, using the two data preprocessing methods, removal of discrepant data and smoothing, for the 500 unit GRU and LOWESS with 36 days window, for press 4.

Prediction errors for press 4						
	C. Intensity	Hydraulic	Torque	VAT	Velocity	Temperature
MAPE	1.2	1.12	2.32	1.6	2.77	1.36
MAE	0.27	0.8	0.18	0.51	0.26	0.5
RMSE	0.30	1.00	0.20	0.61	0.30	0.69

4.5 Conclusion

In this chapter, the prediction results using classical models and recurrent neural networks of variables were presented. Although the predictions of the classical models are good, they do not allow long-term prediction or prediction by multivariate data.

Recurrent neural networks showed a good ability to forecast 30 days ahead and demonstrated the ability to forecast with multivariate data.

The chapter also showed the importance of preprocessing the data at the input of the recurrent neural network and reinforced the conclusion of the importance of adjusting the hyperparameters to realise a prediction with good accuracy.

The scientific results of this chapter led to four articles published in journals indexed in Scopus, where they already have a number of relevant citations, as this is one of the first studies to use this type of recurrent neural network architecture to predict pulp press failure (Appendix B, C, D and E).

Figure 4.20: Plot of the predictions with different combinations of activation functions.

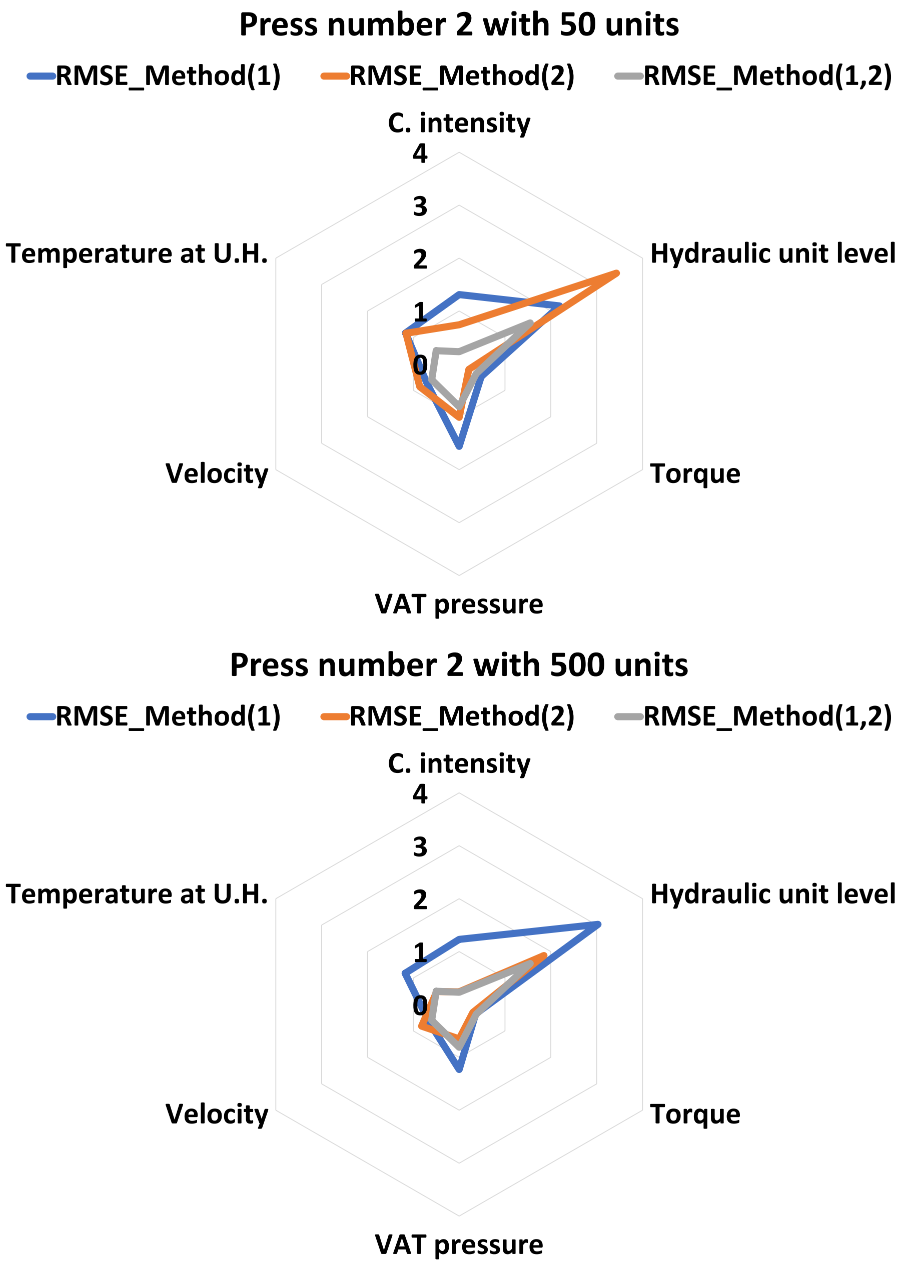
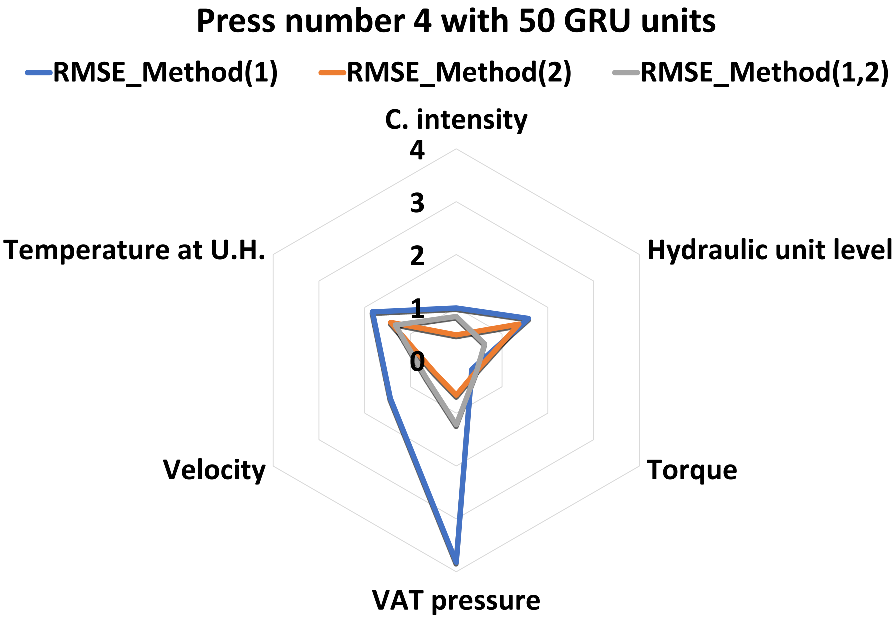


Figure 4.21: RMSE of the best models for press 2, using the two different methods for pre-processing data, for the smaller and larger GRU networks. Method 1 only removes discrepant data. Method 2 smoothes the data using a LOWESS filter. Method (1,2) is the application of both.



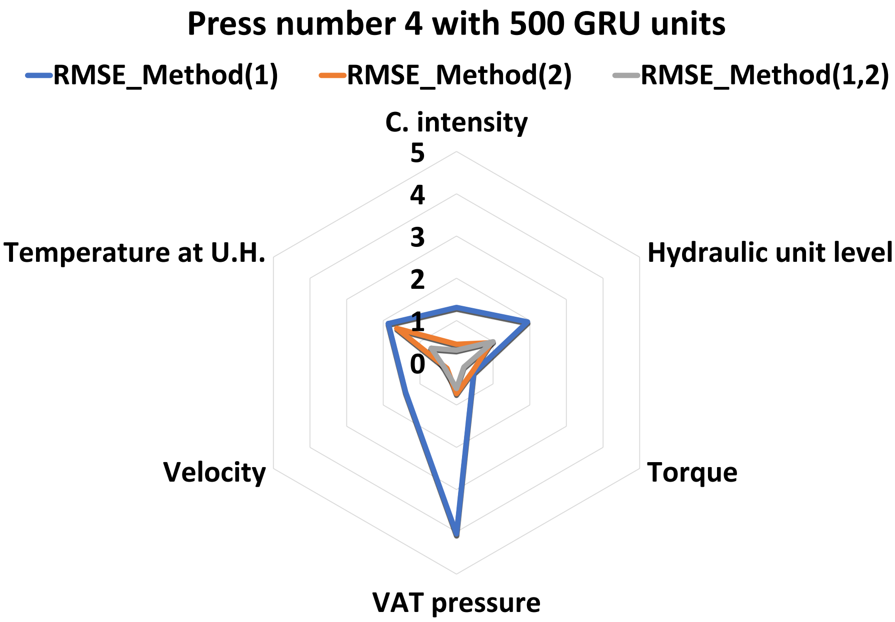


Figure 4.22: RMSE for predictions of press 4 using the different data pre-processing methods. LOWESS filtering and 500 GRU units result in smaller RMSE errors.

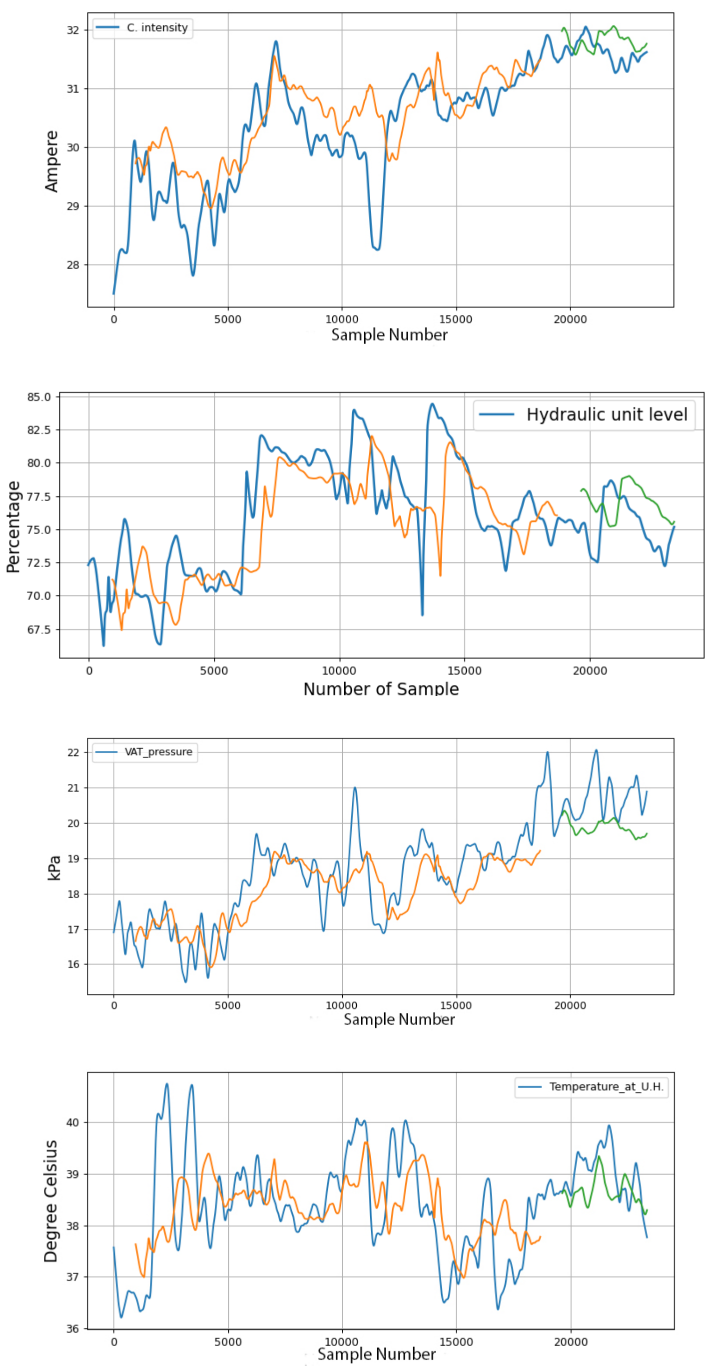


Figure 4.23: Signals and forecast results for press 2, with 30 day advance, using the two data processing methods, both removal of discrepant data and data smoothing using LOWESS filtering with 5 days window. The blue lines represent the actual value. The orange and green lines are predictions, respectively, in the train and test subsets.

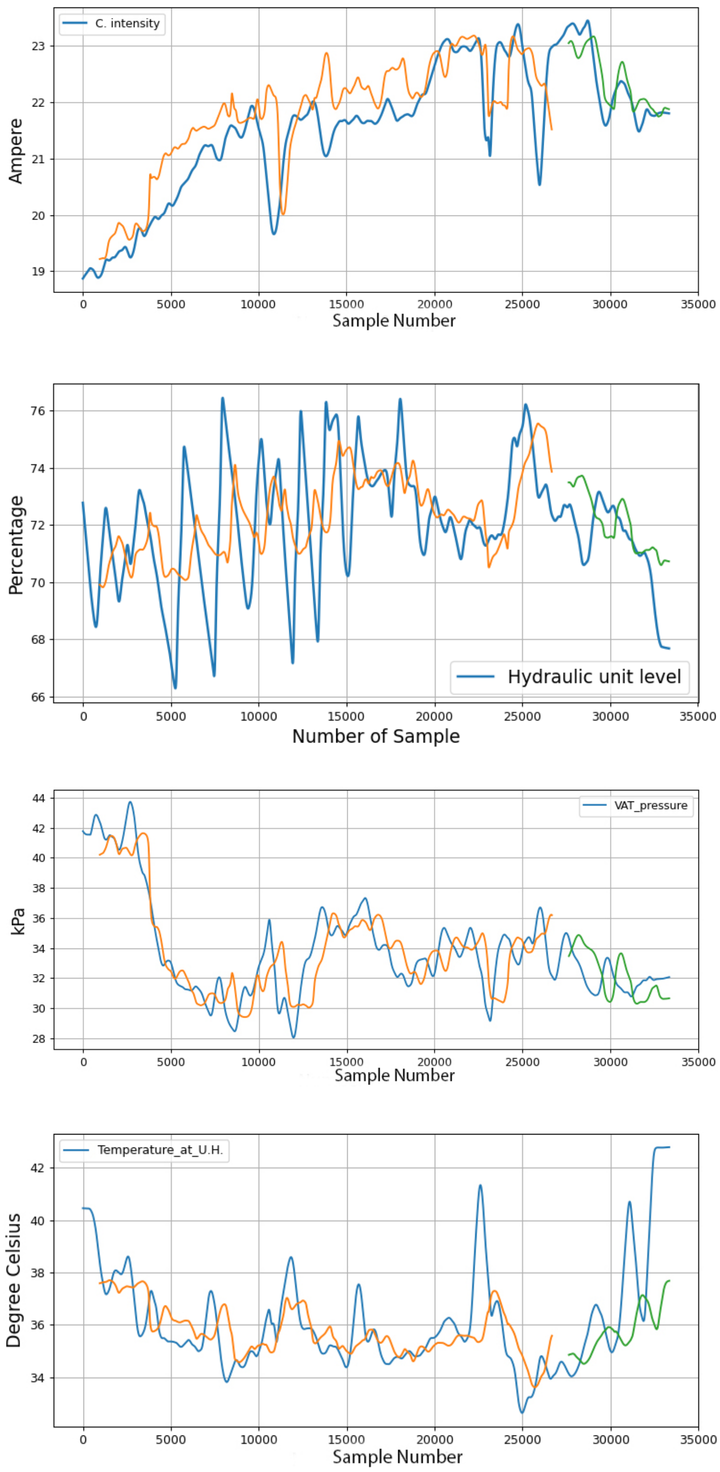


Figure 4.24: Signals and forecast results for press 4, with 30 day advance, using the two data processing methods, both removal of discrepant data and data smoothing using LOWESS filtering with 36 days window. The blue lines represent the actual value. The orange and green lines are predictions, respectively, in the train and test subsets.

Chapter 5

Tests and Results for Algorithm in Production

5.1 Prediction indexed to future's stock market

5.1.1 Use the Dataset for Steel Production

For the study of the forecast of the steel production variable we considered the variables of the future stock market and population growth. Figure 5.1 presents the description of the variables in which we can verify that they present a temporal behavior since the sampling rate is in a period of one year.

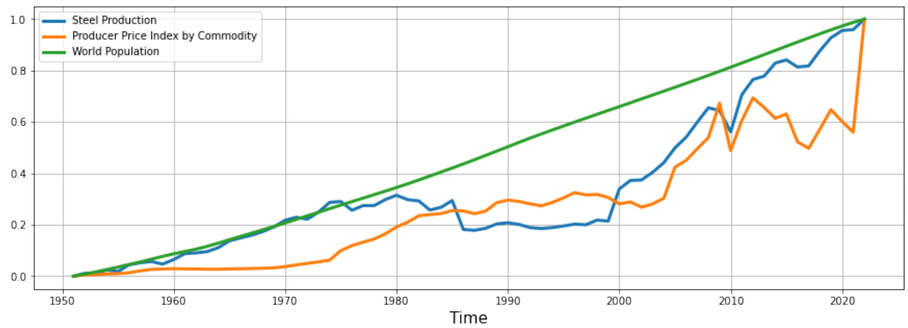


Figure 5.1: Annual crude steel production in the world, in millions of metric ton.

In Figure 5.1 it is possible to verify how the behavior of the variables has an upward tangential growth and it is also verified that when the price index per commodity lowers had an increase in the production of steel, but some points come to land the same behavior, already for the variable of population growth it is possible to verify that there is a constant growth.

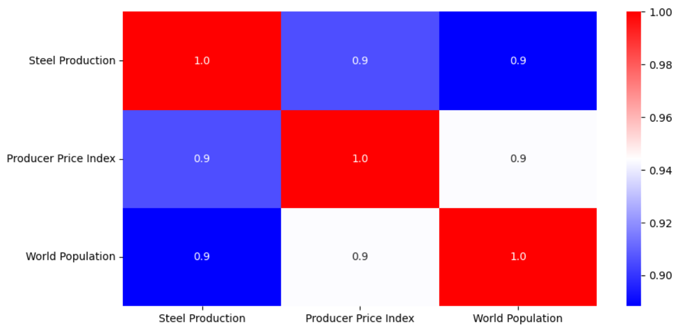


Figure 5.2: Correlation among variables.

Performing the correlation study between the variables it is possible to verify that there is a correlation of great relevance among the three variables and these correlations reach a value of 0.9 being 1 the highest value as shown in Figure 5.2.

It was also made an autocorrelation study starting from the steel production variable whose correlation falls when it reaches a lag of 10, in the same way this behavior is verified for other 2 as shown in Figure 5.3.

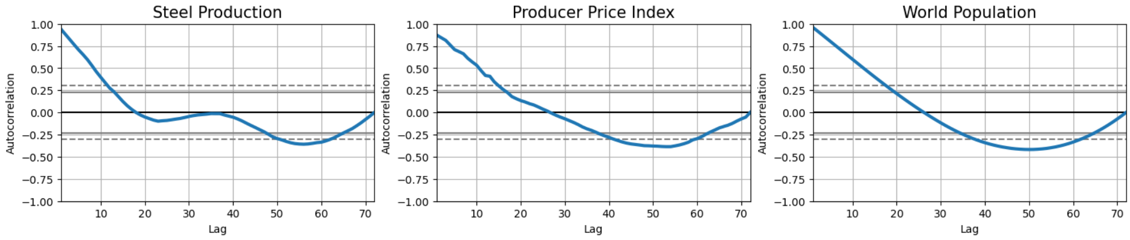


Figure 5.3: Autocorrection of variables.

Having characterised the variables with the same forecasting model, an early prediction of 1 year was performed as the sampling rate for both variables is 1 year. The first test consists of adding the model input one variable at a time until the last one, having the steel production variable fixed and the model output with the feathered one variable forecast result for steel production. In Table 5.1, and Table 5.2 is present the errors of the total steel production forecast. Figure 5.4 and the Figure 5.5 shows the result of the best model against the input.

For test 1, when a variable is incremented to the input we notice that the error increases with the exception of the Steel Prodution and World Population variable that decreases.

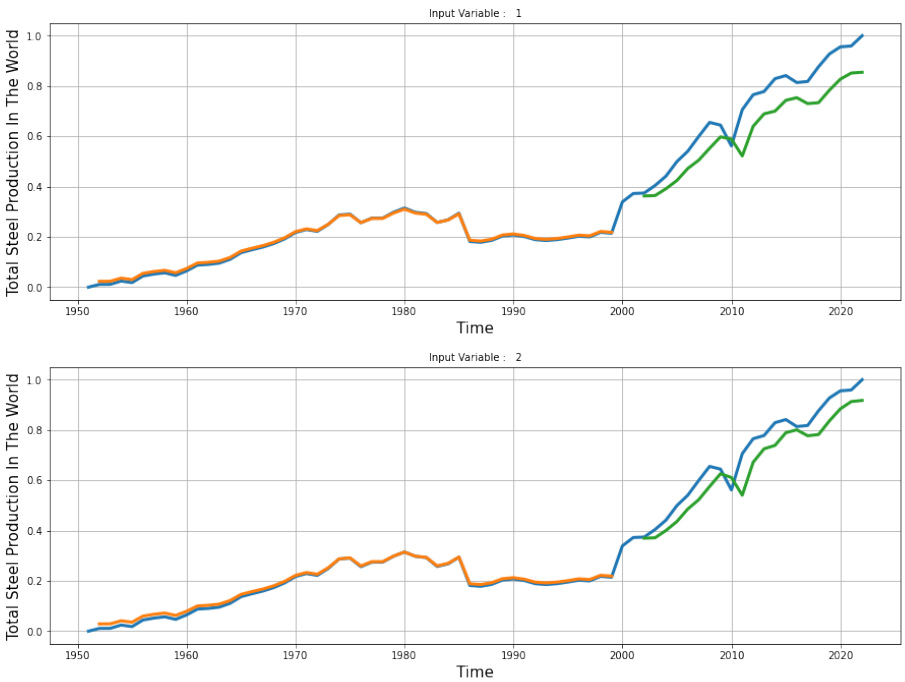


Figure 5.4: Best prediction of Test 1 result for the steel production variable with two neural net inputs.

As can be seen from Figure 5.6, the variables have a good correlation. In this case of the derivatives the correlations are weak, the derivative that has a reasonable correlation in relation to steel production is the world population.

Table 5.1: Test 1 error of the forecast Steel Production with Sampling rate per Year.

	RMSE	MAPE	MAE	R^2
Steel Production	0.07	8.13	0.061	0.87
Producer Price Index	0.03	3.76	0.03	0.97
World Population	0.16	19.98	0.15	0.23
Steel Production Der	0.13	14.38	0.11	0.56
Producer Price Index Der	0.13	16.32	0.12	0.52
World Population Der	0.17	18.56	0.14	0.32

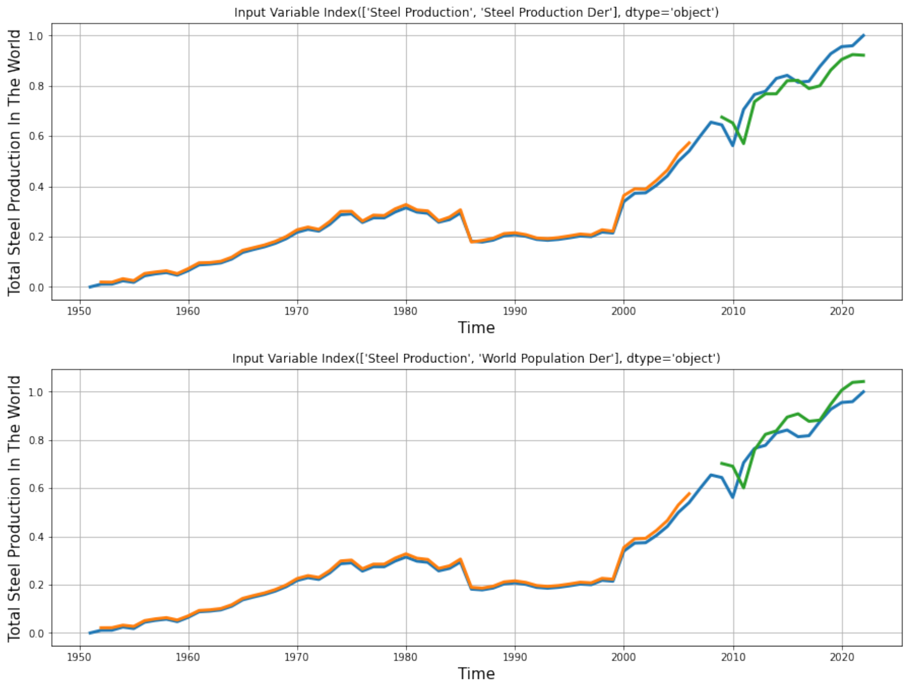


Figure 5.5: Best prediction of Test 2 result for the steel production variable with two neural net inputs.

Table 5.2: Test 2 error of the forecast Steel Production with Sampling rate per Year.

	RMSE	MAPE	MAE	R2
Steel Production	0.07	8.83	0.07	0.60
Producer Price Index	0.12	14.48	0.12	-0.10
World Population	0.08	9.89	0.08	0.50
Steel Production Der	0.02	2.27	0.02	0.97
Producer Price Index Der	0.08	9.32	0.08	0.55
World Population Der	0.06	7.86	0.06	0.69

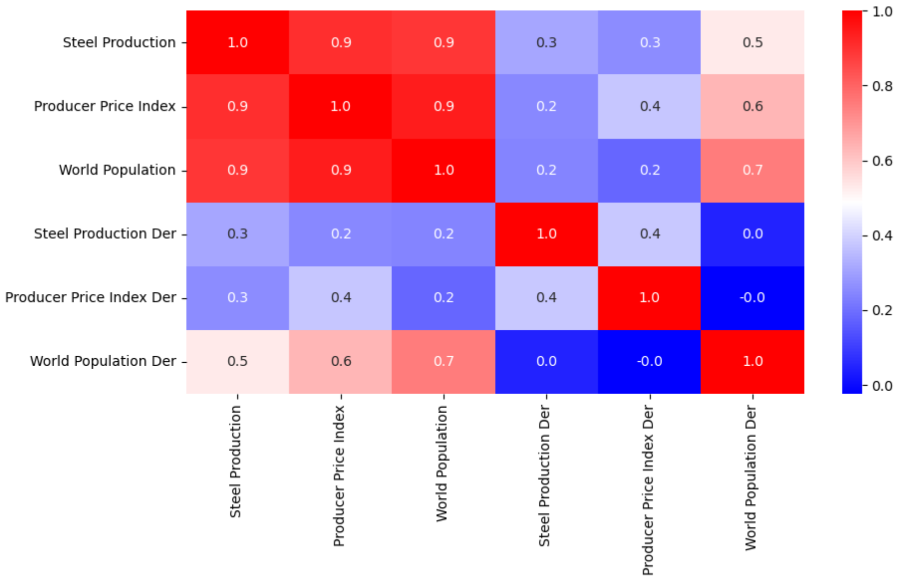


Figure 5.6: Correlation between variables and derivatives.

Using the same forecasting model and methodology with the same defined parameters, the forecast was carried out again, but with the particularity of having as input the derivatives of other variables for the steel production forecast.

5.1.2 Market Futures

Market Futures are used to trade an underlying asset at a future date at a set price, protecting buyers from changes in Asset prices.

The Commodities are the raw inputs used in the production of goods and of Market Futures, that can be bought and sold on specialized exchanges as Financial Assets - these Financial Assets are produced by Physical Assets. The Commodities are highly speculative and are especially sensitive to economic changes.

Like all kinds of Assets, the Commodity prices are determined by Supply and Demand: Supply and Demand can be impacted in many ways, such as economic shocks, natural disasters, and investor enthusiasm. The Commodities can be: Hard – which are usually classified as those that are mined or extracted from the earth (for example: metals, gold and petroleum); Soft - such as agricultural products (for example: wheat, cotton, coffee, sugar and soybeans).

The importance of Market Futures, having into account the preceding, relates directly to the production to be predicted by the plants managers in the future, year after year.

When the plants managers know what to produce in the future, they know the Availability of the Physical Assets they need to fulfil those objectives. If the plants managers Know the Physical Asset's Availability, they can plan the best Maintenance Policy to reach the Key Performance Indicators (KPI) most adequate to perform the production objectives aiming to reach the Market Futures forecast.

The preceding is one of the biggest value added of this research, which permits to relate the Future Market with the shop floor planning, correlating Production and Maintenance Policy planning.

5.1.3 Production Forecast Indexed To Futures Market Values

After the test with the steel production data similar tests were carried out with the pulp production data. Since the external variables (PSIALL, Wood Pulp, ALTRI SGPS, Consumer_Prices) have different sampling rates, it was decided to standardize the sampling rates by month, once some samples were in that sampling rate. With the monthly sampling rate, the data were normalised, because they had different magnitudes, having the results shown in Figures 5.7.

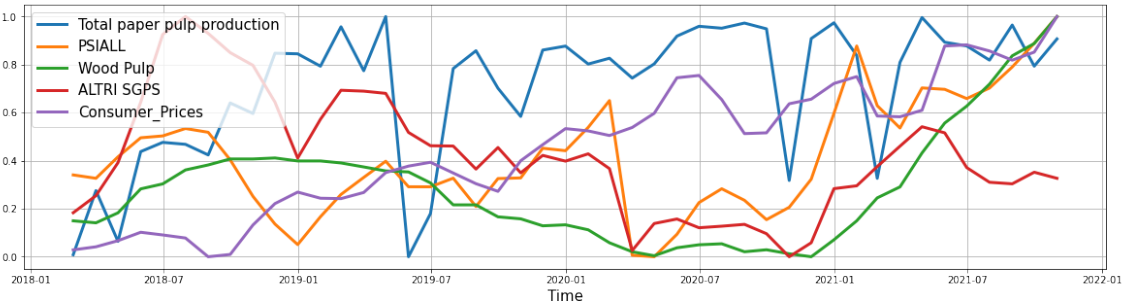


Figure 5.7: Total Paper Pulp Production with Sampling Range per Month.

In Figure 5.8 we can see the respective histogram that show the frequencies in the data, where it is possible to verify that there are some discrepancies between some data, namely the total production data and the company data.

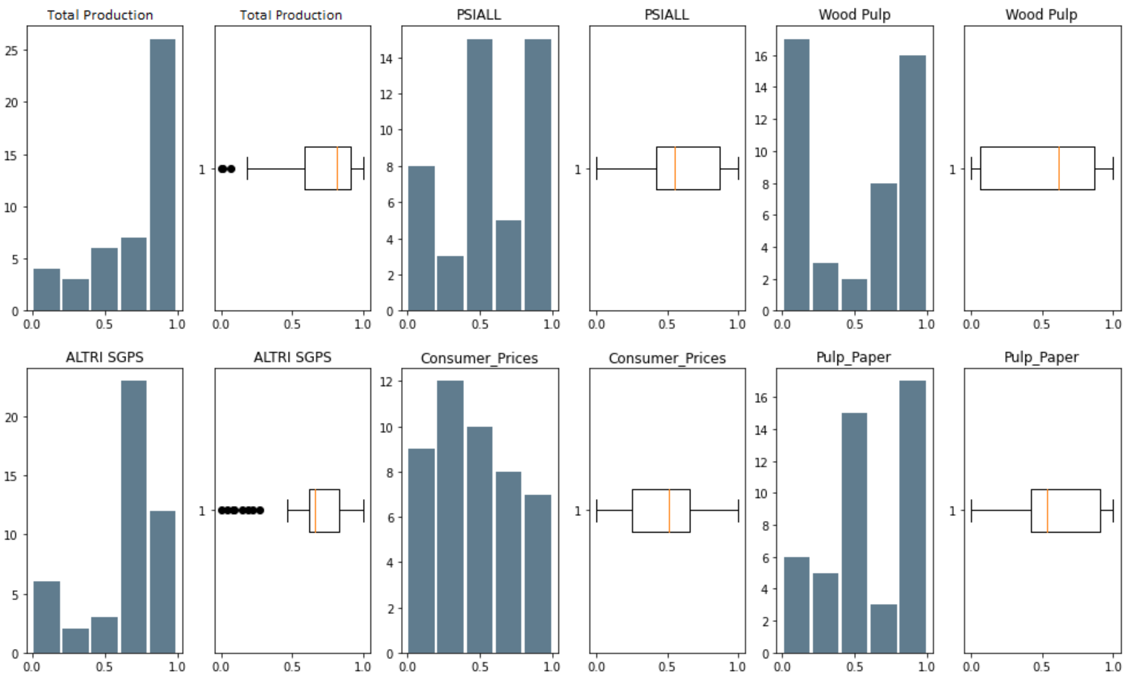


Figure 5.8: Hestogram and boxplot of futures market variables.

Table 5.3 shows the data description, averages of the standard deviation variable, the minimum, maximum, first, second and third quantiles. Being the variables normalized we will have its minimum as zero and its maximum at one.

Table 5.3: Statistical parameters of the variables for the stock Exchange.

	Count	**Mean**	**std**	**min**	Q_1**—25%**	Q_2**—50%**	Q_3**—75%**	**max**
	count	mean	std	min	0.25	0.50	0.75	max
Total Production	46	0.71	0.28	0.00	0.59	0.81	0.91	1.00
PSIALL	46	0.56	0.30	0.00	0.42	0.55	0.87	1.00
Wood Pulp	46	0.50	0.39	0.00	0.07	0.61	0.87	1.00
ALTRI SGPS	46	0.63	0.27	0.00	0.62	0.66	0.83	1.00
Consumer_Prices	46	0.46	0.29	0.00	0.25	0.51	0.65	1.00
Pulp_Paper	46	0.58	0.30	0.00	0.42	0.54	0.91	1.00

Table 5.4: Test 1 error of the forecast Total Paper Pulp Production with Sampling Range per Month.

	RMSE	**MAPE**	**MAE**	**R2**
Total Production	0.18	28.38	0.16	0.33
ALTRI SGPS(volume)	0.21	30.86	0.19	0.16
ALTRI SGPS	0.17	26.13	0.15	0.4
PSI ALL-SHARE	0.24	34.35	0.22	-0.11
Consumer	0.21	31.68	0.17	0.1
PCU32213221	0.18	27.21	0.13	0.33
Total Production Der	0.23	34.83	0.2	-0.09
ALTRI SGPS(volume) Der	0.24	35.22	0.18	-0.11
ALTRI SGPS Der	0.24	35.89	0.22	-0.17
PSI ALL-SHARE Der	0.23	33.78	0.18	-0.07
Consumer Dear	0.28	40.43	0.21	-0.55

It was verified that, there is a correlation between the variable the paper pulp production with the variables of the futures market as shown in Figure 5.9. The production of pulp paper have a correlation with the 3 variables Company, Consumer Prices and Pulp Paper.

With the same architecture used in the previous studies texts were carried out with a main objective to add information in the prediction model by increasing the input variables of the neural network in order to improve the prediction of the total paper pulp production from 30 days forward.

The first test, which consists of adding a variable in each training and testing process, was performed using the same architecture. It was found that the results improved when applied to an encoder and decoder architecture with 100 Gru units with a training process of 200 epochs, with the combination of activation functions in the first 'tangent' layer and in the second 'relu' layer.

Table 5.4 shows the results of the prediction of the variable combinations, in Figure 5.10 show the best prediction the paper pulp total production.

Using the same architecture, we moved on to the second test, which has net input a pair of variables that corresponds a total pulp production as a fixed variable. Table 5.5 shows the prediction errors, which do not prove that the prediction has good accuracy. Figure 5.11 shows the pair of variables that gave the best results.

Figure 5.12 shows the result of the technique of increasing the sample size from one sample size per month to one sample size per day.

In Figure 5.13 it can be seen that the correlation between the variables and derivatives remains low in relation to the total paper production variable.

The first test added a variable in each training and testing process to predict the value of all pulp production using the same architecture. The results were better with the encoder and decoder

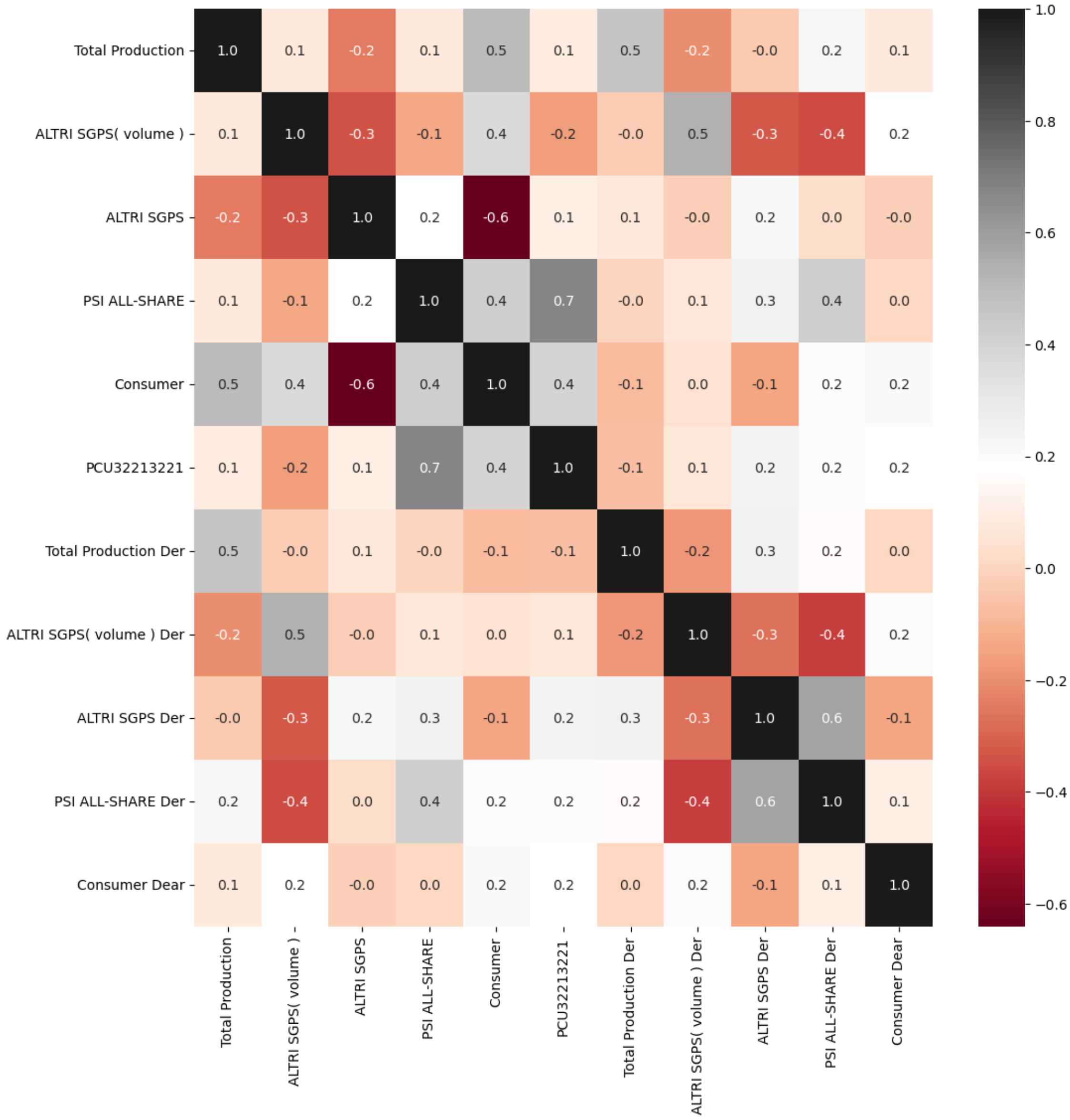


Figure 5.9: Correlation between stock market variables and total production.

Table 5.5: Test 2 error of the forecast Total Paper Pulp Production with Sampling rate per Month.

	RMSE	MAPE	MAE	R2
Total Production	0.17	26.55	0.15	0.42
ALTRI SGPS(volume)	0.20	30.28	0.18	0.19
ALTRI SGPS	0.22	32.05	0.15	0.03
PSI ALL-SHARE	0.27	39.14	0.26	-0.50
Consumer	0.24	36.25	0.17	-0.13
PCU32213221	0.21	30.67	0.19	0.15
Total Production Der	0.23	35.92	0.19	-0.06
ALTRI SGPS(volume) Der	0.18	27.50	0.16	0.38
ALTRI SGPS Der	0.20	30.74	0.18	0.18
PSI ALL-SHARE Der	0.20	30.34	0.18	0.21
Consumer Dear	0.17	26.46	0.15	0.40

Figure 5.10: Best Model of Test 1 in Forecasting Total Pulp Production with Sampling Range per Month.

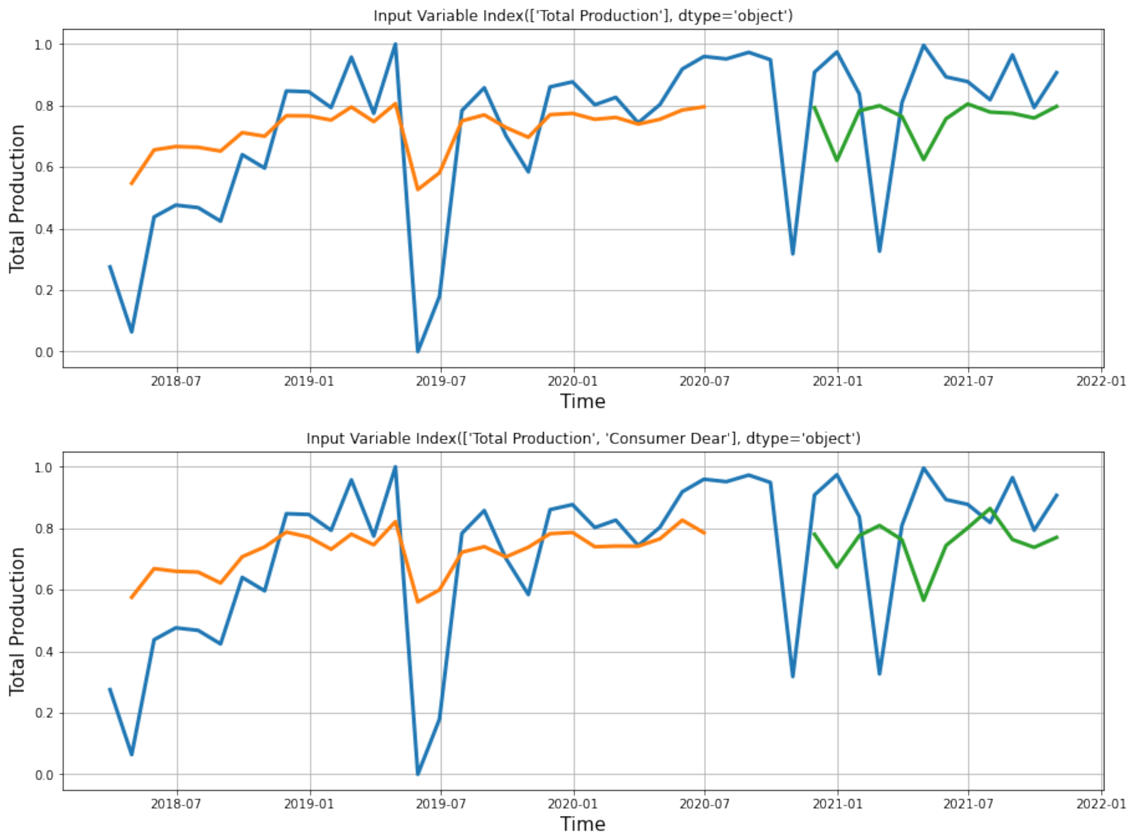


Figure 5.11: Best Model of Test 1 in Forecasting Total Pulp Production with Sampling Rate per Day.

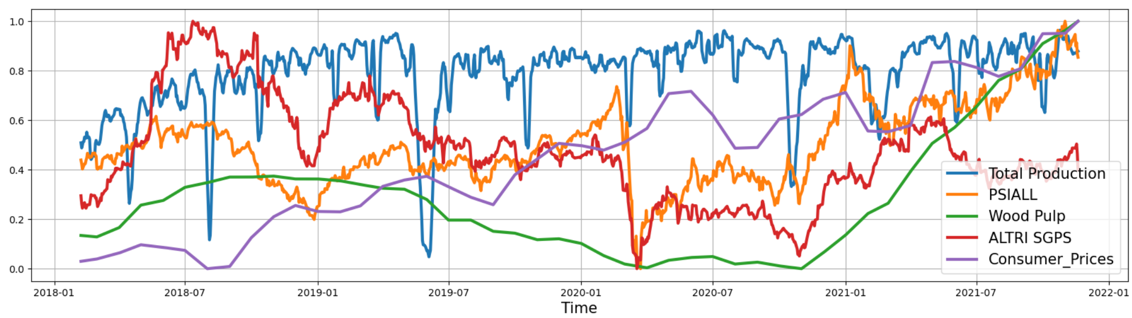


Figure 5.12: Total Paper Pulp Production with Sampling Range per Day.

architecture with 100 units of GRU, with the combination of the activation function in the first layer TanH (Hyperbolic Tangent) and in the second layer ReLU (Rectified Linear Unit).

Using the same architecture, the second test was conducted, which consisted of the net input being two variables and the output is the prediction of the total pulp production variable. Table 5.6, and Table 5.7 shows the prediction errors as a function of the variables entered into the network. Figure 5.14 shows the best prediction result of test 2.

Using the same architecture, we moved to the second test, which has net input a pair of variables that have the total pulp production as a fixed variable. Figure 5.15 shows the pair of variables that gave the best results.

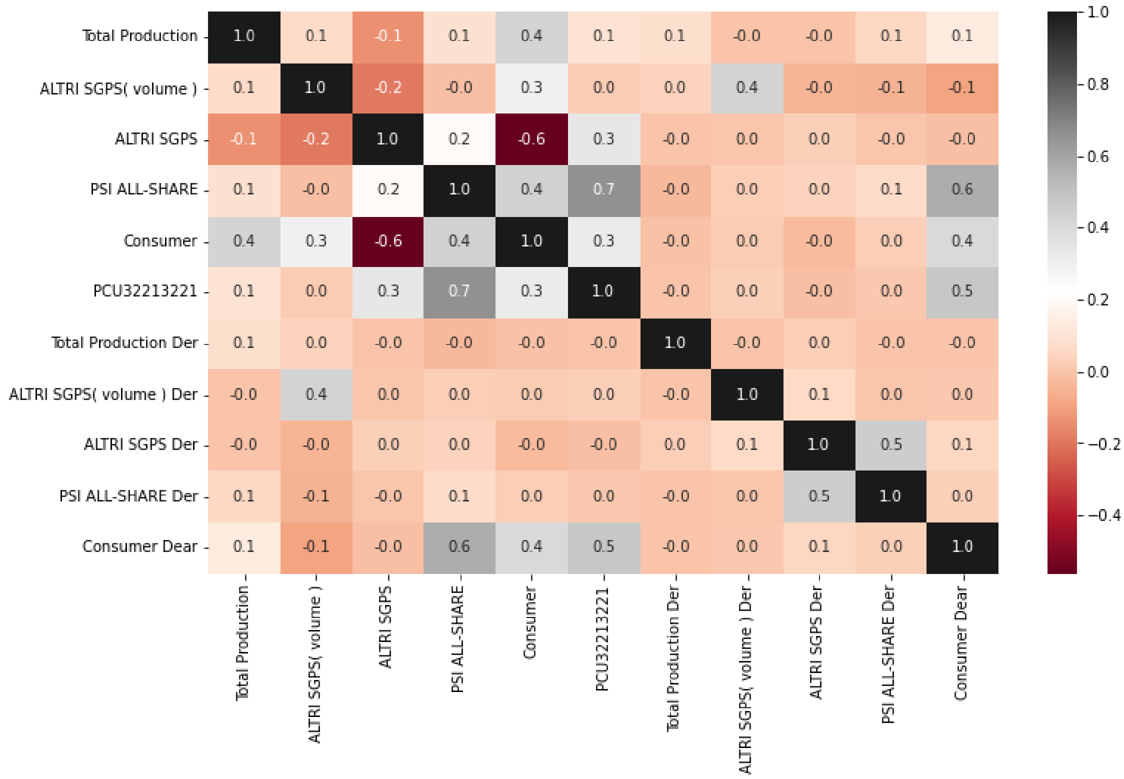


Figure 5.13: Correlation of all variables with the Sampling rate per Day.

Table 5.6: Test 1 error of the forecast Total Paper Pulp Production with Sampling rate per Day.

	RMSE	MAPE	MAE	R2
Total Production	0.06	6.84	0.05	0.75
ALTRI SGPS (volume)	0.05	6.16	0.04	0.79
ALTRI SGPS	0.06	6.76	0.05	0.75
PSI ALL-SHARE	0.06	7.09	0.05	0.73
Consumer	0.13	14.91	0.11	-0.12
PCU32213221	0.09	9.99	0.07	0.41
Total Production Der	0.12	13.94	0.10	-0.04
ALTRI SGPS (volume) Der	0.12	12.58	0.09	0.05
ALTRI SGPS Der	0.12	13.07	0.10	0.03
PSI ALL-SHARE Der	0.16	18.02	0.14	-0.75
Consumer Dear	0.10	10.84	0.08	0.34

5.2 Conclusion

Forecasting represents a major importance in adjusting production management, as informed decisions can be made to minimise risks based on the correlation of variables.

This chapter shown the importance of the stock exchanges in adjusting the forecasting model; for the steel production data, it was found that the forecast was adjusted with the increase of the stock exchange variables; this importance is confirmed in the correlation curve between the variables.

For paper pulp production, although some variables show significant correlation, there is another factor that did not allow the expected results, namely the low sampling rate per day of paper pulp production.

Figure 5.14: Best Model of Test 1 in Forecasting Total Pulp Production with Sampling Rate per Day.

Table 5.7: Test 2 error of the forecast Total Paper Pulp Production with Sampling rate per Day.

	RMSE	**MAPE**	**MAE**	**R2**
Total Production	0.05	5.71	0.04	0.82
ALTRI SGPS(volume)	0.05	5.91	0.04	0.80
ALTRI SGPS	0.06	6.16	0.05	0.79
PSI ALL-SHARE	0.06	6.20	0.05	0.80
Consumer	0.12	13.52	0.10	0.08
PCU32213221	0.06	7.01	0.05	0.72
Total Production Der	0.05	5.08	0.04	0.86
ALTRI SGPS(volume) Der	0.05	5.55	0.04	0.82
ALTRI SGPS Der	0.06	6.55	0.05	0.77
PSI ALL-SHARE Der	0.05	5.43	0.04	0.83
Consumer Dear	0.05	5.23	0.04	0.84

This chapter led to the participation in the National Congress of Asset Management in Coimbra (Congrega 2022), publication in a Scopus indexed journal and winning third place in the Innovation Prize for Young Engineers - PIJE 2021 (Appendix F).

Figure 5.15: Best Model of Test 2 in Forecasting Total Pulp Production with Sampling Rate per Day.

Chapter 6

Discussion

6.1 Summary of main results

Predicting the future behavior of industrial assets is a long-awaited goal, as it enables predictive maintenance to take the right action at the right time. Therefore, the application of time series and other AI models to predict device health is a new and growing field.

This thesis compiles a set of results concerning the state of the art and the results of the proposed predictive models.

The model comparison becomes important since to select the best model that fits the data it is necessary to test them, and if they are well parameterized this effort can present advantages in the optimization of computational resources.

Among all the models proposed in this work, the GRU recurrent neural network presented the best result compared to the others. The study shows that with more information in the GRU neural network, it is possible to obtain good prediction results. The sampling rate plays an important role in reaching a good prediction result.

6.2 Comparison of these results with the state of the art

For short-term forecasting, the models are satisfactory, emphasizing the need to clean the discrepant data. According to new studies in this area, they show superiority in the growth of the use of Neural Networks for those objectives, namely Recurrent Neural Networks that have greater long-term and short-term forecasting efficiency due to their Long-Short Term Memory capacity [213, 214, 215].

Based on the research proposed in the prior art, the usefulness of deep networks for time series variable forecasting can be verified. The field of prediction using deep neural networks has grown rapidly due to the development of new models and the development of computing power.

The stochastic models AR, ARMA and SARIMA showed good results although for either high sampling rate or low sampling rate in a 30-day forecast, the forecast became stable which is not intended.

Using these three models (AR, ARMA and SARIMA), it can be verified using evaluation data that both provide acceptable prediction errors. The ARIMA model outperformed the AR model, and SARIMA model.

There are no hyperparameters to optimize for regression model performance. However, in order to find the best model, it is necessary to evaluate multiple models and use the AIC information criterion to choose the parameters that best fit the data. Based on data showing moderate variation, it can be concluded that these models have good predictive.

The results show that it is possible to optimize neural models LSTM to forecast future values 30 days in advance. The model with LSTM unit experimented uses as input a vector consisting of concatenation of a number of samples of all variables. The output is a vector with the predictions of all samples too. The performance of the models is generally better for some variables and worse for others.

In the literature review, no other studies were found to deal with forecast for industrial paper pulp presses using encoder-decoder architectures and recurrent neural units.

LSTM and GRU models are two of the best forecast models. They have gained popularity recently, even though most of the state-of-the-art models are more traditional architectures. The GRU network is simpler than the LSTM, supports higher resampling rates, and it can work on smaller and larger datasets.

The experiments performed showed that the best results are based on the GRU neural network: it is easier and faster to train and achieve good results. A GRU network, with encoding and decoding layers, is able to forecast future behavior of an industrial paper press, 30 days in advance, with MAPE in general less than 10%.

An optimized GRU model offers better results with a 12-day sampling sliding window, with a sampling period of 1 h, and 50 units in the hidden layer.

The best activation functions depend on the model. However, the ReLU–tanh is perhaps one of the best models, on average. The results also demonstrate that training the models using just one output variable, thus optimizing a model for each variable separately, is not advantageous when compared to training one model to predict all six variables at the same time.

Data processing removing discrepant data simplifies the learning process of the RNN model and also leads to an improvement in the prediction results. The results obtained showed an improvement with data from both presses when discrepant data samples were replaced by the average. An analysis of autocorrelations shows that the use of data processing methods results in higher correlations for larger periods of time, when compared to untreated data.

Using data smoothed with the LOWESS filter, the learning process is highly facilitated. The prediction errors obtained in a 30 days advance forecast are smaller, with MAPE in general less than 10 %.

Compared to first results the LSTM predict model, the MAPE for the Current Intensity for press 2 decreased from 2.30% to 0.62%. For the Hydraulic oil level the MAPE decreased from 2.8% to 1.85%. For the Torque, the MAPE decreased from 2.85% to 2.24%. For the VAT pressure, the MAPE comes from 9.87% to 3.91%. For the Velocity, MAPE decreased from 11.8% to 10.27%. Finally, for the Temperature MAPE decreased from 2.66% to 0.96%.

The quality of the results is confirmed visually which are in general easy to read and show the main trends of the variables.

6.3 Advantages and limitations of proposals

Resorting to the futures market database, we extracted the most relevant data to our study. Having as objective to add information to the model for the prediction of the paper pulp production in the future in the period of 30 days ahead, with the same model it was not possible to perform the prediction since the stock exchange data are in a sample per day making it difficult since it is necessary to reduce the sampling rate of the base and production data of the paper pulp having as result very low sampling rates and in what concerns few samples for the model.

Using other data the same test previously intended was performed and it was found that there is an improvement when information is added to the model input, in this case the derivatives of the respective variables.

Chapter 7

Conclusions and Future Work

7.1 Problem summary

Predicting the future behavior of industrial machinery is key to the success of predictive maintenance. This study aims to find predictive models suitable for accurately predicting the future behavior of industrial plants.

In the industrial world, minimizing downtime is very important. Equipment downtime due to malfunction or curative maintenance means lost production time. To solve this problem, predictive maintenance is the best solution today. Artificial intelligence models have been deployed with the aim of predicting the future behavior of machines and thus avoiding potential failures.

For a prediction to be satisfactory, several factors must be considered, including sensor readings, database records, model processing and tuning. The processing showed a great support to the quality of the results. When processing the data, some variables also showed instability in the learning process and its respective prediction.

In modern industries, prediction algorithms can anticipate future trends and contribute for better management decisions, namely in predictive maintenance.

The predictive model used was based on LSTM networks, with encoding and decoding layers as the input and output, respectively. In this study, different data pre-processing techniques, network architectures, and hyperparameters were tested, in order to determine the best models.

The predictive model used was based on LSTM network, with encoding and decoding layers as the input and output, respectively.

The results show that the model proposed is able to learn and forecast the behavior of the six variables studied: torque, pressure, current intensity, velocity, oil level and temperature. The best results were obtained using a window of samples of the last 10 days at two samples per day. The MAPE errors varied in the range of 2 to 17% for one of the best models for different variables.

7.2 Research Limitations

The research seeks to explore recurrent neural networks, which have shown good efficiency in forecasting time series, and the data acquired present a characteristic that can be called time series. During the research, it was difficult to overcome the following limitations:

- Models developed fail in data that do not have high sampling rates;
- Computational capacity is limited for certain data size;

- Although the models are for prediction for diagnostics, they are only performed to a certain extent period;
- The study is focused on predicting failures without considering the damage to other components of the paper press.The approach is not completely holistic.

7.3 Ideas for future work

The training process of a recurrent neural network becomes time consuming when it comes to high sampling rate. In future work it is intended to make a prediction in real time so that employees can have information of the market behaviour with a head start of 30 days ahead. This will be possible with the increase in computational power and with the methodologies that are emerging, thus making the training process a faster process.

Bibliography

[1] Z. M. Bi, S. Y. Lang, W. Shen, and L. Wang, “Reconfigurable manufacturing systems: the state of the art,” *http://dx.doi.org/10.1080/00207540600905646*, vol. 46, pp. 967–992, 2 2007. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00207540600905646

[2] K. Wang, “Intelligent predictive maintenance (IPdM) system–industry 4.0 scenario,” vol. 113, pp. 259–268, 2016, ISBN: 9781784661694.

[3] R. K. Mobley, *An introduction to predictive maintenance*, 2nd ed. Butterworth-Heinemann, 9 2002, vol. 1.

[4] F. Pardo-Bosch and A. Aguado, “Investment priorities for the management of hydraulic structures,” *Structure and Infrastructure Engineering*, vol. 11, no. 10, pp. 1338–1351, 2015. [Online]. Available: https://doi.org/10.1080/15732479.2014.964267

[5] J. O. Riis, J. T. Luxhøj, and U. Thorsteinsson, “A situational maintenance model,” *International Journal of Quality & Reliability Management*, 1997.

[6] A. Bousdekis, K. Lepenioti, D. Apostolou, and G. Mentzas, “A review of data-driven decision-making methods for industry 4.0 maintenance applications,” vol. 10, no. 7.

[7] M. Pech, J. Vrchota, and J. Bednář, “Predictive maintenance and intelligent sensors in smart factory: Review,” vol. 21, no. 4, pp. 1–39.

[8] A. Martins, I. Fonseca, J. T. Farinha, J. Reis, and A. J. M. Cardoso, “Maintenance prediction through sensing using hidden markov models—a case study,” vol. 11, no. 16, p. 7685. [Online]. Available: https://www.mdpi.com/2076-3417/11/16/7685

[9] A. Berrichi, L. Amodeo, F. Yalaoui, E. Châtelet, and M. Mezghiche, “Bi-objective optimization algorithms for joint production and maintenance scheduling: application to the parallel machine problem,” *Journal of Intelligent Manufacturing*, vol. 20, pp. 389–400, 8 2009.

[10] J. Wang and T. Zhang, “A degradation prediction method by use of autoregressive algorithm,” *Proceedings of the IEEE International Conference on Industrial Technology*, 2008.

[11] S. Duffuaa, A. Kolus, U. Al-Turki, and A. El-Khalifa, “An integrated model of production scheduling, maintenance and quality for a single machine,” *Computers and Industrial Engineering*, vol. 142, 4 2020.

[12] S. V. Amari and L. Mclaughlin, “Optimal design of a condition-based maintenance model,” 2004.

[13] F. M. Bianchi, E. De Santis, A. Rizzi, and A. Sadeghian, “Short-term electric load forecasting using echo state networks and PCA decomposition,” vol. 3, pp. 1931–1943, 2015, conference Name: IEEE Access.

[14] J. Pati, B. Kumar, D. Manjhi, and K. K. Shukla, “A comparison among ARIMA, BP-NN, and MOGA-NN for software clone evolution prediction,” vol. 5, pp. 11 841–11 851, 2017, conference Name: IEEE Access.

[15] H. Akaike, “Autoregressive model fitting for control,” in *Selected Papers of Hirotugu Akaike*, ser. Springer Series in Statistics, E. Parzen, K. Tanabe, and G. Kitagawa, Eds. Springer, pp. 153–170. [Online]. Available: https://doi.org/10.1007/978-1-4612-1694-0_12

[16] S. Ray, S. S. Das, P. Mishra, and A. M. G. Al Khatib, “Time series SARIMA modelling and forecasting of monthly rainfall and temperature in the south asian countries.” [Online]. Available: https://doi.org/10.1007/s41748-021-00205-w

[17] K. Wang and Y. Wang, “How AI affects the future predictive maintenance: A primer of deep learning,” in *Advanced Manufacturing and Automation VII*, ser. Notas de aula sobre engenharia elétrica, K. Wang, Y. Wang, J. O. Strandhagen, and T. Yu, Eds. Springer, 2018, pp. 1–9.

[18] B. C. Mateus, M. Mendes, J. T. Farinha, and A. M. Cardoso, “Anticipating future behavior of an industrial press using LSTM networks,” vol. 11, no. 13, p. 6101, 2021-01. [Online]. Available: https://www.mdpi.com/2076-3417/11/13/6101

[19] M. S. Ahmed and A. R. Cook, “Analysis of freeway traffic time-series data by using box-jenkins techniques,” no. 722, 1979, ISBN: 9780309029728. [Online]. Available: https://trid.trb.org/view/148123

[20] A. Majeed, Y. Zhang, S. Ren, J. Lv, T. Peng, S. Waqar, and E. Yin, “A big data-driven framework for sustainable and smart additive manufacturing,” vol. 67, p. 102026, 2021-02. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0736584520302374

[21] S. Ferreiro, E. Konde, S. Fernández, and A. Prado, “Industry 4.0: predictive intelligent maintenance for production equipment,” in *European Conference of the Prognostics and Health Management Society, no*, 2016, pp. 1–8.

[22] M. C. Monte, E. Fuente, A. Blanco, and C. Negro, “Waste management from pulp and paper production in the european union,” *Waste Management*, vol. 29, pp. 293–308, 1 2009.

[23] (2019) European paper recycling council (eprc) monitoring report 2019. [Online]. Available: https://www.cepi.org/eprc-monitoring-report-2019/

[24] A. Dogan and D. Birant, “Machine learning and data mining in manufacturing,” *Expert Systems with Applications*, vol. 166, 3 2021.

[25] A. S. Murty and V. N. Naikan, "Availability and maintenance cost optimization of a production plant," *International Journal of Quality Reliability Management*, vol. 12, pp. 28–35, 1995.

[26] L. M. S. Gouveia, "Impacto da internet of things no lean manufacturing," 2015-06-23. [Online]. Available: https://ubibliorum.ubi.pt/handle/10400.6/5868

[27] A. Riel, C. Kreiner, G. Macher, and R. Messnarz, "Integrated design for tackling safety and security challenges of smart products and digital manufacturing," *CIRP Annals*, vol. 66, pp. 177–180, 1 2017.

[28] Y. Yin, K. E. Stecke, and D. Li, "The evolution of production systems from industry 2.0 through industry 4.0," *International Journal of Production Research*, vol. 56, no. 1-2, pp. 848–861, 2018. [Online]. Available: https://doi.org/10.1080/00207543.2017.1403664

[29] A. Luque, M. E. Peralta, A. de las Heras, and A. Córdoba, "State of the industry 4.0 in the andalusian food sector," *Procedia Manufacturing*, vol. 13, pp. 1199–1205, 1 2017.

[30] "Human-centric artificial intelligence architecture for industry 5.0 applications," 2022. [Online]. Available: https://arxiv.org/abs/2203.10794

[31] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang, "Industry 4.0 and industry 5.0—inception, conception and perception," *Journal of Manufacturing Systems*, vol. 61, pp. 530–535, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278612521002119

[32] UE, "Indústria 5.0 - serviço das publicações da ue," 2021. [Online]. Available: https://op.europa.eu/en/publication-detail/-/publication/468a892a-5097-11eb-b59f-01aa75ed71a1/

[33] L. Reh, "Process engineering in circular economy," *Particuology*, vol. 11, no. 2, pp. 119–133, 2013, measurement Technology for Particulate System. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1674200113000023

[34] V. Roldão and J. Ribeiro, *Gestão das Operações*, 1st ed. Monitor.

[35] M. Soroush, D. Kaviani, and J. L. Jensen, "Interwell connectivity evaluation in cases of changing skin and frequent production interruptions," *Journal of Petroleum Science and Engineering*, vol. 122, pp. 616–630, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0920410514002770

[36] M. Ben-Daya, "The economic production lot-sizing problem with imperfect production processes and imperfect maintenance," *International Journal of Production Economics*, vol. 76, no. 3, pp. 257–264, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925527301001682

[37] H. Zijm, M. Klumpp, S. Heragu, and A. Regattieri, "Operations, logistics and supply chain management: Definitions and objectives," *Lecture Notes in Logistics*, pp. 27–42, 2019. [Online]. Available: https://research.utwente.nl/en/publications/operations-logistics-and-supply-chain-management-definitions-and-

[38] W. H. Zijm, "Towards intelligent manufacturing planning and control systems," *OR-Spektrum 2000 22:3*, vol. 22, pp. 313–345, 2000. [Online]. Available: https://link.springer.com/article/10.1007/s002919900032

[39] P. E. Green and A. M. Krieger, "Product design strategies for target-market positioning," *Journal of Product Innovation Management*, vol. 8, pp. 189–202, 9 1991.

[40] K. Iwata and Y. Fukuda, "A new proposal of dynamic process planning in machine shop," 1989.

[41] A. H. Tsang, "Strategic dimensions of maintenance management," pp. 7–39, 3 2002.

[42] "Integrating maintenance and production decisions in a hierarchical production planning environment," *Computers Operations Research*, vol. 26, no. 10, pp. 1059–1074, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0

[43] K. K. Boyer and C. McDermott, "Strategic consensus in operations strategy," vol. 17, no. 3, pp. 289–305. [Online]. Available: http://doi.wiley.com/10.1016/S0272-6963(98)00042-4

[44] F. R. Jacobs and R. Chase, *Operations and Supply Chain Management 15th Edition*, 15th ed. McGraw Hill, 2 2017.

[45] R. S. Bojanowski, "Improving factory performance with service requirements planning(srp)." *PRODUCT. INVENT. MANAGE.*, vol. 25, no. 2, pp. 31–44, 1984.

[46] E. G. Hinkelman, *Dictionary of International Trade: Handbook of the Global Trade Community Includes 19 Key Appendices*, 6th ed. World Trade Press.

[47] L. Yang, G. Cai, and J. Chen, "Push, pull, and supply chain risk-averse attitude," *Production and Operations Management*, vol. 27, no. 8, pp. 1534–1552, 2018.

[48] J. P. Peter and J. H. Donnelly, *A Preface to Marketing Management*. McGraw-Hill/Irwin, google-Books-ID: KhQjP6KZ9GQC.

[49] G. R. G. R. Dowling, *The art and science of marketing : marketing for marketing managers*. Oxford : Oxford University Press. [Online]. Available: http://archive.org/details/artscienceofmark0000dowl

[50] K. Venkatesh, M.-C. Zhou, M. Kaighobadi, and R. Caudill, "A petri net approach to investigating push and pull paradigms in flexible factory automated systems," *International Journal of Production Research*, vol. 34, no. 3, pp. 595–620, 1996. [Online]. Available: https://doi.org/10.1080/00207549608904922

[51] A. De Toni, M. Caputo, and A. Vinelli, “Production management techniques: push-pull classification and application conditions,” *International Journal of Operations & Production Management*, vol. 8, no. 2, pp. 35–51, 1988.

[52] . P. P. D. Team, *Pull Production for the Shopfloor*, 1st ed. Productivity Press.

[53] G. Merkuryeva, A. Valberga, and A. Smirnov, “Demand forecasting in pharmaceutical supply chains: A case study,” vol. 149, pp. 3–10. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050919301061

[54] S. Boon-itt, “Using a simulation game approach to teach pull and push production system concepts,” *Engineering Management Research*, vol. 1, no. 1, p. 110, 2012.

[55] J. Olhager and B. Östlund, “An integrated push-pull manufacturing strategy,” *European Journal of Operational Research*, vol. 45, no. 2, pp. 135–142, 1990, oR for Engineers Expert Systems and Decision-Aid. [Online]. Available: https://www.sciencedirect.com/science/article/pii/037722179090180J

[56] M. Bonney, Z. Zhang, M. Head, C. Tien, and R. Barson, “Are push and pull systems really so different?” *International Journal of Production Economics*, vol. 59, no. 1, pp. 53–64, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925527398000942

[57] J. ASHAYERI, A. TEELEN, and W. SELENJ, “A production and maintenance planning model for the process industry,” *International Journal of Production Research*, vol. 34, no. 12, pp. 3311–3326, 1996. [Online]. Available: https://doi.org/10.1080/00207549608905092

[58] M. J. Rosenblatt and H. L. Lee, “Economic production cycles with imperfect production processes,” *IIE Transactions*, vol. 18, no. 1, pp. 48–55, 1986. [Online]. Available: https://doi.org/10.1080/07408178608975329

[59] N. Rezg, A. Chelbi, and X. Xie, “Modeling and optimizing a joint inventory control and preventive maintenance strategy for a randomly failing production unit: Analytical and simulation approaches,” *International Journal of Computer Integrated Manufacturing*, vol. 18, no. 2-3, pp. 225–235, 2005. [Online]. Available: https://doi.org/10.1080/0951192052000288152

[60] M. Ben-Daya and M. Makhdoum, “Integrated production and quality model under various preventive maintenance policies,” *Journal of the Operational Research Society*, vol. 49, no. 8, pp. 840–853, 1998. [Online]. Available: https://doi.org/10.1057/palgrave.jors.2600586

[61] J. Martin, “Maintenance stores and inventory control,” *Maintenance Engineering Handbook, McGraw-Hill, Inc., New York, NY, USA*, 1988.

[62] J. Xie, X. Zhao, and T. Lee, “Freezing the master production schedule under single resource constraint and demand uncertainty,” *International Journal of Production Economics*, vol. 83, no. 1, pp. 65–84, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925527302002621

[63] J. Kimemia and S. B. Gershwin, “An algorithm for the computer control of a flexible manufacturing system,” *IIE Transactions*, vol. 15, no. 4, pp. 353–362, 1983. [Online]. Available: https://doi.org/10.1080/05695558308974659

[64] R. Akella, Y. Choong, and S. Gershwin, “Performance of hierarchical production scheduling policy,” *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, vol. 7, no. 3, pp. 225–240, 1984.

[65] X. Jin, L. Li, and J. Ni, “Option model for joint production and preventive maintenance system,” *International Journal of Production Economics*, vol. 119, no. 2, pp. 347–353, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925527309000875

[66] T. Nakagawa and K. Yasui, “Periodic-replacement models with threshold levels,” *IEEE Transactions on Reliability*, vol. 40, no. 3, pp. 395–397, 1991.

[67] P. Jonsson, “Company-wide integration of strategic maintenance: An empirical analysis,” *International Journal of Production Economics*, vol. 60-61, pp. 155–164, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925527398001480

[68] M. Matsumoto and S. Komatsu, “Demand forecasting for production planning in remanufacturing,” *The International Journal of Advanced Manufacturing Technology*, vol. 79, no. 1, pp. 161–175, 2015.

[69] B. Dong, C. Cao, and S. E. Lee, “Applying support vector machines to predict building energy consumption in tropical region,” *Energy and Buildings*, vol. 37, no. 5, pp. 545–553, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778804002981

[70] A. Azadeh, S. Ghaderi, S. Tarverdian, and M. Saberi, “Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption,” *Applied Mathematics and Computation*, vol. 186, no. 2, pp. 1731–1741, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0096300306011088

[71] Ümmühan Başaran Filik, Ömer Nezih Gerek, and M. Kurban, “A novel modeling approach for hourly forecasting of long-term electric energy demand,” *Energy Conversion and Management*, vol. 52, no. 1, pp. 199–211, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0196890410002803

[72] F.-L. Chu, “Forecasting tourism demand with arma-based methods,” *Tourism Management*, vol. 30, no. 5, pp. 740–751, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0261517708001568

[73] F. De Carlo, O. Borgia, and M. Tucci, "Bucket brigades to increase productivity in a luxury assembly line," *International Journal of Engineering Business Management*, vol. 5, p. 28, 2013.

[74] F. De Carlo, M. A. Arleo, O. Borgia, and M. Tucci, "Layout design for a low capacity manufacturing line: a case study," *International Journal of Engineering Business Management*, vol. 5, no. Godište 2013, pp. 5–35, 2013.

[75] A. A. Mir, M. Alghassab, K. Ullah, Z. A. Khan, Y. Lu, and M. Imran, "A review of electricity demand forecasting in low and middle income countries: The demand determinants and horizons," *Sustainability*, vol. 12, no. 15, 2020. [Online]. Available: https://www.mdpi.com/2071-1050/12/15/5931

[76] R. Dekker, "Applications of maintenance optimization models: a review and analysis," vol. 51, no. 3, pp. 229–240. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0951832095000763

[77] A. Garg and S. Deshmukh, "Maintenance management: literature review and directions," vol. 12, no. 3, pp. 205–238. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/13552510610685075/full/html

[78] A. Sharma, G. Yadava, and S. Deshmukh, "A literature review and future perspectives on maintenance optimization," vol. 17, no. 1, pp. 5–25, publisher: Emerald Group Publishing Limited. [Online]. Available: https://doi.org/10.1108/13552511111116222

[79] R. Dekker, "Applications of maintenance optimization models: a review and analysis," *Reliability Engineering System Safety*, vol. 51, no. 3, pp. 229–240, 1996, maintenance and reliability. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0951832095000763

[80] R. González-Bravo, J. M. Ponce-Ortega, and M. M. El-Halwagi, "Optimal design of water desalination systems involving waste heat recovery," *Industrial & Engineering Chemistry Research*, vol. 56, no. 7, pp. 1834–1847, 2017. [Online]. Available: https://doi.org/10.1021/acs.iecr.6b04725

[81] J. Yulan, J. Zuhua, and H. Wenrui, "Multi-objective integrated optimization research on preventive maintenance planning and production scheduling for a single machine," *International Journal of Advanced Manufacturing Technology*, vol. 39, pp. 954–964, 11 2008.

[82] K. O. Cua, K. E. Mckone, and R. G. Schroeder, "Relationships between implementation of tqm, jit, and tpm and manufacturing performance," pp. 675–694, 2001.

[83] L. Monostori, B. Kádár, T. Bauernhansl, S. Kondoh, S. Kumara, G. Reinhart, O. Sauer, G. Schuh, W. Sihn, and K. Ueda, "Cyber-physical systems in manufacturing," *CIRP Annals*, vol. 65, pp. 621–641, 1 2016.

[84] L. Monostori, B. Kádár, A. Pfeiffer, and D. Karnok, “Solution approaches to real-time control of customized mass production,” *CIRP Annals*, vol. 56, no. 1, pp. 431–434, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0007850607001047

[85] H. Talpaz, M. Cohen, B. Fancher, and J. Halley, “Applying complex models to poultry production in the future-economics and biology,” *Poultry Science*, vol. 92, pp. 2541–2549, 2013.

[86] A. L. Solano-Blanco, J. E. González, L. O. Gómez-Rueda, J. J. Vargas-Sánchez, and A. L. Medaglia, “Integrated planning decisions in the broiler chicken supply chain,” *International Transactions in Operational Research*, 2020.

[87] R. V. Ramani, “Application of computer methods in the mineral industry, proceedings of the symposium, 14th, 1976.” 1977.

[88] J. Wang, Y. Lu, Y. Yang, and T. Mao, “Thermodynamic performance analysis and optimization of a solar-assisted combined cooling, heating and power system,” vol. 115, pp. 49–59, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544216312154

[89] L. G. Hernández-Pérez, C. Ramírez-Márquez, J. G. Segovia-Hernández, and J. M. Ponce-Ortega, “Simultaneous structural and operating optimization of process flowsheets combining process simulators and metaheuristic techniques: The case of solar-grade silicon process,” p. 106946, 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0098135419312402

[90] H. R. Rao and B. P. Lingaraj, “Expert systems in production and operations management: Classification and prospects,” vol. 18, no. 6, pp. 80–91. [Online]. Available: http://pubsonline.informs.org/doi/abs/10.1287/inte.18.6.80

[91] M. Dorigo, G. D. Caro, and L. M. Gambardella, “Ant algorithms for discrete optimization,” vol. 5, no. 2, pp. 137–172, 1999. [Online]. Available: http://www.mitpressjournals.org/doi/10.1162/106454699568728

[92] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed. Springer Science & Business Media.

[93] H. Wang, “A survey of maintenance policies of deteriorating systems,” vol. 139, no. 3, pp. 469–489. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0377221701001977

[94] C. S. Syan and G. Ramsoobag, “Maintenance applications of multi-criteria optimization: A review,” vol. 190, p. 106520. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0951832018313759

[95] B. Jonge and P. A. Scarf, "A review on maintenance optimization," p. S0377221719308045. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0377221719308045

[96] A. Bousdekis, N. Papageorgiou, B. Magoutas, D. Apostolou, and G. Mentzas, "A proactive event-driven decision model for joint equipment predictive maintenance and spare parts inventory optimization," vol. 59, pp. 184–189. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2212827116309490

[97] R. K. Mobley, *An Introduction to Predictive Maintenance (Plant Engineering)*, 2nd ed. Butterworth-Heinemann, google-Books-ID: SjqXzxpAzSQC.

[98] D. J. Edwards, G. D. Holt, and F. Harris, "Predictive maintenance techniques and their relevance to construction plant," vol. 4, no. 1, pp. 25–37. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/13552519810369057/full/html

[99] B. Zhou, Y. Qi, and Y. Liu, "Proactive preventive maintenance policy for buffered serial production systems based on energy saving opportunistic windows," vol. 253, p. 119791. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095965261934661X

[100] Z. Zhao, F.-l. Wang, M.-x. Jia, and S. Wang, "Predictive maintenance policy based on process data," vol. 103, no. 2, pp. 137–143. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016974391000119X

[101] J. M. T. Farinha, *Asset Maintenance Engineering Methodologies*. CRC Press, google-Books-ID: 3SJWDwAAQBAJ.

[102] H. M. Hashemian, "State-of-the-art predictive maintenance techniques," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, pp. 226–236, 1 2011.

[103] M. C. Eti, S. O. T. Ogaji, and S. D. Probert, "Reducing the cost of preventive maintenance (PM) through adopting a proactive reliability-focused culture," vol. 83, no. 11, pp. 1235–1248. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261906000079

[104] C. H. Lie and Y. H. Chun, "An algorithm for preventive maintenance policy," *IEEE Transactions on Reliability*, vol. 35, no. 1, pp. 71–75, 1986.

[105] J. Farinha, *Manutenção - A Terologia e as Novas Ferramentas de Gestão*, monitor ed.

[106] J. Wan, S. Tang, D. Li, S. Wang, C. Liu, H. Abbas, and A. V. Vasilakos, "A manufacturing big data solution for active preventive maintenance," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2039–2047, 2017.

[107] M. Zaeri, J. Shahrabi, M. Pariazar, and A. Morabbi, "A combined multivariate technique and multi criteria decision making to maintenance strategy selection," in *2007 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE, pp. 621–625. [Online]. Available: http://ieeexplore.ieee.org/document/4419264/

[108] R. Masoni, F. Ferrise, M. Bordegoni, M. Gattullo, A. E. Uva, M. Fiorentino, E. Carrabba, and M. Di Donato, "Supporting remote maintenance in industry 4.0 through augmented reality," *Procedia Manufacturing*, vol. 11, pp. 1296–1302, 2017, 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2351978917304651

[109] J. P. Gilbert and B. J. Finch, "Maintenance management: Keeping up with production's changing trends and technologies," *Journal of Operations Management*, vol. 6, pp. 1–12, 11 1985.

[110] J. Kershaw and B. Robertson, "Condition-based maintenance program increases production, reduces costs," pp. 34–36.

[111] Q. Chang, J. Ni, P. Bandyopadhyay, S. Biller, and G. Xiao, "Maintenance opportunity planning system," vol. 129, no. 3, pp. 661–668, publisher: American Society of Mechanical Engineers Digital Collection. [Online]. Available: https://asmedigitalcollection.asme.org/manufacturingscience/article/129/3/661/462149/Maintenance-Opportunity-Planning-System

[112] M. Carnero, "Selection of diagnostic techniques and instrumentation in a predictive maintenance program. a case study," vol. 38, no. 4, pp. 539–555. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167923603001283

[113] "Urea cycle dysregulation in non-alcoholic fatty liver disease," vol. 69, no. 4, pp. 905–915. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168827818321767

[114] Q. Chang, J. Ni, P. Bandyopadhyay, S. Biller, and G. Xiao, "Maintenance opportunity planning system," *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, vol. 129, pp. 661–668, 6 2007.

[115] T. Bedford, I. Dewan, I. Meilijson, and A. Zitrou, "The signal model: A model for competing risks of opportunistic maintenance," *European Journal of Operational Research*, vol. 214, pp. 665–673, 2011.

[116] X. Zhou, K. Huang, L. Xi, and J. Lee, "Preventive maintenance modeling for multi-component systems with considering stochastic failures and disassembly sequence," *Reliability Engineering System Safety*, vol. 142, pp. 231–237, 10 2015.

[117] R. Sipos, D. Fradkin, F. Moerchen, and Z. Wang, "Log-based predictive maintenance," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1867–1876, 2014.

[118] A. Gill, "Optimisation of the technical object maintenance system taking account of risk analysis results," vol. 19, no. 3, pp. 420–431. [Online]. Available: http://www.ein.org.pl/sites/default/files/2017-03-13.pdf

[119] A. H. Tsang, “Condition-based maintenance: tools and decision making,” *Journal of quality in maintenance engineering*, 1995.

[120] O. Aydin and S. Guldamlasioglu, “Using LSTM networks to predict engine condition on large scale data processing framework,” in *2017 4th International Conference on Electrical and Electronic Engineering (ICEEE)*, 2017-04, pp. 281–285.

[121] D. Dong, X.-Y. Li, and F.-Q. Sun, “Life prediction of jet engines based on LSTM-recurrent neural networks,” in *2017 Prognostics and System Health Management Conference (PHM-Harbin)*, 2017-07, pp. 1–6, ISSN: 2166-5656.

[122] V. Mathew, T. Toby, V. Singh, B. M. Rao, and M. G. Kumar, “Prediction of remaining useful lifetime (RUL) of turbofan engine using machine learning,” in *2017 IEEE International Conference on Circuits and Systems (ICCS)*, 2017-12, pp. 306–311.

[123] H. V. Düdükçü, M. Taşkıran, and N. Kahraman, “LSTM and WaveNet implementation for predictive maintenance of turbofan engines,” in *2020 IEEE 20th International Symposium on Computational Intelligence and Informatics (CINTI)*, 2020-11, pp. 000 151–000 156, ISSN: 2471-9269.

[124] S. Song, D. W. Coit, and Q. Feng, “Reliability for systems of degrading components with distinct component shock sets,” *Reliability Engineering System Safety*, vol. 132, pp. 115–124, 12 2014.

[125] K. S. Moghaddam, “Multi-objective preventive maintenance and replacement scheduling in a manufacturing system using goal programming,” *International Journal of Production Economics*, vol. 146, pp. 704–716, 12 2013.

[126] L. A. Hadidi, U. M. Al-Turki, and M. A. Rahim, “Joint job scheduling and preventive maintenance on a single machine,” *International Journal of Operational Research*, vol. 13, pp. 174–184, 1 2012.

[127] J.-H. Shin and H.-B. Jun, “On condition based maintenance policy,” vol. 2, no. 2, pp. 119–127. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2288430014000141

[128] H. Ab-Samat and S. Kamaruddin, “Opportunistic maintenance (om) as a new advancement in maintenance approaches: A review,” *Journal of Quality in Maintenance Engineering*, 2014.

[129] T. Xia, L. Xi, E. Pan, and J. Ni, “Reconfiguration-oriented opportunistic maintenance policy for reconfigurable manufacturing systems,” *Reliability Engineering and System Safety*, vol. 166, pp. 87–98, 10 2017.

[130] B. Rao, “Opportunistic maintenance of multi-equipment system: a case study,” p. 14.

[131] J. Koochaki, J. A. Bokhorst, H. Wortmann, and W. Klingenberg, “Condition based maintenance in the context of opportunistic maintenance,” *International Journal of Production Research*, vol. 50, pp. 6918–6929, 12 2012.

[132] H. W. Javid Koochaki, Jos A.C. Bokhorst and W. Klingenberg, "Condition based maintenance in the context of opportunistic maintenance," *International Journal of Production Research*, vol. 50, no. 23, pp. 6918–6929, 2012. [Online]. Available: https://doi.org/10.1080/00207543.2011.636924

[133] A. N. Rao and B. Bhadury, "Opportunistic maintenance of multi-equipment system: a case study," *Quality and Reliability Engineering International*, vol. 16, no. 6, pp. 487–500, 2000.

[134] X. Zhou, L. Xi, and J. Lee, "Opportunistic preventive maintenance scheduling for a multi-unit series system based on dynamic programming," vol. 118, no. 2, pp. 361–366. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925527308003472

[135] F. Ding and Z. Tian, "Opportunistic maintenance for wind farms considering multi-level imperfect maintenance thresholds," vol. 45, pp. 175–182. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148112001802

[136] C. Letot, I. Soleimanmeigouni, A. Ahmadi, and P. Dehombreux, "An adaptive opportunistic maintenance model based on railway track condition prediction," vol. 49, no. 28, pp. 120–125. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2405896316324454

[137] H. Truong Ba, M. E. Cholette, P. Borghesani, Y. Zhou, and L. Ma, "Opportunistic maintenance considering non-homogenous opportunity arrivals and stochastic opportunity durations," vol. 160, pp. 151–161. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095183201631002X

[138] B. Iung, E. Levrat, and E. Thomas, "'odds algorithm'-based opportunistic maintenance task execution for preserving product conditions," vol. 56, no. 1, pp. 13–16. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0007850607000054

[139] B. de Jonge and P. A. Scarf, "A review on maintenance optimization," *European Journal of Operational Research*, vol. 285, no. 3, pp. 805–824, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377221719308045

[140] H. Wang, "A survey of maintenance policies of deteriorating systems," *European Journal of Operational Research*, vol. 139, no. 3, pp. 469–489, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377221701001977

[141] F. Besnard, M. Patrikssont, A.-B. Strombergt, A. Wojciechowskit, and L. Bertling, "An optimization framework for opportunistic maintenance of offshore wind power system," in *2009 IEEE Bucharest PowerTech*, pp. 1–7.

[142] J. Hu, L. Zhang, and W. Liang, "Opportunistic predictive maintenance for complex multi-component systems based on dbn-hazop model," *Process Safety and Environmental Protection*, vol. 90, no. 5, pp. 376–388, 2012, special Issue Celebrating Trevor Kletz's 90th Birthday. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957582012000614

[143] I. Daniyan, K. Mpofu, M. Oyesola, B. Ramatsetse, and A. Adeodu, “Artificial intelligence for predictive maintenance in the railcar learning factories,” vol. 45, pp. 13–18, 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2351978920310702

[144] K. Benke and G. Benke, “Artificial intelligence and big data in public health,” *International Journal of Environmental Research and Public Health*, vol. 15, no. 12, 2018. [Online]. Available: https://www.mdpi.com/1660-4601/15/12/2796

[145] M. Mitchell, “Artificial intelligence hits the barrier of meaning,” *Information*, vol. 10, no. 2, 2019. [Online]. Available: https://www.mdpi.com/2078-2489/10/2/51

[146] R. Liu, B. Yang, E. Zio, and X. Chen, “Artificial intelligence for fault diagnosis of rotating machinery: A review,” *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327018300748

[147] A. C. Müller and S. Guido, *Introduction to machine learning with Python: a guide for data scientists*. ” O’Reilly Media, Inc.”, 2016.

[148] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. Basto, and S. G. S. Alcalá, “A systematic literature review of machine learning methods applied to predictive maintenance,” *Computers & Industrial Engineering*, vol. 137, p. 106024, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360835219304838

[149] T. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben, “Machine learning in manufacturing: advantages, challenges, and applications,” *Production & Manufacturing Research*, vol. 4, no. 1, pp. 23–45, 2016.

[150] “Ensemble learning methodologies for soft sensor development in industrial processes,” accepted: 2015-02-23T10:59:05Z. [Online]. Available: https://estudogeral.sib.uc.pt/handle/10316/28313

[151] J.-H. Shin, H.-B. Jun, and J.-G. Kim, “Dynamic control of intelligent parking guidance using neural network predictive control,” vol. 120, pp. 15–30, 2018-06-01. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360835218301591

[152] M. Paolanti, L. Romeo, A. Felicetti, A. Mancini, E. Frontoni, and J. Loncarski, “Machine learning approach for predictive maintenance in industry 4.0,” in *2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, pp. 1–6.

[153] S. Deng, L. Huang, G. Xu, X. Wu, and Z. Wu, “On deep learning for trust-aware recommendations in social networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 5, pp. 1164–1177, 2017.

[154] X. Bampoula, G. Siaterlis, N. Nikolakis, and K. Alexopoulos, “A deep learning model for predictive maintenance in cyber-physical production systems using LSTM autoencoders,” vol. 21, no. 3, p. 972, 2021-02, place: Basel Publisher: Mdpi WOS:000615495200001.

[155] K. Yasaka, H. Akai, A. Kunimatsu, S. Kiryu, and O. Abe, “Deep learning with convolutional neural network in radiology,” vol. 36, no. 4, pp. 257–272, 2018-04-01. [Online]. Available: https://doi.org/10.1007/s11604-018-0726-3

[156] N. Kriegeskorte and T. Golan, “Neural network models and deep learning,” vol. 29, no. 7, pp. R231–R236. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960982219302040

[157] L. Zhang, F. Yang, Y. Daniel Zhang, and Y. J. Zhu, “Road crack detection using deep convolutional neural network,” in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016-09, pp. 3708–3712, ISSN: 2381-8549.

[158] Y. Bengio, *Learning deep architectures for AI.* Now Publishers Inc, 2009.

[159] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, “Extracting and composing robust features with denoising autoencoders,” in *Proceedings of the 25th international conference on Machine learning*, ser. ICML ’08. Association for Computing Machinery, 2008-07-05, pp. 1096–1103. [Online]. Available: https://doi.org/10.1145/1390156.1390294

[160] H. Lee, P. Pham, Y. Largman, and A. Ng, “Unsupervised feature learning for audio classification using convolutional deep belief networks,” *Advances in neural information processing systems*, vol. 22, pp. 1096–1104, 2009.

[161] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[162] Y.-Y. Hsu, T.-T. Tung, H.-C. Yeh, and C.-N. Lu, “Two-stage artificial neural network model for short-term load forecasting,” vol. 51, no. 28, pp. 678–683, 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2405896318335043

[163] M. Balduíno, J. Torres Farinha, and A. Marques Cardoso, “Production optimization versus asset availability – a review,” vol. 15, pp. 320–332, 2020-08-17. [Online]. Available: https://www.wseas.org/multimedia/journals/control/2020/a665103-968.pdf

[164] P. Lara-Benítez, M. Carranza-García, J. M. Luna-Romera, and J. C. Riquelme, “Temporal convolutional networks applied to energy-related time series forecasting,” vol. 10, no. 7, pp. 1–17, 2020.

[165] J. Yeomans, S. Thwaites, W. S. Robertson, D. Booth, B. Ng, and D. Thewlis, “Simulating time-series data for improved deep neural network performance,” vol. 7, pp. 131 248–131 255, 2019.

[166] Z. Yu, D. S. Moirangthem, and M. Lee, “Continuous timescale long-short term memory neural network for human intent understanding,” vol. 11, p. 42, 2017-08-23. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnbot.2017.00042/full

[167] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder-decoder architecture," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 1672–1678. [Online]. Available: https://ieeexplore.ieee.org/document/8500658/

[168] K. Cho, B. van Merrienboer, Ç. Gülçehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *CoRR*, vol. abs/1406.1078, 2014. [Online]. Available: http://arxiv.org/abs/1406.1078

[169] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," 2014-12-14. [Online]. Available: http://arxiv.org/abs/1409.3215

[170] T. Wang, P. Chen, K. Amaral, and J. Qiang, "An experimental study of lstm encoder-decoder model for text simplification," 2016.

[171] S. Bengio, O. Vinyals, N. Jaitly, and N. Shazeer, "Scheduled sampling for sequence prediction with recurrent neural networks," 2015.

[172] S. Du, T. Li, Y. Yang, and S.-J. Horng, "Multivariate time series forecasting via attention-based encoder–decoder framework," vol. 388, pp. 269–279, 2020-05-07. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231220300606

[173] P. Bangalore and L. B. Tjernberg, "An artificial neural network approach for early fault detection of gearbox bearings," vol. 6, no. 2, pp. 980–987, 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7012091/

[174] A. Beshr and F. Zarzoura, "Using artificial neural networks for GNSS observations analysis and displacement prediction of suspension highway bridge," vol. 6, no. 2, 2021.

[175] M. S. Jahan, B. G. Gunter, and A. Rahman, "Substituting wood with nonwood fibers in papermaking: A win-win solution for bangladesh," 2009.

[176] P. Bajpai, S. P. Mishra, O. Mishra, S. Kumar, P. K. Bajpai, and S. Singh, "Biochemical pulping of wheat straw," *Tappi Journal*, vol. 3, pp. 3–6, 2004.

[177] L. Szabó, A. Soria, J. Forsström, J. Keränen, and E. Hytönen, "A world model of the pulp and paper industry: Demand, energy consumption and emission scenarios to 2030," *Environmental Science Policy*, vol. 12, no. 3, pp. 257–269, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1462901109000227

[178] P. Shenoy and P. Aithal, "A study on history of paper and possible paper free world," *International Journal of Management, IT and Engineering*, vol. 6, no. 1, pp. 337–355, 2016.

[179] A. Mohd Aripin, "Potential of non-wood fibres for pulp and paper-based industries," Ph.D. dissertation, Universiti Tun Hussein Onn Malaysia, 2014.

[180] W. Sridach, "The environmentally benign pulping process of non-wood fibers." *Suranaree Journal of Science & Technology*, vol. 17, no. 2, 2010.

[181] M. Chandra IV, "Use of nonwood plant fibers for pulp and paper industry in asia: Potential in china," 1998.

[182] M. Valenzuela, J. Bentley, and R. Lorenz, "Evaluation of torsional oscillations in paper machine sections," *IEEE Transactions on Industry Applications*, vol. 41, no. 2, pp. 493–501, 2005.

[183] K. Holmberg, R. Siilasto, T. Laitinen, P. Andersson, and A. Jäsberg, "Global energy consumption due to friction in paper machines," *Tribology International*, vol. 62, pp. 58–77, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301679X13000443

[184] A. Zvolinschi, E. Johannessen, and S. Kjelstrup, "The second-law optimal operation of a paper drying machine," *Chemical Engineering Science*, vol. 61, no. 11, pp. 3653–3662, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0009250906000078

[185] G. Stewart, D. Gorinevsky, and G. Dumont, "Feedback controller design for a spatially distributed system: the paper machine problem," *IEEE Transactions on Control Systems Technology*, vol. 11, no. 5, pp. 612–628, 2003.

[186] W. S. Cleveland, "Lowess: A program for smoothing scatterplots by robust locally weighted regression," *The American Statistician*, vol. 35, no. 1, pp. 54–54, 1981. [Online]. Available: http://www.jstor.org/stable/2683591

[187] A. Agresti, *Categorical data analysis*. New Jersey: John Wiley & Sons, 2003.

[188] G. M. R. G. C. L. G. M. Box, George E. P.; Jenkins, *Time series analysis : forecasting and control*. John Wiley Sons, 2016.

[189] R. H. Shumway and D. S. Stoffer, *Time series analysis and its applications*. New York: Springer, 2000.

[190] I. Khandelwal, R. Adhikari, and G. Verma, "Time series forecasting using hybrid arima and ann models based on dwt decomposition," *Procedia Computer Science*, vol. 48, pp. 173–179, 2015, international Conference on Computer, Communication and Convergence (ICCC 2015). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050915006766

[191] H.-l. Yip, H. Fan, and Y.-h. Chiang, "Predicting the maintenance cost of construction equipment: Comparison between general regression neural network and box–jenkins time series models," vol. 38, pp. 30–38. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0926580513001921

[192] R. Sarikaya, G. E. Hinton, and A. Deoras, "Application of deep belief networks for natural language understanding," vol. 22, no. 4, pp. 778–784, 2014-04, conference Name: IEEE/ACM Transactions on Audio, Speech, and Language Processing.

[193] M. Sundermeyer, R. Schlüter, and H. Ney, "LSTM neural networks for language modeling," in *Thirteenth annual conference of the international speech communication association*, 2012.

[194] Y. Hu, X. Sun, X. Nie, Y. Li, and L. Liu, "An enhanced LSTM for trend following of time series," vol. 7, pp. 34 020–34 030, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8630920/

[195] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with LSTM," pp. 850–855, 1999-01-01, publisher: IET Digital Library. [Online]. Available: https://digital-library.theiet.org/content/conferences/10.1049/cp_19991218

[196] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," vol. 28, no. 10, pp. 2222–2232, 2017-10, conference Name: IEEE Transactions on Neural Networks and Learning Systems.

[197] X. Kong, D. Kong, J. Yao, L. Bai, and J. Xiao, "Online pricing of demand response based on long short-term memory and reinforcement learning," vol. 271, p. 114945. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261920304578

[198] S. Ayvaz and K. Alpay, "Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time," vol. 173, p. 114598, 2021-07-01. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421000397

[199] A. Essien and C. Giannetti, "A deep learning model for smart manufacturing using convolutional LSTM neural network autoencoders," vol. 16, no. 9, pp. 6069–6078, 2020-09, conference Name: IEEE Transactions on Industrial Informatics.

[200] J. Schmidhuber, "Deep learning in neural networks: An overview," vol. 61, pp. 85–117, 2015-01-01. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608014002135

[201] H. Palangi, L. Deng, Y. Shen, J. Gao, X. He, J. Chen, X. Song, and R. Ward, "Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval," vol. 24, no. 4, pp. 694–707, 2016-04, conference Name: IEEE/ACM Transactions on Audio, Speech, and Language Processing.

[202] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," vol. 323, no. 6088, pp. 533–536, 1986-10, bandiera_abtest: a Cg_type: Nature Research Journals Number: 6088 Primary_atype: Research Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/323533a0

[203] Y. Li and Y. Lu, "LSTM-BA: DDoS detection approach combining LSTM and bayes," in *2019 Seventh International Conference on Advanced Cloud and Big Data (CBD)*, 2019-09, pp. 180–185.

[204] Z. Alameer, A. Fathalla, K. Li, H. Ye, and Z. Jianhua, "Multistep-ahead forecasting of coal prices using a hybrid deep learning model," vol. 65, p. 101588, 2020. [Online]. Available: https://doi.org/10.1016/j.resourpol.2020.101588

[205] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," 2014.

[206] H. M. Lynn, S. B. Pan, and P. Kim, "A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks," vol. 7, pp. 145 395–145 405, conference Name: IEEE Access.

[207] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[208] F. Chollet, "keras," https://github.com/fchollet/keras, 2015.

[209] M. Chai, F. Xia, S. Hao, D. Peng, C. Cui, and W. Liu, "PV power prediction based on LSTM with adaptive hyperparameter adjustment," vol. 7, pp. 115 473–115 486, conference Name: IEEE Access.

[210] N. Reimers and I. Gurevych, "Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks." [Online]. Available: http://arxiv.org/abs/1707.06799

[211] S. Merity, N. S. Keskar, and R. Socher, "An analysis of neural language modeling at multiple scales." [Online]. Available: http://arxiv.org/abs/1803.08240

[212] B. C. Mateus, M. Mendes, J. T. Farinha, and A. M. Cardoso, "Anticipating future behavior of an industrial press using lstm networks," *Applied Sciences*, vol. 11, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/13/6101

[213] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction," *Journal of Big Data*, vol. 8, no. 1, pp. 1–28, 2021.

[214] C. Chen, Y. Liu, S. Wang, X. Sun, C. Di Cairano-Gilfedder, S. Titmus, and A. A. Syntetos, "Predictive maintenance using cox proportional hazard deep learning,"

Advanced Engineering Informatics, vol. 44, p. 101054, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474034620300239

[215] Z. Gui, Y. Sun, L. Yang, D. Peng, F. Li, H. Wu, C. Guo, W. Guo, and J. Gong, "Lsi-lstm: An attention-aware lstm for real-time driving destination prediction by considering location semantics and location importance of trajectory points," *Neurocomputing*, vol. 440, pp. 72–88, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092523122100134X

Appendix A

Production Optimization versus Asset Availability – a Review

[1,2]Balduíno Mateus; [3,4]José Torres Farinha; [2]António Marques Cardoso
balduino.mateus@ubi.pt; torres.farinha@dem.uc.pt; ajmarcar@gmail.com
[1]Industrial Eng. and Management, Univ. Lusofona. Lisbon, Portugal
[2]CISE – Electromechatronic Systems Research Centre, Univ. Beira Interior, Portugal
[3]CEMMPRE - Centre for Mechanical Engineering, Materials and Processes, Univ. of Coimbra, Portugal
[4]ISEC/IPC - Polytechnic Institute of Coimbra, Portugal

Abstract: -Nowadays, companies want to give a quick answer in order to face their market competitors. These quick responses must be reflected in the quality of the products; to this be possible, it is necessary to manage a number of factors that will bring benefits in its market positioning. As technology grows, there is the possibility, at a computational level, to create a combination of mathematical and technological tools that were not implemented in the past due to the lack of resources, since they have high robustness about their analytical resolution.
This paper presents mathematical and computer tools that have potential great benefits when applied to industrial problems solving, such as operation management.
Along the paper it is made a temporal location of all tools with their main objectives about optimizing industrial processes, focusing on maintenance costs, contributing directly to the rationalization of global costs of the processes.
Analytical and technological methods that have had great success regarding to the reduction costs of production in industries are presented. The approaches of this paper bring a literary review of process optimization, namely about Neural Networks and multivariate analysis for prediction.

Key-Words: Optimization; Production; Forecast; Neural Networks; Multivariate Analysis; Industrial Maintenance

1 Introduction

The research on maintenance optimization has been a priority [1], also having been a big trend in the area of optimization based on maintenance simulation [2], [3]. The research in this context has as its main objective to find the best maintenance plans that minimize the general maintenance cost or maximize a system performance measure bringing cost reductions [4]. Maintenance plays an important role in the industry, being responsible for improving the availability of assets, thus reducing the downtime in industrial plants. The cost of maintenance can vary between the values of 15% and 70% of the production costs [5]. With the technological evolution of industrial processes, maintenance has increased its complexity [5], [6]. This is mostly due to modern manufacturing systems that involve numerous interactions and dependencies among components [7]. Corrective and preventive maintenance aim to take the systems "as good as new". Regarding the predictive substitution model policy, under some assumptions, the one that fits most practical situations, has proved to have a limit control policy [8].
J. Khalil [9] Predicting the failure rate of each machine part is possible. Because of this, the following actions must be executable, spending the minimum effort:

- get a clear view of the industrial domain (usually vast, confusing and complicated) in an industry, in terms of its most basic units (the machine parts);
- obtain an economical and scientifically calculated service life for each machine part in the industrial domain;
- to adapt preventive actions as opportunities;
- to be able to refine restricted production, subject to the relationship between recent and historical trends in machine part failures; and
- to obtain a mathematical formulation of the costs and the probabilities of survival of each part of the machine throughout its useful life, in order to be able to change the availability of that part with full knowledge of the financial consequences.

The structure of maintenance cycles in the preventive strategy makes the objective function of the decision-making problem discontinuous. Therefore, it is suggested to solve the problem with the use of dynamic programming methods.
In the development of a dedicated computer program, it was possible to highlight possible resolutions for decision-making problems in relation to: manner, scope and schedules of replacements, repairs and regular maintenance of elements of technical objects, mode and schedules of diagnosis and preventive replacement of elements and problems of supplying spare parts to the maintenance subsystem [10].

Based on the inspiration of [11], three slightly different decision dimensions are proposed: ”product dimension”; ”risk dimension”; and ”resource dimension”. This research concludes that these decision dimensions must be considered simultaneously, at the same time, in which they optimize maintenance in an integrated way. The structure of this paper is the following:

- The first section presents the introduction of the theme;
- The second section presents the state of the art, namely a global approach about optimization;
- The third section presents proposals for possible resolution of the problems in this area;
- The fourth section presents a discussion based on the results of the research;
- Finally, they are presented the conclusions.

The degree of adaptation of a competitive organization imply priorities in its primary decisions related to structural and infrastructural investments, providing the key to the progress of the full potential of their operations as a competitive weapon. Figure 1 illustrates a graphic model about the operations strategy generally accepted [12].

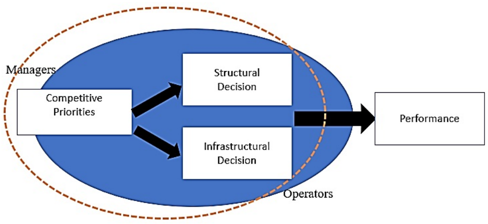


Figure 1: Operations strategy model [12]

The paper is focused on the aggregation of knowledge for future implementations of the tools covered in the scope of operations management; in the second section it is made a retrospective of the premises of the maintenance policies focused on production costs that supports the paper case studies. In this same section, the forecasting tools, based on multivariate analysis and neural networks are addressed. In section three the proposal for the next phase of the study is presented and, in third section, the state of the art of these tools is presented; in section five a conclusion of the study is made.

2 State of the Art

2.1 Global approach

The poor maintenance of our car can lead its poor performance, causing malfunction at an inopportune moment, even causing a bad flow on the road. An operation’s manager does not want this happen on his manufacturing line. Therefore, it can be said that maintenance is a fundamental pillar for the fluidity of the production plant. This paper focuses on maintenance management and in what it can prevent failure and increase equipment availability. Compared to the costs associated with maintenance, the production system, according to Dhillon [13] becomes increasingly important, because, in many industrial plants, the maintenance costs can exceed 30% of the operational costs and, in the context of the service life cycle, maintenance and support of manufacturing systems, represent 60 to 75% of the total cost of the life cycle. According to Mobley [14], the production capacity of a plant is partly demarcated by the availability of production systems and their auxiliary equipment, being the main function of the maintenance organization to ensure that all equipment and systems in the plant are up to date, always “online” and in good operating conditions, without jeopardizing profitability and, always possible using other tools such as optimization. Although it seems that optimization is an activity present in people’s daily lives, according to Cua, McKone & Schroeder [15], this is the best option to carry out a task without compromising restrictions: a fast and efficient task that companies use to improve and make more profitable processes. The same authors say that, the optimization techniques seek the best solution for each problem (maximum or minimum of measurable quantities in their domains of definition); they are necessary tools in many areas of engineering, such as:

- Operational research - optimization of technical and economic systems, stock control, production planning, etc.;
- Process engineering - process sizing and optimization, parameter estimation, data grouping, flexibility analysis, etc.;
- Process control - system identification, optimal control, adaptive control, predictive control, state estimators, etc.;
- Numerical analysis - approximations, regression, solution of linear and non-linear systems, etc.

Wang *et al.* [16] say that an optimization process is needed to determine the optimal capabilities and operational strategy. There has been a great contribution in making optimization in a dynamic way by combining process simulators with metaheuristic techniques

for simultaneous optimization of process flowcharts with the corresponding operating conditions [17].
The Operations Management have applications into strategic and tactically oriented applications, namely in next areas: aggregate planning; forecasting; location decisions; scheduling; capacity planning; layout; process and product design; quality control; task design; control inventory; maintenance and reliability [18].
In the industrial area, it is also possible to carry out an adaptation of the optimization model, demonstrating their efficiency for the solution of complex problems; it can start with simpler problems, observing what already exists in nature, for example, [19] retransmitting a study that presents an overview of recent work on ant algorithms, i.e., algorithms for discrete optimization that were inspired by the observation of ant colony forage behaviour and presenting the ant colony optimization metaheuristic.
The field of metaheuristics for applying combinatorial optimization problems is a rapidly growing field of research. This is due to the importance of combinatorial optimization problems for the scientific and industrial world [20]. Alaswad & Xiang [21] present a review of the Condition Based Maintenance (CBM) literature with emphasis on mathematical modelling and optimization approaches. They focused the review on important aspects of the CBM, such as optimization criteria, frequency of inspection, degree of maintenance, solution methodology, etc.
Nocedal & Wright [22] refer that knowledge of the capabilities and limitations of optimization algorithms leads to a better understanding of their impact in various applications and points the way for future research on algorithms and software for improvement and extension. To encompass the optimization of a digital industry, it is necessary to fulfil some requirements such as the robustness of the database and the reliability of the data/samples. According to Dekker [1], there are several applications of maintenance optimization models that generally cover four aspects:

1. A description of a technical system (its function and importance);
2. A model for the deterioration of the system over time and possible consequences for the system;
3. A description of the information available on the system and the actions available to management;
4. An objective function and an optimization technique that helps to find the best balance.

These maintenance optimization models produce different results. First, policies can be evaluated and compared to the characteristics of cost-effectiveness and reliability.

Wang [23] presents an extensive review of maintenance optimization policies. Maintenance optimization studies prior to 2002 mainly considered time-based maintenance configurations. Syan & Ramsoobag [24] state that modern maintenance optimization decisions are complex problems that need to satisfy multiple and conflicting criteria. With the increase in applications and recent advances in Multi-Criteria Optimization (MCO) approaches, a review is needed to group and categorize these advances in the field of maintenance. Jonge & Scarf [25] says that optimization applied to maintenance comprises the development and analysis of mathematical models that aim to improve or optimize maintenance policies. A study on the substantial developments in the field of maintenance optimization is fully demonstrated in [23].
In order to validate the effectiveness of decision models, Bousdekis *et al.* [26] prove that an event-driven proactive decision model is possible for joint predictive maintenance and optimization of the spare parts inventory, which addresses the ”Detect” ”prevent-decide-act” model phase that can be incorporated into an EDA (Event Oriented Architecture) for processing time within the framework of the concept of electronic maintenance. Zhou, Qi & Liu [27] show some drawbacks that the ideal maintenance policy is, in fact, a monotony of control limits, in which the control limits decrease monotonously with the age of the system; but, other studies expose some solutions like Zhao *et al.* [28], that propose a predictive maintenance policy based on process data, demonstrating that, when compared to traditional preventive maintenance strategies, their strategy have adaptability and effectiveness to the deterioration of the system. Among the techniques presented in [14], there are five non-invasive techniques used for the management of predictive maintenance such as monitoring vibrations, monitoring process parameters, thermography, tribology and visual inspection. Predictive techniques can vary, as mentioned in [29]: lubricant analysis; vibration analysis; thermography; penetrating liquids; radiography; ultrasound; corrosion control; etc.
There is a current concern in making predictive diagnostics to avoid unwanted costs of maintenance and production, having industries and researchers bet on their resources to solve this dilemma. One of the main focuses of multivariate analysis is the search for models that best apply to a forecast or even explain the behaviour of the variables studied. In this section, applications of multivariate models are presented in the scope of engineering, with a focus on asset management during the life cycle monitoring production and maintenance management to reduce unwanted costs. The reduction of expenses in the industrial sphere is

a great challenge in which it involves a chain of responsibility from the worker who is fully with the machine passing through the managers and reaching the researchers, who have as challenge the way to respond effectively to these events unwanted. Vatn, Hokstad & Bodsberg [30] present a global approach to maintenance optimization that requires knowledge in several fields; for example, decision theory, risk analysis, reliability and maintenance modelling. The authors also refer that, in order to perform maintenance optimization, it is generally not feasible to carry out a complete risk analysis, and the effect of the chosen maintenance program on safety, that can be assessed by a somewhat simplified approach, considering only a very limited number of scenario representatives. Besnard *et al.* [31] present an opportunistic maintenance optimization model for offshore wind energy systems. The model takes advantage of wind forecasts and corrective maintenance activities to perform preventive low cost maintenance tasks. Pinto & Nascif [32] refer that, sometimes, the increase in reliability is done through the expense of availability. This suggests a threefold restriction between quality, availability, and reliability of maintenance, directly influencing costs (Figure 2).

Figure 2: Triple maintenance constraint [32]

Maintenance plays a key role in reliability, availability, product quality, risk reduction, greater equipment efficiency and safety [33]. According to Kershaw & Robertson [34], predictive maintenance works with periodic monitoring of component conditions, instead of replacing them, which means better data, increasing plant productivity and preventing disastrous failures. For example, Chang *et al.* [35] explore the optimization of maintenance, presenting a method that incorporates information in real time about production conditions and machine failure. Carnero [36] states that Predictive Maintenance can provide an increase in safety, quality and availability in industrial plants. The graph shown in Figure 3 illustrates that continuous investments in preventive maintenance reduce failure costs and, as a consequence, a decrease in the total maintenance cost, in which preventive maintenance costs are added to the failure costs. However, the graph also shows that, from the ideal point of investment in preventive maintenance, more investments bring few benefits to reduce the cost of failures and end up increasing the total cost, which is what the maintenance policy takes into account.

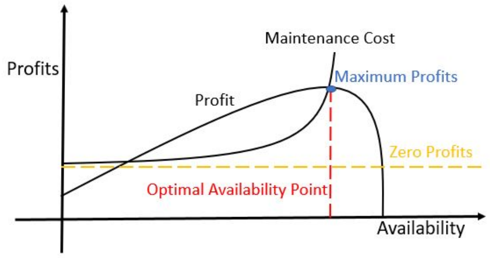


Figure 3: Graph of cost versus maintenance level [37]

However, the creation of a Predictive Maintenance Program is a strategic decision that, until now, lacks an analysis of the problems related to installation, management and control. According to Shin & Jun [38], when it is a high-value asset, the Operation and Maintenance (O&M) phase requires heavy charges and more effort than in the installation (construction) phase, as these assets have a useful life that any unexpected event of the asset during that period causes catastrophic damage to the industry.

3 Some specific approaches

3.1 Opportunistic Maintenance

In order to fulfill the maintenance objectives, the company needs management skills to integrate people, policies, equipment and practices. It also needs adequate engineering and technology [39]. Colledani, Magnanini & Tolio [40] state that Opportunistic Maintenance (OM) is an effective strategy to reduce interference between maintenance and production operations in multi-stage manufacturing systems and its application in the industry is still limited due to the complexity of predicting its impact on system performance. The maintenance of opportunities is theoretically adjusted automatically; if insufficient opportunities arise, the average delay increases and failures increase until a balance is reached, but there are minimal conditions for a given age renewal schedule and the natural balance may not be economically ideal [41]. Takahashi [42], thorough investigation of these opportunities and their occurrences, comes to next questions:

- When opportunities arise, which machines allow for other simultaneous repairs?

- What opportunities are needed, when do they arise, and for how long?

For Takahashi [42], Maintenance Management must consider some points to restructure a company and prepare it for future challenges, always with the participation of all employees:

1. Restrict investments in unnecessary equipment;
2. Make the most of existing equipment;
3. Improve the rate of use of equipment for production;
4. Guarantee the quality of the product, through the use of the equipment;
5. Reduce low-cost labour, improving equipment;
6. Reduce the cost of energy and material purchased, through innovations in equipment and improvements in methods of use.

Iung, Levrat & Thomas [43] show that the numerical results of the study presented in the paper, by properly defining the parameters of opportunistic maintenance actions, it is possible to obtain an effective synchronization of preventive maintenance and production operations, preserving the conditions of the machine and meeting the production goals. The superiority of Condition Based Maintenance remains uncertain in multicomponent systems, in which opportunistic maintenance strategies can be applied [44].

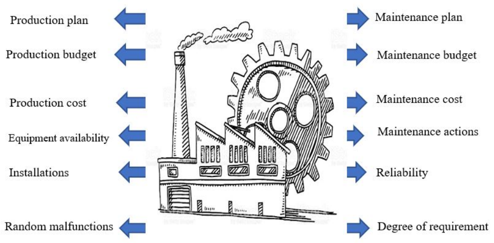


Figure 4: production / maintenance interface [20]

Compared to the static opportunistic maintenance strategy and the strategy without considering opportunistic maintenance, the total cost of maintenance and inventory of the dynamic opportunistic maintenance strategy shows a decline of 10.927% and 18.304%, respectively [45]. Vu *et al.* [46] mention that the structure of the system, the maintenance opportunity and its support, as well as the economic dependence, are important issues that must be considered when making maintenance decisions.

In CBM, the opportunistic zone is defined as (part of) the P-F interval (Figure 5) that is part of the degradation curve. It starts at the point where a potential failure can be detected (P) and ends at the moment when the failure occurs (F) [47].

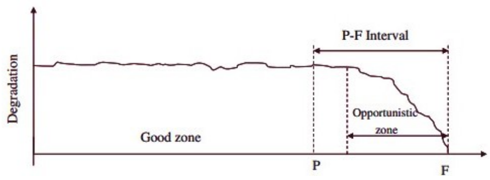


Figure 5: Opportunistic maintenance zone in CBM policy [44]

In case of opportunistic maintenance, the maintenance activities can begin at an arbitrary point within the opportunistic zone (which is equal to or less than the P-F interval). This opportunistic zone is the period during which the degradation started, without leading to a fatal shutdown of the component. Within the opportunistic zone, Planned Maintenance (PM) activities can be carried out against PM costs [44]. Rao [48] demonstrates that policies with various opportunistic maintenance ages for each increasing failure rate component are better with respect to policies with a single opportunistic maintenance age for each component. Performing preventive maintenance, even when there is no opportunity, can have a detrimental effect. It can be said that opportunistic maintenance policies of the type (n, ∞) are sufficient in the case of systems with a large number of components [48]. Zhou, Xi & Lee [49] studied a dynamic programming method in which decisions are based on a combination of OM cost savings and penalty costs and stated that an ideal maintenance practice is determined by maximizing the savings of cumulative short-term costs. Ding & Tian [50] proposed a method for making opportunistic maintenance decisions by comparing the age of a given component with a limit defined by some percentage of the Mean Time To Failure (MTTF). Dekker [1] developed a model to determine the ideal age for opportunistic maintenance when the opportunity follows the Homogeneous Process of Poisson (HPP) [51] with the adaptive opportunistic maintenance model, based on the forecast of the conditions of a railroad, demonstrating that the adaptive opportunistic maintenance strategy has a lower cost per unit of time than systematic preventive maintenance. Truong Ba *et al.* [52], in their results indicate that significant savings can be achieved considering OM. In addition, it is shown that the new consideration of partial opportunities significantly increases the benefit of OM.

Among preventive maintenance control policies, opportunistic maintenance is an effective strategy for reducing the impact of maintenance operations on multistage manufacturing systems [43]. Under the scope of Operations Management, one of the great challenges is to give them as occurrences of unexpected failures, which is why numerous studies have emerged with the main objective of avoiding these unexpected events such as Multivariate Analysis and Neural Networks, among others.

3.1.1 Multivariate Analysis

One of the main focuses of Multivariate Analysis is the search for models that best apply to a forecast or even explain the behaviour of the variables studied. Multivariate data consists of observations about several different variables for various individuals or objects. Such data immerse in all branches of science, such as Psychology, Education, Geology, Social Sciences, Engineering, Ergonomics, etc. The multivariate method has become an increasingly important area of statistics. In fact, the vast majority of the data is multivariate [53]. The multivariate model also aims at data reduction or structural simplification, classifying, and grouping, investigating the dependency among variables, predicting, and formulating hypotheses [54]. Simoglou, Martin & Morris [55] extended the existing methodologies to monitor dynamic processes; as it is multivariate, they considers the effect of exogenous inputs, providing additional monitoring methods and appropriate control limits when the serial correlation is present in the data system. There are several techniques for analysing multivariate data. Among them, factor analysis, multiple regression and multiple correlations, multiple discriminant analysis, multivariate analysis of variance and covariance, joint analysis, canonical correlation, cluster analysis, and scheduling. Barker & Newby [56] presented a study focused on the problem of determining the inspection and maintenance strategy of a system whose state is described by a multivariate stochastic process.

Chatfield & Collins [53] show that the general point that multivariate analyses tend to be concerned with finding relationships, not only among variables but also among individuals. Based on this, the opportunity arises to include computer technology as Internet of Things (IoT).There is also an extremely important tool, the multivariate control charts, which provides powerful methods to detect out of control situations and to diagnose causes; for example, Hotelling [57] presents a multivariate control plot procedure that is based on the most recent observation - it is insensitive to small and moderate changes in the mean vector. The difficulty of interpreting an out-of-control signal on a multivariate control chart was widely discussed by Chua and Montgomery [58], Alt [59], Doganaksoy *et al.* [60], Lowry *et al.* [61] and Haq [62]. The procedures of the multivariate control chart are often considered for use in cascade-type processes, such as those found in the process industries [63].

Multivariate models are a family of several models in which each one has its applicability depending on the behaviour of the data to be analysed. Wang [64] presents a design and an optimization of simulation-based multivariate Bayesian control graph for maintenance applications based on conditions. It combines the use of the concept of delay time and Bayesian theory to develop the posterior probability function of the underlying state, given the history of observed monitoring information.

3.1.2 Neural Networks

Within the scope of the development of Artificial Intelligence, there was a promising development, with a high degree of success, with regard to the applications of neural networks in the detection of defects in energy systems, as mentioned in [65], [66] and [67]. According to Lippmann [68], Artificial Neural Networks have been widely used for pattern recognition due to their ability to generalize and respond to unexpected patterns. The main strength of Neural Networks is the ability to recognize patterns in incomplete or "noisy" data. Regarding the forecast, this tool has presented satisfactory answers. Cadenas & Rivera [69] present a study of wind speed forecast in the region of La Venta, Oaxaca, Mexico, using various techniques of Artificial Neural Networks, with data resources collected, making possible to verify the importance of this tool in relation to the precision.

Jolliffe [70] mentions in his work that the most common application in Principal Component Analysis (PCA) is to reduce, with minimal loss of information, the dimensions of the data varieties. Generally, these data sets constitute a large set of correlated variables that are transformed into a new set of variables, called Principal Components. Not everything is 100% ideal. Crupi, Guglielmino & Milazzo [71] present one of the disadvantages of the neural network that requires training.

Bansal, Evans & Jones [72] exposed a predictive maintenance system in real-time for machine systems based on the neural network approach. The use of simulated data, generated by an experimentally validated simulation model, proved to be effective.

In the study of hydraulic pumps presented in [73], there was a preference by the choice of MultiLayer Perceptron (MLP) neural networks, with greater capacity to be suitable for the classification of patterns. The study provided by Ni & Wang [74] mentions the effectiveness of the prediction of the Multilayer Feed-forward Neural Networks (MFNNs); the Neural Network models showed a high precision in their results. Firat & Gungor [75] go in the same research line reaching the results, indicating that the General regression neural network can be successfully applied to predict the depth of cleaning around the pillars of the circular bridge.

AlGhazzawi & Lennox [76], in their study, exposed that static PCA techniques were not suitable for the development of a simple process monitoring system that would allow process operators to quickly and easily identify any sources of abnormalities in the process. Zaranezhad, Mahabadi & Dehghani [77], in a numerical analysis with respect to the result of neural networks, present a comparison of the precision model demonstrating that the perceptron neural networks had a prediction accuracy of 90.9% with a prediction accuracy rate of 96.19%; the neural-GA model obtained the highest forecast accuracy of 95.9% and an accuracy of 96.7%. In a forecasting study presented in [78], it is said that its quality of RNA prediction can be improved by expanding the training data sets and optimizing the construction of the network. Rao [79] also states that the results will depend on the collection of high-quality data.

Gilabert & Arnaiz [80] present a case study based on a Neural Network, where it is possible to find a predictive maintenance solution for non-critical machines. The results indicate the feasibility of partial solutions in monitoring and diagnosis.

Fu *et al.* [81] present a form of identification and diagnosis based on Artificial Neural Network (ANR) in the electro-hydraulic servomechanisms of a hydroelectric unit. Experimental tests show that the proposed strategy can guarantee optimal performance. According to the study presented in [82], predictive maintenance is already used or will be used by 83% of production companies; it has been a valuable application of the Internet of Things (IoT) mainly on the factory floor.

According to the CXP Group report, the Digital Industrial Revolution with Predictive Maintenance showed that the level of use of predictive maintenance is being used by 91% of manufacturers [83]; it can be verified that the reduction in the repair time of 93% of the companies pointed to the improvement of the old industrial infrastructure as the main objective of their predictive maintenance initiatives.

Javadpour & Knapp [84] present an implementation of a predictive neural network for use as an operator assistance fault diagnosis with high forecasting accuracy in an automated manufacturing environment. The network was able to correctly classify three different fault categories with a performance rate greater than 99% in standards and 100% in failures. As a result of implementing adequate tools, ans according to a PWC report [85], predictive maintenance in factories could:

- Reduce the cost by 12%;
- Improve uptime by 9%;
- Reduce safety, health, environment and quality risks by 14%;
- Increase the useful life of an old asset by 20

4 Proposed approaches

The previous sections show that maintenance policies play an essential role in the operations management. One of the big challenges of Operations Management science is to avoid several losses in the goods' manufacturing process, avoiding unwanted losses; most of these failures are classified as random but, through the reliability function and related tools, it is possible to exploit the maintenance capacity through its policies; for example, the predictive maintenance policy, supported by the math tools referred in the past section, can make predictions with results with a high degree of rigour – but, it must be emphasized that these results depends on the quality of the data available.

With these tools, the proposal is to improve the asset monitoring system using the portfolios provided by the asset management system; in this perspective, it is possible to automate the maintenance system using the concept that emerged in 1999, and whose main objective is to make equipment autonomous and intelligent enough, aiming they do not need human intervention: this concept is supported by the IoT [86]. Applying IoT in the industry, it becomes possible, for maintenance stops, to converge to an optimal reduction in intervention times and, consequently, in costs, which implies making the process more profitable.

The purpose of future developments is to explore the predictive maintenance policy in order to avoid bottleneck phenomena, making the production process more versatile, regarding to capacity readjustment.

Neural Networks will be used to support the classification of groups of variables in order to understand them, and the multivariate analysis applied to forecast, with the main focus on Machine Learning (ML). Since maintenance is an important branch in the production process, Big Data analysis, Machine Learning and the IoT, using real-time forecasting in the manufacturing equipment, will make possible to predict failures at run time, aiming to launch Working Orders before the fault happens [87]. For a first test, the technologies of Figure 6 will be applied.

5 Discussion

The present article brings an overview of the linked literature Optimizing production by maximizing the availability of assets, when it comes to optimizing

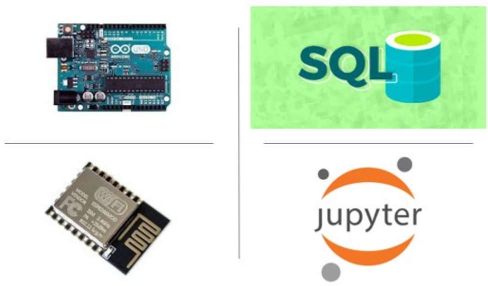


Figure 6: Support technologies.

production it is necessary to have accessibility than that the trailer behind, for this reason our study brings the reader a location of concepts that directly influences the study presented on this topic.

What can be seen in many studies in this area about a conflict between the investment limit and the return that a prediction maintenance system can bring to an entity? Regarding the prediction methods by multivariate models, there are very few, and there is a gap regarding the automation of prediction through this model. Regarding neural networks, there is a wide range of studies, this does not mean that it presents a significant advantage over multivariate analysis, each of which has particular advantages.

When it is possible to have a timely response to make decisions that avoid making bad decisions in any department of a company, the big challenge is to reduce the time to act in the face of events such as failures. Therefore, there are mathematical and computational tools that, together, can bring a great benefit, such as obtaining data through IoT technology and exploring them for possible event prediction. Finally, we will have an optimized process, as these failures can be controlled and resolved within a restricted time window. As the failure of a machine will depend on many variables to be extinguished, such as vibration, temperature and noise, among others; one variable may be related to another or not, which makes multivariate analysis an ideal tool for this study. Currently, some “software” and their respective languages show that many have limitations in terms of their programming.

Based on the tools presented in this paper, it is possible to avoid unwanted events; if we work with probabilities, hence the great strength of the predictive maintenance policy, that arises because it reduces the probability of equipment failures; its domain is covering the production system and the values associated with the production is the availability of the equipment that is related with its faults. Since the total cost is the sum of all costs related to production, this means that the reduction in one of them reflecting the reduction of the total cost. According to Yip, Fan & Chiang [88], the time series models General Regression Neural Network (GRNN) and the Box-Jenkins, usually describe the behaviour and predict the costs associated with maintaining different categories of equipment at an acceptable level of precision. Mateus, Margalho & Farinha [89] presented in their studies the disadvantages that the ARMA time series model presents in relation to the forecast when faced with oscillatory data (dummy variables).

During this research, it was possible to verify that the multivariate analyses showed promising results in the researches done, solving problems that were previously unthinkable because they had a solution or approaches to solutions of a great result. With this multivariate analysis, these solutions are a good part of the performance and versatility of the ”software” developed for these purposes. During the problem-solving process, it is necessary to go through several assumptions. According Renwick [90] Most benefits the predictive maintenance programs include not only evident cost benefits, such as reducing machinery downtime and production losses, but also the more subtle long-term cost benefits which can result from accurate maintenance scheduling.

The importance of Maintenance in a production system has been validated and a concern to respond in a timely manner to unwanted stops on a manufacturing line. To solve problems, knowing the variables in question must be one of the important factors, knowing the right time for intervention, without forget the life cycles of all parts that make up a complex system, with complex problems joining simple problems.

6 Conclusions

The paper presents a global approach about the maintenance in production results, and that it is possible to validate the use and efficiency of several tools in solving problems related to the increase maintenance efficiency. In the Maintenance space, some studies were applied to solutions in which it was possible to detect and diagnose the failure based on this approach, and also verifying how efficient the Neural Networks and the Principal Component Analysis are. The accumulated costs associated with factory failure have a high level of significance. For this reason, maintenance approaches have evolved to respond to these dilemmas, such as Predictive Maintenance (PM), which has always shown the ability to evolve to maintain integrity in companies, generating information about the conditions of the equipment; this data allows maintenance to be effective.

The real cost of implementing and maintaining a predictive maintenance program is not the initial cost of the system. Instead, it is the annual labour costs and indirect costs associated with the acquisition, storage, trends, and analysis of the data necessary to determine the operational condition of the plant's facilities. In this area, predictive maintenance systems present a greater variation in their capacity, automation in data acquisition and storage, etc.; this will reduce operating costs.

References

[1] R. Dekker, "Applications of maintenance optimization models: a review and analysis," vol. 51, no. 3, pp. 229–240, 1996-03. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0951832095000763

[2] A. Garg and S. Deshmukh, "Maintenance management: literature review and directions," vol. 12, no. 3, pp. 205–238, 2006-07. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/13552510610685075/full/html

[3] A. Sharma, G. Yadava, and S. Deshmukh, "A literature review and future perspectives on maintenance optimization," vol. 17, no. 1, pp. 5–25, 2011-01-01, publisher: Emerald Group Publishing Limited. [Online]. Available: https://doi.org/10.1108/13552511111116222

[4] A. Alrabghi and A. Tiwari, "State of the art in simulation-based optimisation for maintenance systems," vol. 82, pp. 167–182, 2015-04. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0360835214004513

[5] L. Wang, J. Chu, and W. Mao, "An optimum condition-based replacement and spare provisioning policy based on markov chains," vol. 14, no. 4, pp. 387–401, 2008-09-26. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/13552510810909984/full/html

[6] S. Duffuaa, A. Kolus, U. Al-Turki, and A. El-Khalifa, "An integrated model of production scheduling, maintenance and quality for a single machine," vol. 142, p. 106239, 2020-04. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0360835219307089

[7] S. Amari and L. McLaughlin, "Optimal design of a condition-based maintenance model," in *Annual Symposium Reliability and Maintainability, 2004 - RAMS*. IEEE, 2004, pp. 528–533. [Online]. Available: http://ieeexplore.ieee.org/document/1285501/

[8] C. Chu, J.-M. Proth, and P. Wolff, "Predictive maintenance: The one-unit replacement model," vol. 54, no. 3, pp. 285–295, 1998-05. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925527398000048

[9] J. Khalil, S. M. Saad, and N. Gindy, "An integrated cost optimisation maintenance model for industrial equipment," vol. 15, no. 1, pp. 106–118, 2009-03-27. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/13552510910943912/full/html

[10] A. Gill, "Optimisation of the technical object maintenance system taking account of risk analysis results," vol. 19, no. 3, pp. 420–431, 2017-06-12. [Online]. Available: http://www.ein.org.pl/sites/default/files/2017-03-13.pdf

[11] A. H. C. Tsang, A. K. S. Jardine, J. D. Campbell, and J. V. Picknell, *Reliability centred maintenance : a key to maintenance excellence*. Dept. of Manufacturing Engineering and Engineering Management, City University of Hong Kong, 2000, accepted: 2016-07-29T04:04:40Z Journal Abbreviation: Reliability-centered maintenance: towards excellence. [Online]. Available: http://ira.lib.polyu.edu.hk/handle/10397/54245

[12] K. K. Boyer and C. McDermott, "Strategic consensus in operations strategy," vol. 17, no. 3, pp. 289–305, 1999-03. [Online]. Available: http://doi.wiley.com/10.1016/S0272-6963(98)00042-4

[13] B. Dhillon, *Maintainability, Maintenance, and Reliability for Engineers*, 0th ed. CRC Press, 2006-03-27. [Online]. Available: https://www.taylorfrancis.com/books/9781420006780

[14] R. K. Mobley, *An Introduction to Predictive Maintenance(Plant Engineering)*, 2nd ed. Butterworth-Heinemann, 2002-10-24, google-Books-ID: SjqXzxpAzSQC.

[15] K. O. Cua, K. E. McKone, and R. G. Schroeder, "Relationships between implementation of TQM, JIT, and TPM and manufacturing performance," vol. 19, no. 6, pp. 675–694, 2001-11. [Online]. Available: http://doi.wiley.com/10.1016/S0272-6963(01)00066-3

[16] J. Wang, Y. Lu, Y. Yang, and T. Mao, "Thermodynamic performance analysis and optimization of a solar-assisted combined cooling, heating and power system," vol. 115, pp. 49–59, 2016-11-15. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544216312154

[17] L. G. Hernández-Pérez, C. Ramírez-Márquez, J. G. Segovia-Hernández, and J. M. Ponce-Ortega, "Simultaneous structural and operating optimization of process flowsheets combining process simulators and metaheuristic techniques: The case of solar-grade silicon process," p. 106946, 2020-05. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0098135419312402

[18] H. R. Rao and B. P. Lingaraj, "Expert systems in production and operations management: Classification and prospects," vol. 18, no. 6, pp. 80–91, 1988-12. [Online]. Available: http://pubsonline.informs.org/doi/abs/10.1287/inte.18.6.80

[19] M. Dorigo, G. D. Caro, and L. M. Gambardella, "Ant algorithms for discrete optimization," vol. 5, no. 2, pp. 137–172, 1999-04. [Online]. Available: http://www.mitpressjournals.org/doi/10.1162/106454699568728

[20] C. Blum and A. Roli, "Metaheuristics in combinatorial optimization: Overview and conceptual comparison," vol. 35, no. 3, pp. 268–308, 2003-09-01. [Online]. Available: http://portal.acm.org/citation.cfm?doid=937503.937505

[21] S. Alaswad and Y. Xiang, "A review on condition-based maintenance optimization models for stochastically deteriorating system," vol. 157, pp. 54–63, 2017-01. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0951832016303714

[22] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed. Springer Science & Business Media, 2006-06-06.

[23] H. Wang, "A survey of maintenance policies of deteriorating systems," vol. 139, no. 3, pp. 469–489, 2002-06. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0377221701001977

[24] C. S. Syan and G. Ramsoobag, "Maintenance applications of multi-criteria optimization: A review," vol. 190, p. 106520, 2019-10. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0951832018313759

[25] B. de Jonge and P. A. Scarf, "A review on maintenance optimization," p. S0377221719308045, 2019-09. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0377221719308045

[26] A. Bousdekis, N. Papageorgiou, B. Magoutas, D. Apostolou, and G. Mentzas, "A proactive event-driven decision model for joint equipment predictive maintenance and spare parts inventory optimization," vol. 59, pp. 184–189, 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2212827116309490

[27] B. Zhou, Y. Qi, and Y. Liu, "Proactive preventive maintenance policy for buffered serial production systems based on energy saving opportunistic windows," vol. 253, p. 119791, 2020-04-20. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095965261934661X

[28] Z. Zhao, F.-l. Wang, M.-x. Jia, and S. Wang, "Predictive maintenance policy based on process data," vol. 103, no. 2, pp. 137–143, 2010-10-15. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016974391000119X

[29] D. J. Edwards, G. D. Holt, and F. Harris, "Predictive maintenance techniques and their relevance to construction plant," vol. 4, no. 1, pp. 25–37, 1998-03. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/13552519810369057/full/html

[30] J. Vatn, P. Hokstad, and L. Bodsberg, "An overall model for maintenance optimization," vol. 51, no. 3, pp. 241–257, 1996-03. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0951832095000550

[31] F. Besnard, M. Patrikssont, A.-B. Strombergt, A. Wojciechowskit, and L. Bertling, "An optimization framework for opportunistic maintenance of offshore wind power system," in *2009 IEEE Bucharest PowerTech*, 2009-06, pp. 1–7.

[32] A. K. Pinto and J. Nascif, *Maintenance Strategic Function*, 2nd ed. Qualitymark, 2001.

[33] M. Zaeri, J. Shahrabi, M. Pariazar, and A. Morabbi, "A combined multivariate technique and multi criteria decision making to maintenance strategy selection," in *2007 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE, 2007-12, pp. 621–625. [Online]. Available: http://ieeexplore.ieee.org/document/4419264/

[34] J. Kershaw and B. Robertson, "Condition-based maintenance program increases production, reduces costs," pp. 34–36, 1985.

[35] Q. Chang, J. Ni, P. Bandyopadhyay, S. Biller, and G. Xiao, “Maintenance opportunity planning system,” vol. 129, no. 3, pp. 661–668, 2007-06-01, publisher: American Society of Mechanical Engineers Digital Collection. [Online]. Available: https://asmedigitalcollection.asme.org/manufacturingscience/article/129/3/661/462149/Maintenance-Opportunity-Planning-System

[36] M. Carnero, “Selection of diagnostic techniques and instrumentation in a predictive maintenance program. a case study,” vol. 38, no. 4, pp. 539–555, 2005-01. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167923603001283

[37] A. Murty and V. Naikan, “Availability and maintenance cost optimization of a production plant,” vol. 12, no. 2, pp. 28–35, 1995-03. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/02656719510080596/full/html

[38] J.-H. Shin and H.-B. Jun, “On condition based maintenance policy,” vol. 2, no. 2, pp. 119–127, 2015-04. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2288430014000141

[39] L. Pintelon and L. Gelders, “Maintenance management decision making,” vol. 58, no. 3, pp. 301–317, 1992-05. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/037722179290062E

[40] M. Colledani, M. C. Magnanini, and T. Tolio, “Impact of opportunistic maintenance on manufacturing system performance,” vol. 67, no. 1, pp. 499–502, 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0007850618301021

[41] D. J. Sherwin, “Age-based opportunity maintenance,” vol. 5, no. 3, pp. 221–235, 1999-01-01, publisher: MCB UP Ltd. [Online]. Available: https://doi.org/10.1108/13552519910282674

[42] Y. Takahashi, *TPM/MPT: Total Productive Maintenance*, 5th ed. IMAM, 2010.

[43] B. Iung, E. Levrat, and E. Thomas, “‘odds algorithm’-based opportunistic maintenance task execution for preserving product conditions,” vol. 56, no. 1, pp. 13–16, 2007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0007850607000054

[44] J. Koochaki, J. A. Bokhorst, H. Wortmann, and W. Klingenberg, “Condition based maintenance in the context of opportunistic maintenance,” vol. 50, no. 23, pp. 6918–6929, 2012-12. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00207543.2011.636924

[45] C. Zhang, W. Gao, T. Yang, and S. Guo, “Opportunistic maintenance strategy for wind turbines considering weather conditions and spare parts inventory management,” vol. 133, pp. 703–711, 2019-04. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0960148118312606

[46] H. C. Vu, P. Do, M. Fouladirad, and A. Grall, “Dynamic opportunistic maintenance planning for multi-component redundant systems with various types of opportunities,” vol. 198, p. 106854, 2020-06-01. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0951832019309421

[47] J. Moubray, *Reliability-Centered Maintenance*, 2nd ed. Butterworth-Heinemann, 1997.

[48] B. Rao, “Opportunistic maintenance of multi-equipment system: a case study,” p. 14, 2000.

[49] X. Zhou, L. Xi, and J. Lee, “Opportunistic preventive maintenance scheduling for a multi-unit series system based on dynamic programming,” vol. 118, no. 2, pp. 361–366, 2009-04-01. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925527308003472

[50] F. Ding and Z. Tian, “Opportunistic maintenance for wind farms considering multi-level imperfect maintenance thresholds,” vol. 45, pp. 175–182, 2012-09-01. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148112001802

[51] C. Letot, I. Soleimanmeigouni, A. Ahmadi, and P. Dehombreux, “An adaptive opportunistic maintenance model based on railway track condition prediction,” vol. 49, no. 28, pp. 120–125, 2016-01-01. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2405896316324454

[52] H. Truong Ba, M. E. Cholette, P. Borghesani, Y. Zhou, and L. Ma, “Opportunistic maintenance considering non-homogenous opportunity arrivals and stochastic opportunity durations,” vol. 160, pp. 151–161, 2017-04-01. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095183201631002X

[53] C. Chatfield and A. J. Collins, *Introduction to Multivariate Analysis*, a c r c press company ed. Northwestern University, 2000.

[54] R. A. Johnson and D. W. Wichern, *Applied multivariate statistical analysis*, 6th ed. Pearson Prentice Hall, 2007, OCLC: ocm70867129.

[55] A. Simoglou, E. Martin, and A. Morris, "Statistical performance monitoring of dynamic multivariate processes using state space modelling," vol. 26, no. 6, pp. 909–920, 2002-06. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0098135402000121

[56] C. Barker and M. Newby, "Optimal non-periodic inspection for a multivariate degradation model," vol. 94, no. 1, pp. 33–43, 2009-01. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0951832007001123

[57] H. Hotelling, "Multivariate quality control. techniques of statistical analysis," 1947.

[58] M.-K. Chua and D. C. Montgomery, "Investigation and characterization of a control scheme for multivariate quality control," vol. 8, no. 1, pp. 37–44, 1992, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qre.4680080107. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.4680080107

[59] F. B. Alt, "Multivariate quality control," vol. 6, pp. 110–122, 1985, publisher: John Wiley & Sons, Inc.

[60] N. Doganaksoy, F. W. Faltin, and W. T. Tucker, "Identification of out of control quality characteristics in a multivariate manufacturing environment," vol. 20, no. 9, pp. 2775–2790, 1991-01-01, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/03610929108830667. [Online]. Available: https://doi.org/10.1080/03610929108830667

[61] C. A. Lowry, W. H. Woodall, C. W. Champ, and S. E. Rigdon, "A multivariate exponentially weighted moving average control chart," vol. 34, no. 1, pp. 46–53, 1992-02-01, publisher: Taylor & Francis. [Online]. Available: https://amstat.tandfonline.com/doi/abs/10.1080/00401706.1992.10485232

[62] A. Haq and M. B. C. Khoo, "Memory-type multivariate control charts with auxiliary information for process mean," vol. 35, no. 1, pp. 192–203, 2019, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qre.2391. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.2391

[63] C. A. Lowry and D. C. Montgomery, "A review of multivariate control charts," vol. 27, no. 6, pp. 800–810, 1995-12. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/07408179508936797

[64] W. Wang, "A simulation-based multivariate bayesian control chart for real time condition-based maintenance of complex systems," vol. 218, no. 3, pp. 726–734, 2012-05. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0377221711010873

[65] T. Sidhu, H. Singh, and M. Sachdev, "Design, implementation and testing of an artificial neural network based fault direction discriminator for protecting transmission lines," vol. 10, no. 2, pp. 697–706, 1995-04. [Online]. Available: http://ieeexplore.ieee.org/document/400862/

[66] S. A. Kalogirou, "Applications of artificial neural-networks for energy systems," vol. 67, no. 1, pp. 17–35, 2000.

[67] R. Das and S. Kunsman, "A novel approach for ground fault detection," in *57th Annual Conference for Protective Relay Engineers, 2004*. IEEE, 2004, pp. 97–109. [Online]. Available: http://ieeexplore.ieee.org/document/1287088/

[68] R. P. M. Lippmann, "An introduction to computing with neural nets," vol. 4, pp. 4–22, 1987.

[69] E. Cadenas and W. Rivera, "Wind speed forecasting in the south coast of oaxaca, m□xico," vol. 32, no. 12, pp. 2116–2128, 2007-10. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0960148106002801

[70] I. T. Jolliffe, "Principal component analysis," pp. 150–166, 2002.

[71] V. Crupi, E. Guglielmino, and G. Milazzo, "Neural-network-based system for novel fault detection in rotating machinery," vol. 10, no. 8, pp. 1137–1150, 2004-08. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1077546304043543

[72] D. Bansal, D. J. Evans, and B. Jones, "A real-time predictive maintenance system for machine systems," vol. 44, no. 7, pp. 759–766, 2004-06. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0890695504000392

Appendix B

Data Analysis for Predictive Maintenance Using Time Series and Deep Learning Models – A Case Study in a Pulp Paper Industry

[1,2]Balduíno Mateus; [3,4]Mateus Mendes; [3,5]José Torres Farinha, [1,2]Alexandre Batista Martins [2]António Marques Cardoso
[1]EIGeS - Research Centre in Industrial Engineering, Management and Sustainability. Lusófona University,Campo Grande, 376, 1749-024 Lisboa-Portugal
[2]CISE – Electromechatronic Systems Research Centre, University of Beira Interior, Calçada Fonte do Lameiro, P–62001-001 Covilhã, Portugal
[3]Polytechnic of Coimbra – Polytechnic of Coimbra, ISEC, 3045-093 Coimbra, Portugal
[4]Institute of Systems and Robotics, University of Coimbra, 3004-531 Coimbra, Portugal
[5]CEMMPRE - Centre for Mechanical Engineering, Materials and Processes—CEMMPRE, 3030-788 Coimbra, Portugal

balduino.mateus@ubi.pt; mmendes@isr.uc.pt; torres-farinha@dem.uc.pt; p5922@ulusofona.pt; ajmarcar@gmail.com

Abstract. Predictive maintenance is fundamental for modern industries, in order to improve the physical assets availability, decision making and rationalize costs. That requires deployment of sensor networks, data storage and development of data treatment methods that can satisfy the quality required in the forecasting models. The present paper describes a case study where data collected in an industrial pulp paper press was pre-processed and used to predict future behavior, aiming to anticipate potential failures, optimize predictive maintenance and physical assets availability. The data were processed and analyzed, outliers identified and treated. Time series models were used to predict short-term future behavior. The results show that it is possible to predict future values up to ten days in advance with good accuracy.

Keywords: Data Analysis, Autoregressive Models, ARIMA, Deep Learning, Time Series Forecasting, Predictive Maintenance.

1. Introduction

Life cycle optimization has been a concern for decades; it becomes clear that a physical asset with an adequate maintenance will have a longer life with a greater return for the organization [1]. Monitoring industrial equipment is essential to anticipate and avoid potential failures, which can endanger people and assets. Sensors are deployed and data are collected to facilitate and automate the process. The methods applied to treat and analyse the data are relevant for improving the fault detection performance, predictions and decision making. Data cleaning is one of the key challenges [2], [3], so that excess data or wrong data can be removed out of the analysis process. Using the data collected and properly treated, machine learning models can be trained, parameters can be calculated and obtained, so that actions, decision making, control, supervision and planning can be implemented to optimize manufacturing plants production processes [4], [5].

One of the biggest challenges is the elimination of duplicate data and noise. Gong *et al*. (2017) propose a simple binary classifier to separate useful data from bad data with 99% accuracy[6]. Veit *et al*. (2017) propose an approach that consists of combining clean and noisy data, pre-training a network using a large noisy data set, and then fine-tune it with the clean data set [7]. Plutowski & White in 1993 use a multi-layer feedforward neural network architecture to find patterns of bad quality data in a dataset [8].

Sensor data recorded along the time can be processed using time series methods. According to Zhang (2016), the classical decomposition method of the time series is, for example, to decompose a seasonal time series into trend, seasonal, cyclical, and irregular components [9]. After the components are known, the data can then be used to adjust or train suitable machine learning models.

Time series prediction models such as Autoregressive Integrated Moving Average (ARIMA) models, as well as Artificial Neural Networks (ANN) are frequently used and compared, with mixed conclusions about the superiority in forecasting performance [9], [10]. Mateus *et al*.(2020) discuss the disadvantages that the Autoregressive Moving Average (ARMA) time series model presents to forecast when faced with oscillatory data (dummy variables) [11]. In the case of complex problems that have both linear and non-linear correlation structures, the combination of ARMA with ANN is an effective way to improve forecasting performance. Although ANN are essentially nonlinear models, they have a capacity of modelling linear processes as well [9].

Deep learning methods are capable of identifying the structure and patterns of data, such as non-linearity and problems of complexity in time series forecasting [10]. Backpropagation networks (BP) are good at solving a wide variety of problems, and are used in time series forecasting [12]. According Hecht-Nielsen (1989), the standard BP network using a subjective transfer function can learn any measurable function in a very precise manner when a sufficient number of hidden neurons are used to [13].

The paper is structured as follows: Section two gives an overview of prediction problems on maintenance and some problem reviews about the prediction model and some solutions; In the third section it is presented Data Characterization and Pre-processing; In the fourth section a case study, to evaluate and validate the forecasting models, is presented; Finally, the conclusions of the study are made.

2. Related Work

Machine learning methods are increasingly popular in predictive maintenance. Jimenez *et al*. (2020) showed that there exists potential in the development of predictive models for application in predictive maintenance [14]. Rodrigues et al. use neural networks and principal component analysis to assess diesel engine oil degradation and determine the optimal point for oil replacement [15].

Daniyan *et al*. (2020) combine ANN with a dynamic time series model in diagnosing failures, to optimize maintenance intervention time in industrial equipment [16]. Ayvaz & Alpay (2021) propose a method to improve maintenance planning to minimize unexpected stops, through the combined use of Ensemble Empirical Mode Decomposition and Long Short-Term Memory [17]. Huang *et al*. (2019) apply Long Short-Term Memory (LSTM) neural network approaches to forecast real production data, obtaining satisfactory results, superior to conventional models [18].

Using deep networks to carry out stock market forecasting, Nti *et al*. (2021) reach a fairly satisfactory result of forecasting. They concluded that the efficient fusion of information from different sample indicators offers greater precision than individual data [19]. Liu *et al*. (2021), using an elastic mesh algorithm and LSTM to calculate the remaining bearing life, demonstrate that this algorithm can achieve good stability in terms of problem prediction [20]. Still, Aydin & Guldamlasioglu (2017) used an LSTM network to predict the current situation of an engine - their model demonstrated good forecasting reliability [21].

3. Data Characterization and Pre-processing

3.1 Dataset and framework

Data used in the present work are the result of monitoring an industrial paper press system. Six sensors are monitoring the functioning of the press, with a sampling period of 1 minute. The variables monitored are: 1) Electrical Current Intensity; 2) Hydraulic Unit Oil Level; 3) VAT Pressure; 4) Rotation Velocity, 5) Temperature in the Hydraulic Unit; and 6) Torque.

The dataset contains the sensor readings from 1 February 2018 to 30 October 2020. There are 1,490,400 samples in the dataset. Fig. 1 shows a plot of the values of all variables in the original dataset. This dataset was loaded in python and processed, using ScyPy libraries.

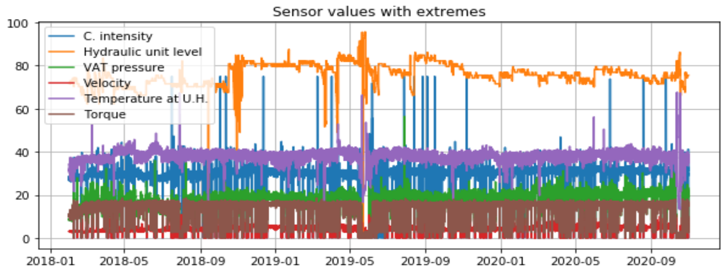


Fig. 1 Plot of the original dataset values. The variables are: C. intensity, Hydraulic unit level, VAT pressure, velocity, Temperature u. l., and Torque.

3.2. Data characterization and identification of discrepant data

As Fig. 1 shows, there are some sensor readings which show extreme levels. The very large values may be reading errors or overload moments. The very low values may be when the press was stopped, malfunctioning, underused, or they may also be reading errors. Those extremely low or extremely high values provide information about abnormal functioning of the press. They may negatively impact the performance of the forecasting algorithms.

Table 1 shows some statistical values of the data, namely the mean, standard deviation, minimum and maximum values: Fig. 2 shows histograms of the variables' quartiles.

Table 1. Statistical parameters of the variables: C. intensity, Hydraulic unit level, VAT pressure, velocity, Temperature u. l., and Torque.

	C. intensity	Hydraulic	Torque	VAT	Velocity	Temperature
mean	30.26	75.90	15.28	18.25	4.59	38.22
std	1.32	4.54	0.69	2.67	0.977	1.62
min	26.34	62.93	13.59	9.67	1.27	33.19
25%	29.30	72.86	14.90	17.13	3.92	37.17
50%	30.46	75.53	15.43	18.72	4.57	38.33
75%	31.28	79.52	15.78	19.97	5.28	39.35
max	34.26	88.97	17.09	26.17	7.87	43.10

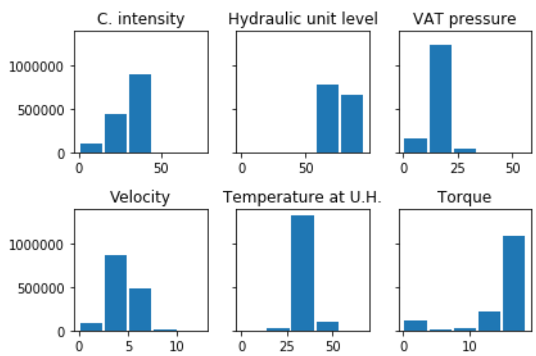


Fig. 2 Histogram of variables showing the number of samples per quartile.

Fig. 3 shows the amplitude of each sample concerning the lower and upper bounds for each variable. As the figure shows, the distribution of data is skewed for all variables.

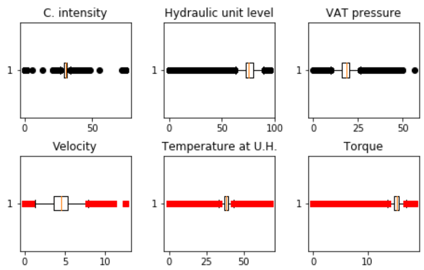


Fig. 3 Distribution of data points of all the sensors, with Low and High extremes.

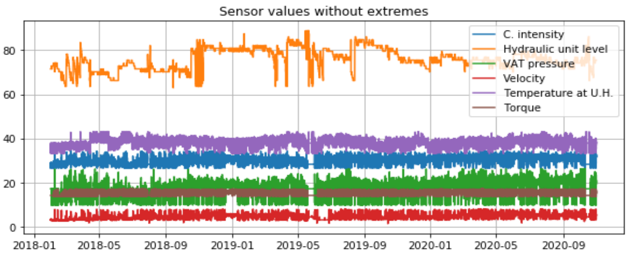


Fig. 4 Plot of the dataset variables without extreme values: Current intensity, Hydraulic unit level, VAT pressure, Rotation velocity, Temperature in the Hydraulic Unit, and Torque.

In order to achieve best performance in training predictive machine learning models, discrepant data (Ning & You, 2017) must be identified and possibly removed. The method used was the quartile approach, as explained in Formulae (1) to (5). In the formulae, $Q_{\frac{1}{4}}$ is the first quartile, $Q_{\frac{3}{4}}$ is the end quartile, n is sample number and IQ Inter-quartile Range.

$$Q_{\frac{1}{4}}=\frac{1}{4}\,(n+1) \tag{1}$$

$$Q_{\frac{3}{4}}=\frac{3}{4}\,(n+1) \tag{2}$$

$$IQ=Q_{\frac{1}{4}} - Q_{\frac{3}{4}} \tag{3}$$

$$Down_{limit}=Q_{\frac{1}{4}}-k.IQ \tag{4}$$

$$Up_{limit}=Q_{\frac{3}{4}}+k.IQ \tag{5}$$

$Down_{limit}$ is the lower bound limit accepted for the variable, calculated by subtracting of the constant k multiplied IQ to $Q_{\frac{1}{4}}$. Up_{limit} is the upper bound limit accepted for the variable, calculated by adding the constant k multiplied IQ to $Q_{\frac{3}{4}}$, where k is the variation constant of the limits.

After the application of the quartile method described above, the discrepant samples are taken out of the dataset. Namely, samples which are not in the interval $Down_{limit}$ were removed.

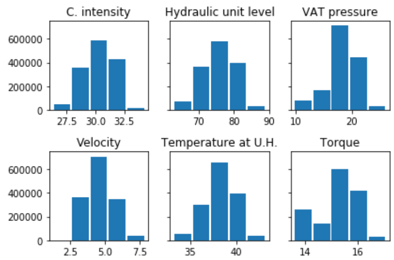


Fig. 5 Histogram of variables after removing discrepant data. The variables are Current intensity, Hydraulic unit level, VAT pressure, Rotation velocity, Temperature in the Hydraulic Unit, and Torque.

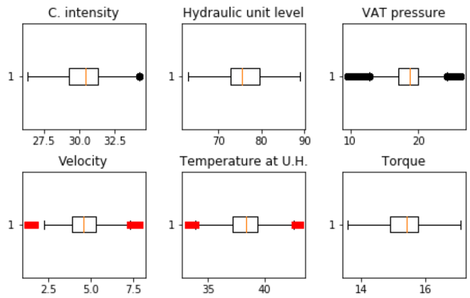


Fig. 6 Distribution of samples for all the sensors after removal of discrepant data. The variables are: Current intensity, Hydraulic unit level, VAT pressure, Rotation velocity, Temperature in the Hydraulic Unit, and Torque.

3.3 Study of correlations

Correlations between variables, as well as autocorrelations, are very important to have a better insight into the dependence of variables and determine which data models can be applied with higher probability of success. shows the matrix of correlations between variables. It is possible to verify some strong correlations between the variables

Current intensity and Velocity, among others that are presented in the graphs below. Most of the correlations, however, are weak.

Fig. 7, shows the autocorrelation of each variable. As the figure shows, the autocorrelations decay very quickly to less than 0.5. The charts show, therefore, that the correlation and autocorrelation of variables are very weak.

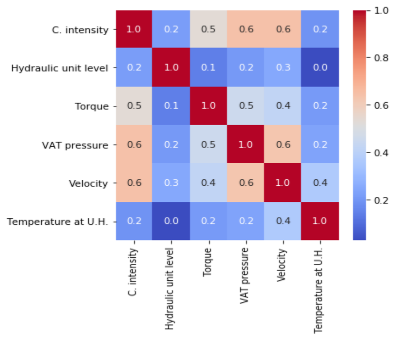


Fig. 7 Correlations among all variables.

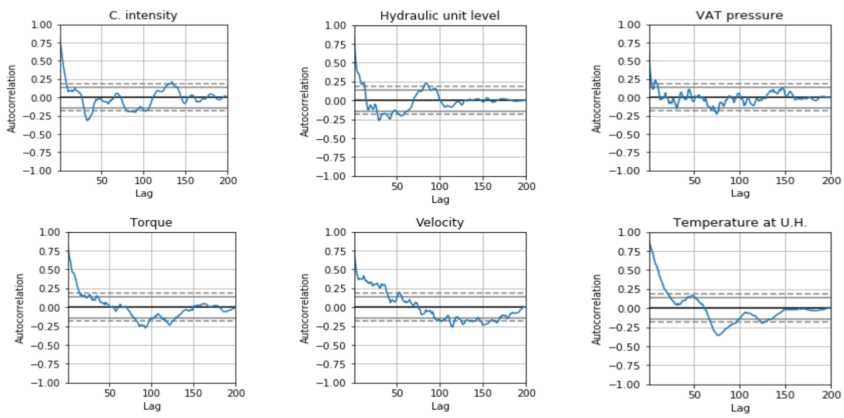


Fig. 8 Autocorrelation between samples of all variables, calculated for 200 days.

In Autocorrelation there is a decay in the period making the correlation increasingly lower. The graph in Fig. 8, shows only a correlation of up to 200 samples that also served for the test and for the forecast. Since the number of the samples is very high (1,490,400), there was a need to down sample the dataset, from a period of minutes to a period of days, in order to have a forecast in days. That was done by averaging the samples of each day using the python pandas function "df.resample ('D'). Mean ()".

4. Modelling using time series

4.1 Autoregressive model

As a first approach to predict future behaviour, an autoregressive model was applied. Autoregressive models are adequate to model variables that depend mostly on their previous behavior and a stochastic value, thus satisfying the following equation:

$$AR = \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \emptyset_p X_{t-p} + \in_t \tag{6}$$

Where $\emptyset_1, \ldots, \emptyset_p$ are real parameters and $\in_t$ is a white noise process independent and identically distributed.

4.2 ARIMA and SARIMA models

Some time series present a seasonal periodic component. A seasonal autoregressive model is characterized by the existence of a significant correlation between observations spaced by a multiple time interval [22].

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is a general case of the models proposed by [23] Box And Jenkins at 1976, for the adjustment of stationary time series. However, when there is a seasonal component in the data, the model class is called SARIMA (p, d, q) (P, D, Q), given by:

$$MA=-\theta_1 \in_{t-1} - \theta_2 \in_{t-2} - \ldots - \theta_q \in_{t-q} \tag{7}$$

Where $\theta_1, \ldots, \theta_p$, are parameters of an order of structures, $\in_t$ is white noise with zero mean.

$$AR_s = \Phi_1 X_{t-1s} + \Phi_2 X_{t-2s} +, \ldots, + \Phi_p X_{t-P} \tag{8}$$

$$MA_s = -\Theta_1 \in_{t-1} - \Theta_2 \in_{t-2} - \Theta_Q \in_{t-Q} \tag{9}$$

$$\nabla^D \Delta^d X_t = AR - MA + AR_s + MA_s \tag{10}$$

Where (p, d, q) refer to the model orders of the seasonal part: p is trend autoregression order, d is trend difference order and q is trend moving average order. (P, D, Q) is the same but with the Seasonal component. The parameters $\Phi_1, \ldots, \Phi_p$, are the parameters referring to the seasonal autoregressive part and $\Theta_1, \ldots \Theta_Q$, are the parameters of moving averages, and i is an error that cannot be estimated from the model and D indicates the number of seasonal differences made in the series to park it. The calculation of the parameters of the models that best fits was made using the most frequent Akaike Information Criterion (AIC), which is defined by:

$$AIC= 2\,log(L.k)+2(k) \tag{11}$$

where *L.k* is the maximized log-likelihood and *k* is the number of parameters in the model.

4. Experiments and Results

4.1 Results of the Autoregressive model

The model was applied with a 20-day sliding window, thus corresponding to 1440 x 10 = 14400 data samples and a forecast window with the same size, thus predicting the values for the next 10 days.

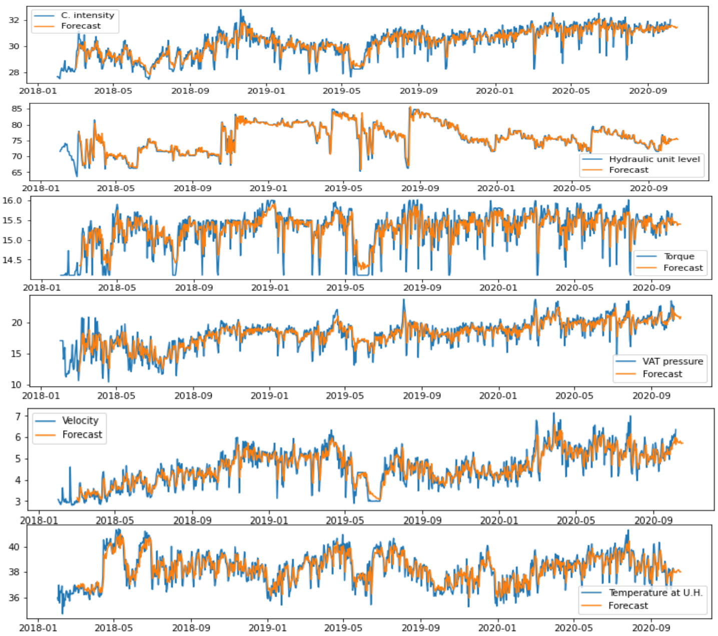


Fig. 9 Prediction of the six variables using a retrogressive method, with 20 days lag and 10 days predict window.

After eliminating the discrepant data samples, some irregularities in the samples, which may be momentary or prolonged damage, are still visible in **Fig.** 9. Nonetheless, the autoregressive model shows a good fit to all variables. The prediction errors between the forecasted values and the actual values are given in Table 2.

4.2 Results of the SARIMA model

The AIC was used to define the hyperparameters p, d, q, P, D and Q of the SARIMA model. The seasonal period was fixed at 12 for all-time series.

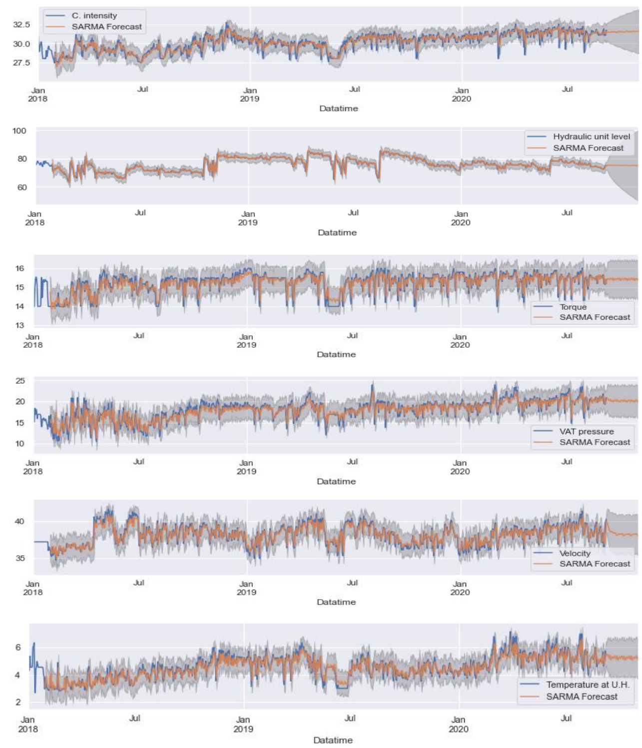


Fig. 10 Prediction of variables using the SARIMA (0, 1, 2) (1, 1, 2) method, with 20 days lag and 10 days predict window.

shows that the SARIMA model gives stable predictions, when using 20 days sliding window and parameters SARIMA (0, 1, 2) (1, 1, 2).

Table 2 shows the results of the forecasting errors in the period referring to the two models, AR and SARIMA. The table shows the Mean Average Error (MAE), the Mean Squared Error (MSE) and the Mean Average Percent Error (MAPE).

Table 2. Summary of the MAE, MSE and MAPE errors for the autoregressive and SARIMA models tested.

		C. intensity	Hydraulic unit level	Torque	VAT pressure	Velocity	Temp. at U.H.
AR	MAE	3.60	3.06	0.93	2.97	1.15	0.36
	MSE	13.08	18.37	1.15	10.54	1.47	0.24
	MAPE	12.92	3.76	6.57	16.95	25.17	0.97
SARIMA	MAE	0.21	1.05	0.15	0.57	0.25	0.64
	MSE	0.07	1.32	0.03	0.59	0.11	0.66
	MAPE	0.68	1.40	0.95	2.83	4.50	1.71

5. Discussion

Using the two models, (Auto-regressive and SARIMA), it was possible to verify that both offer acceptable prediction errors, with the data evaluated. The SARIMA model shows better performance than the AR model, what is expectable since it encompasses the three different components (autoregressive, moving averages and seasonal component). However, that implies a cost of an additional processing time. The SARIMA model takes approximately 15 times more computing time than the AR model. As the SARIMA model, its processing lasted 40 seconds and for the AR model 4 seconds.

For the regressive model prediction, there were no hyperparemeters to optimize. However, to find the best model, it was necessary to evaluate several models and to choose the parameters that best fit the data, using the AIC information criterion. It can be concluded that there is a good capacity of these models to predict based on data that presents a moderate variation.

For short-term forecasting, the models are satisfactory, emphasizing the need to clean the discrepant data. According to new studies in this area, they show superiority in the growth of the use of Neural Networks for those objectives, namely Recurrent Neural Networks that have greater long-term and short-term forecasting efficiency due to their Long-Short Term Memory capacity [19], [20] and [26]. That is planned as future work in the present project, where deep neural models will be designed and optimized for prediction.

5. Conclusion

Sensor data is fundamental to monitor industrial equipment and processes. The present paper describes a case study where six variables were sampled during almost three years, with a period of one minute. The data were selected and cleaned of discrepant data samples, analysed and used to forecast future behaviour with time series models, namely Autoregressive and SARIMA. Data processing and experiments were carried out in Python using ScyPy libraries. The SARIMA model showed smaller errors in the test set, so it is more adequate for the data analysed.

Future work includes experiments with Neural Networks and larger forecast range predictions.

Acknowledgements: The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowvska-Curie grant agreement 871284 project SSHARE and the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under Project POCI-01-0145-FEDER-029494, and by National Funds through the FCT—Portuguese Foundation for Science and Technology, under Projects PTDC/EEI-EEE/29494/2017, UIDB/04131/2020, and UIDP/04131/2020.

References

[1] E. Pais, J. T. Farinha, A. J. M. Cardoso, and H. Raposo, 'Optimizing the Life Cycle of Physical Assets – a Review', *WSEAS Trans. Syst. CONTROL*, vol. 15, pp. 417–430, Sep. 2020, doi: 10.37394/23203.2020.15.42.

[2] S. Chaudhuri and U. Dayal, 'An overview of data warehousing and OLAP technology', *ACM SIGMOD Rec.*, vol. 26, no. 1, pp. 65–74, Mar. 1997, doi: 10.1145/248603.248616.

[3] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, 'The KDD process for extracting useful knowledge from volumes of data', *Commun. ACM*, vol. 39, no. 11, pp. 27–34, Nov. 1996, doi: 10.1145/240455.240464.

[4] L. O. Prado, P. F. Ribeiro, C. A. Duque, and S. H. E. Abdel Aleem, 'Chapter 19 - Modeling and processing of smart grids big data: study case of a university research building', in *Decision Making Applications in Modern Power Systems*, S. H. E. Abdel Aleem, A. Y. Abdelaziz, A. F. Zobaa, and R. Bansal, Eds. Academic Press, 2020, pp. 507–538.

[5] A. B. Martins, J. Torres Farinha, and A. Marques Cardoso, 'Calibration and Certification of Industrial Sensors – a Global Review', *WSEAS Trans. Syst. CONTROL*, vol. 15, pp. 394–416, Sep. 2020, doi: 10.37394/23203.2020.15.41.

[6] Z. Gong, W. Wang, and W.-S. Ku, 'Adversarial and Clean Data Are Not Twins', *ArXiv170404960 Cs*, Apr. 2017, Accessed: Mar. 02, 2021. [Online]. Available: http://arxiv.org/abs/1704.04960.

[7] A. Veit, N. Alldrin, G. Chechik, I. Krasin, A. Gupta, and S. Belongie, 'Learning From Noisy Large-Scale Datasets With Minimal Supervision', 2017, pp. 839–847, Accessed: Mar. 02, 2021. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/html/Veit_Learning_From_Noisy_CVPR_2017_paper.html.

[8] M. Plutowski and H. White, ‘Selecting concise training sets from clean data’, *IEEE Trans. Neural Netw.*, vol. 4, no. 2, pp. 305–318, Mar. 1993, doi: 10.1109/72.207618.

[9] Z. Zhang, ‘Neural networks: further insights into error function, generalized weights and others’, *Ann. Transl. Med.*, vol. 4, no. 16, Aug. 2016, doi: 10.21037/atm.2016.05.37.

[10] S. Siami-Namini and A. S. Namin, ‘Forecasting Economics and Financial Time Series: ARIMA vs. LSTM’, *ArXiv180306386 Cs Q-Fin Stat*, Mar. 2018, Accessed: Mar. 09, 2021. [Online]. Available: http://arxiv.org/abs/1803.06386.

[11] B. Mateus, J. T. Farinha, and A. M. Cardoso, ‘Production Optimization versus Asset Availability – a Review’, vol. 15, p. 13, 2020, doi: DOI: 10.37394/23203.2020.15.33.

[12] I. Kaastra and M. Boyd, ‘Designing a neural network for forecasting financial and economic time series’, *Neurocomputing*, vol. 10, no. 3, pp. 215–236, Apr. 1996, doi: 10.1016/0925-2312(95)00039-9.

[13] R. Hecht-Nielsen, ‘Neurocomputer Applications’, in *Neural Computers*, Berlin, Heidelberg, 1989, pp. 445–453, doi: 10.1007/978-3-642-83740-1_45.

[14] V. J. Jimenez, N. Bouhmala, and A. H. Gausdal, ‘Developing a predictive maintenance model for vessel machinery’, *J. Ocean Eng. Sci.*, vol. 5, no. 4, pp. 358–386, Dec. 2020, doi: 10.1016/j.joes.2020.03.003.

[15] J. Rodrigues, I. Cost, J. T. Farinha, M. Mendes, and L. Margalho, ‘Predicting motor oil condition using artificial neural networks and principal component analysis’, *Eksploat. Niezawodn. - Maint. Reliab.*, vol. 22, no. 3, pp. 440–448, Jun. 2020, doi: 10.17531/ein.2020.3.6.

[16] I. Daniyan, K. Mpofu, M. Oyesola, B. Ramatsetse, and A. Adeodu, ‘Artificial intelligence for predictive maintenance in the railcar learning factories’, *Procedia Manuf.*, vol. 45, pp. 13–18, Jan. 2020, doi: 10.1016/j.promfg.2020.04.032.

[17] S. Ayvaz and K. Alpay, ‘Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time’, *Expert Syst. Appl.*, vol. 173, p. 114598, Jul. 2021, doi: 10.1016/j.eswa.2021.114598.

[18] X. Huang, C. Zanni-Merk, and B. Crémilleux, ‘Enhancing Deep Learning with Semantics: an application to manufacturing time series analysis’, *Procedia Comput. Sci.*, vol. 159, pp. 437–446, Jan. 2019, doi: 10.1016/j.procs.2019.09.198.

[19] I. K. Nti, A. F. Adekoya, and B. A. Weyori, ‘A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction’, *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-020-00400-y.

[20] M.-D. Liu, L. Ding, and Y.-L. Bai, ‘Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction’, *Energy Convers. Manag.*, vol. 233, p. 113917, Apr. 2021, doi: 10.1016/j.enconman.2021.113917.

[21] O. Aydin and S. Guldamlasioglu, ‘Using LSTM networks to predict engine condition on large scale data processing framework’, 2017, pp. 281–285, doi: 10.1109/ICEEE2.2017.7935834.

[22] I. Khandelwal, R. Adhikari, and G. Verma, 'Time Series Forecasting Using Hybrid ARIMA and ANN Models Based on DWT Decomposition', *Procedia Comput. Sci.*, vol. 48, pp. 173–179, Jan. 2015, doi: 10.1016/j.procs.2015.04.167.

[23] H. Yip, H. Fan, and Y. Chiang, 'Predicting the maintenance cost of construction equipment: Comparison between general regression neural network and Box–Jenkins time series models', *Autom. Constr.*, vol. 38, pp. 30–38, Mar. 2014, doi: 10.1016/j.autcon.2013.10.024.

[24] M. Kashefi and M. R. Daliri, 'A stack LSTM structure for decoding continuous force from local field potential signal of primary motor cortex (M1)', *BMC Bioinformatics*, vol. 22, no. 1, 2021, doi: 10.1186/s12859-020-03953-0.

[25] C. Sun, S. Hong, M. Song, H. Li, and Z. Wang, 'Predicting COVID-19 disease progression and patient outcomes based on temporal deep learning', *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, 2021, doi: 10.1186/s12911-020-01359-9.

[26] Z. Gui *et al.*, 'LSI-LSTM: An attention-aware LSTM for real-time driving destination prediction by considering location semantics and location importance of trajectory points', *Neurocomputing*, vol. 440, pp. 72–88, 2021, doi: 10.1016/j.neucom.2021.01.067.

Appendix C

applied sciences

MDPI

Article

Anticipating Future Behavior of an Industrial Press Using LSTM Networks

Balduíno César Mateus [1,2,*], Mateus Mendes [3,4,*], José Torres Farinha [3,5] and António Marques Cardoso [2]

1 EIGeS—Research Centre in Industrial Engineering, Management and Sustainability, Lusófona University, Campo Grande, 376, 1749-024 Lisboa, Portugal
2 CISE—Electromechatronic Systems Research Centre, University of Beira Interior, Calçada Fonte do Lameiro, P–62001-001 Covilhã, Portugal; ajmc@ubi.pt
3 Polytechnic of Coimbra, ISEC, 3045-093 Coimbra, Portugal; tfarinha@isec.pt
4 Institute of Systems and Robotics, University of Coimbra, 3004-531 Coimbra, Portugal
5 Centre for Mechanical Engineering, Materials and Processes—CEMMPRE, 3030-788 Coimbra, Portugal
* Correspondence: balduino.mateus@ubi.pt (B.C.M.); mmendes@isec.pt (M.M.)

Abstract: Predictive maintenance is very important in industrial plants to support decisions aiming to maximize maintenance investments and equipment's availability. This paper presents predictive models based on long short-term memory neural networks, applied to a dataset of sensor readings. The aim is to forecast future equipment statuses based on data from an industrial paper press. The datasets contain data from a three-year period. Data are pre-processed and the neural networks are optimized to minimize prediction errors. The results show that it is possible to predict future behavior up to one month in advance with reasonable confidence. Based on these results, it is possible to anticipate and optimize maintenance decisions, as well as continue research to improve the reliability of the model.

Keywords: time series prediction; LSTM prediction; deep learning prediction; predictive maintenance

check for updates

Citation: Mateus, B.C.; Mendes, M.; Farinha, J.T.; Cardoso, A.M. Anticipating Future Behavior of an Industrial Press Using LSTM Networks. *Appl. Sci.* **2021**, *11*, 6101. https://doi.org/10.3390/app11136101

Academic Editors: Marlene Amorim, Yuval Cohen and João Reis

Received: 30 April 2021
Accepted: 25 June 2021
Published: 30 June 2021

1. Introduction

Modern processors, computers and high speed networks make it possible to acquire, transfer and store large quantities of data in real time. Acquisition and combination of data from different sensors makes it possible to gain an insightful view of the state of factories, industrial plants and other facilities. Large datasets can be constructed, stored and processed using information technologies such as Big Data, cloud computing, cutting-edge computing, and artificial intelligence tools. The Internet of Things (IoT) is a recent concept, which provides many benefits to different areas, such as maintenance and production management, because it facilitates the automation of tasks such as monitoring and maintenance. This results in the popularization of intelligent systems, which are highly dependent on Big Data [1] and are an important area of study, since they offer the tools and methods to acquire and process large volumes of data such as historical production processes, including many production and operating parameters.

Modern time-series and other data analysis techniques have been used with success for different tasks, such as freeway traffic analysis [2] and additive manufacturing [3]. Different approaches have also been proposed in the field of predictive maintenance [4,5]. Satisfactory results were obtained using Big Data records as support for PCA models, which resulted in a warning alarm several days before a potential failure happened [6].

Life cycle optimization has been an important concern for decades. A physical asset with proper maintenance will have a longer useful life with a greater return on investment for the organization [7].

Predictive maintenance requires good quality data. The information that is extracted from the online or offline data must be reliable, and so the results must be good enough to justify the investment in data collection and analysis. The process starts from the correct calibration of the reading sensors and equipment [8]. The data are then stored and processed using different models, such as Principal Component Analysis (PCA) and Neural Networks [9]. Maintenance planning involves the use of several algorithms, the most common being time series [10].

Maintenance of equipment in the industry becomes a sensitive and important point that affects the equipment's operating time and efficiency [5]. This makes maintenance one of the strategic points for the development and growth of competitiveness vis-à-vis competitors. Chen and Tseng studied the total expected cost of maintaining a flotation system, including the cost of lost production, the cost of repairs, and the cost of standby machines [11].

Daniyan et al. propose the integration of Artificial Intelligence (AI) systems, which will bring many benefits in diagnosing condition problems of industrial machines [12]. They highlight the viability of AI that combines the use of Artificial Neural Networks (ANNs) with a dynamic time series model, for fault diagnostics, to optimize the equipment intervention time.

Hsu et al. demonstrated that neural networks can be a great technology in the support and decision making of large and small companies [13]. There is a trend to use those tools in predictive maintenance systems with the aim of making the prediction systems more intelligent [14].

According to Jimenez et al., there is a great effort in the development of predictive models for application in predictive maintenance [15]. Ayvaz and Alpay apply Long Short-Term Memory (LSTM) neural network approaches to predict real production data, obtaining satisfactory results, superior to conventional models [16]. In their study to improve maintenance planning to minimize unexpected stops, they apply a new method that consists of the combined use of decomposition in empirical mode of ensemble and long-term memory. Their results showed a performance superior to other state of the art models.

LSTM networks use several ports with different functions to control neurons and to store information. The LSTM cell can retain important information for a longer period in which it is used. This property of information maintenance allows the LSTM to exhibit a good performance in the classification, processing, or forecasting of a complex dynamic sequences [17].

The present work uses different LSTM models to predict future trends of six variables, on a dataset containing three years of data samples grabbed in an industrial press, which aims to operate continuously with minimum downtime. Different data pre-processing techniques, network architectures and hyperparameters were tested in order to determine the models that best fit the data and provide the lowest prediction errors.

Section 2 contains a summary of related work. Section 3 describes the theory of the LSTM networks. Section 4 describes the methods used for the present work. Section 5 describes the results and validation of the predictive model. Section 7 draws some conclusions and suggestions for future work.

2. Related Work

2.1. Predictive Maintenance

In smart industries, predictive maintenance is one of the most used techniques to improve condition monitoring, as it allows one to evaluate the conditions of specific equipment in order to predict problems before failure [18]. For good performance of predictive models, it is important that the sensor data collected are of good quality. Deep neural models have been used with success to improve prediction for condition monitoring of industrial equipment.

Wang et al. [19] use a model of long short-term recurrent neural networks (LSTM-RNN) with the objective of predictive maintenance based on past data. The main objective of predictive maintenance is to make an accurate estimate of a system's Remaining Useful Life (RUL). Traditional systems are only able to warn the user when it is too late and the failure occurs, causing an unpredictable offline period during which the system cannot operate properly with a consequent waste of time and resources [20].

In order to assess the condition of a system, the predictive maintenance approach employs sensors of different kinds. Some examples are temperature, vibration, velocity or noise sensors, which are attached to the main components whose failure would compromise the entire operation of the system. In this sense, predictive maintenance analyzes the history of a system in terms of the measurements collected by the sensors that are distributed among the components, with the objective of extracting a “failure pattern” that can be exploited to plan an optimal maintenance strategy and thus reducing offline periods [21]. In a case related to the steel industry, Ref. [22] used neural networks for classification of maintenance activities, so that interventions are planned according to the actual status of the machine and not in advance. Using multiple neural networks to identify status and RUL at a higher resolution can be very difficult, as the system can predict failure classifications and may not be able to recognize neighboring states. One limitation arises from the need for maintenance records to label datasets and the need for large amounts of data of adequate quality with maintenance events, such as component failures.

When systems start to be very complex or the number of sensor measurements to manage is very large, it can be difficult to estimate a failure. For this reason, in recent years, machine learning techniques are used more and more to predict working conditions of a component. Mathew et al. [23] propose several approaches to machine learning such as support vector machines (SVMs), decision trees (DTs), Random Forests (RFs), and others that show which technique has the best performance in RUL forecast for turbofan engines.

A major challenge in operations management is related to predicting machine speed, which can be used to dynamically adjust production processes based on different system conditions, optimize production performance and minimize energy consumption [24]. Essien and Giannetti [25] use a deep convolutional LSTM encoder–decoder architecture model on real data, obtained from a metal packaging factory. They show that it is possible to perform combinations of LSTM with other networks to significantly improve the results.

2.2. Prediction with LSTM Models

LSTM neural networks achieved the best performance in a number of computational sequence labeling tasks, including speech recognition and machine translation [26]. There are a variety of engineering problems that can be solved using predictive neural models. Beshrand Zarzoura used neural network models to predict problems of suspended road bridge structures based on global navigation satellite system observations [27]. Sak et al. demonstrated that the proposed LSTM architectures exhibit better performances compared to deep neural networks (DNNs) in a large vocabulary speech recognition task with a large number of output states [28]. Chen et al. adopted LSTMs for predicting the failure of heavy truck air compressors [29]. They concluded that the use of LSTMs leads to more consistency in predictions over time compared to models that ignore history, such as random forest models.

Gosh et al. [30] presented an extension that they called Contextual LSTM (CLSTM). This model was also used for the forecasting of pollutants. There is also the proposal for a genetic long short-term memory (GLSTM), which has been used in the study of wind energy forecasting [31]. Guo et al. presented a combination method based on real-time prediction errors in which the support vector regression (SVR) and LSTM outputs are combined in the final results of the model’s prediction, thus obtaining results of greater precision [32].

Ren et al. used a combination of a Convolution Neural Networks (CNNs) and LSTM in order to extract more in-depth information from data to predict the useful life of ion batteries [33]. Niu et al. used an LSTM and developed an effective speed prediction model to solve prediction problems over time [34]. Feng et al. report that the LSTM algorithm is superior and, according to them, it performs better than conventional neural network models [35].

The architecture of an LSTM network includes the number of hidden layers and the number of delay units, which is the number of previous data points that are considered for training and testing. Currently, there is no general rule for selecting the number of delays and hidden layers [36]. A deep LSTM can be built by stacking multiple LSTM layers, which generally works better than a single layer. Deep LSTM networks have been applied to solve many real-world sequence modelling problems [37]. The LSTM can also be used for planning studies [38], namely for planning the analysis of road traffic speed.

To produce a prediction model with good accuracy, it is necessary to optimize neural models' hyperparameters. While simple models can often produce good results with default hyperparemeters, the optimization process can greatly improve the results [39–41]. The selection of hyperparameters often makes the difference between underperformance and state-of-the-art performance. Optimization is often performed using machine learning algorithms, such as grid search, grey wolf optimization or particle swarm optimization. In the present prediction model, however, the hyperparameters were optimized manually, following a trial and error guided process, one variable at a time. This method was followed because it was the most convenient considering the limited computing power available.

2.3. LSTM with Encoder and Decoder

Experiments were performed with a predictive model based on the LSTM with encoder and decoder architecture. The model consists of two LSTMs, in which the first LSTM has the function of processing an input sequence and generating an encoded state. The encoded state compresses the information in the input stream. The second LSTM, called a decoder, uses the encoded state to produce an output sequence. Those input and output sequences can be of different lengths.

This technique has already been used to solve problems such as the prediction of vehicle trajectories based on deep learning [42]. This architecture [43] has shown great performance for tasks of translating from sequence to sequence. LSTM encoder–decoder models have also been proposed for learning tasks such as automatic translation [43,44]. There is the application of this model to solve many practical problems, such as the study of the equipment condition, applications in language translations, among others [45–47].

3. Theoretical Background

The present work uses LSTM networks, considering the referred different studies showing their usefulness for time series predictions [48,49]. The LSTM is a deep learning recurrent neural network architecture that is a variation of traditional recurrent neural networks (RNNs). It was introduced by Hochreiter and Schmidhuber in 1997. The most popular version is a modification refined by many works in the literature [50,51], which is called vanilla LSTM (hereinafter referred to as LSTM). The LSTM is excellent at handling time series data only with its network parameters. For example, weights and polarization are adjusted or optimized [52]. The primary modification of the LSTM when compared to the RNN architecture is the structure of the hidden layer [53]. The LSTM model is a powerful type of recurrent neural network (RNN), capable of learning long-term dependencies [54]. They became popular due to their power of representation and effectiveness in capturing long-term dependencies [55].

Many networks showed instability when dealing with exploding or vanishing gradient problems during learning. Those problems happen when the gradient of the error is too large or too small. If it is too large, it overflows and the errors cannot propagate properly through different layers during learning. If it is too small, it vanishes and the network does not learn. Different methods were proposed to solve those problems, known as a kind of of "door control" that is used in RNN models. For example, Gated Recurrent unit (GRU) algorithms [56,57], as the LSTMs [58,59], are to a large extent immune to the gradient problems and learn well.

The LSTM network structure is based on three ports whose function is to regulate the flow. Those ports are called the entrance door, the forget gate, and the exit door. The main port of entry is to regulate the entry of new memory data; the forget gate has the function of regulating the storage time in the network memory and the output port intends to regulate how much the value retained in memory influences the activation of the output block [60].

Kong et al. demonstrate some relevant conclusions such as (1) LSTM has a good predictive capacity; (2) their use can significantly improve the profit of service providers, so there is an opportunity when it comes to exploring the forecast in real time [61]. LSTM networks are the *de facto* gold standard for deep learning algorithms for analyzing time series data [55].

Figure 1 shows the internal architecture of an LSTM unit cell. According to [62,63], the internal calculation formulae of the LSTM unit are defined as follows:

$$i_t = \sigma(x_t U^i + h_{t-1} W^i + b_i) \quad (1)$$

$$f_t = \sigma(x_t U^f + h_{t-1} W^f + b_f) \quad (2)$$

$$o_t = \sigma(x_t U^o + h_{t-1} W^o + b_o) \quad (3)$$

$$a_t = tan(x_t U^C + h_{t-1} W^C + b_C) \quad (4)$$

where U^i, U^f, U^o and $U^C$are the weight matrices for mapping the current input layer on three ports and the state of the current input cell.

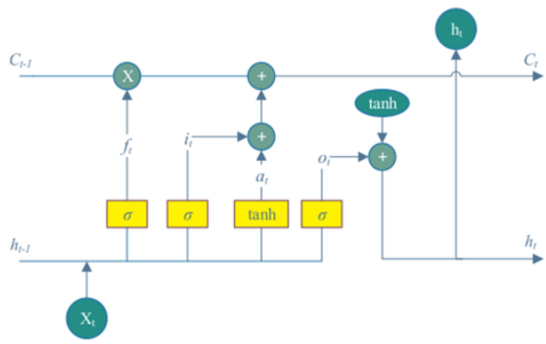


Figure 1. Detailed layout of a long short-term memory unit [63].

W^i, W^f, W^o and W^C are the weight matrices for mapping the previous output layer on three ports and the current state of the input cell. b_f, b_i, b_o, and b_c are polarization vectors for calculating the state of the door and the input cell. σ is the gate activation function, which is normally a sigmoid function. *tan* is the hyperbolic tangent function which is the activation function for the current state of the input cell.

Then, the current state of the output cell and the output layer can be calculated using the following equations.

$$C_t = \sigma(f_t \times C_{t-1} + i_t \times a_t) \quad (5)$$

$$h_t = tanh(C_t) \times o_t \quad (6)$$

To assess the quality of the prediction model, one of the most popular metrics is the Root Mean Square Error (RMSE), which is given by Equation (7):

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y})^2} \quad (7)$$

where Y_t represents the desired (real) value and $\hat{Y}$ is the predicted (obtained from the model) value. The difference between Y and $\hat{Y}$ is the error between the value expected to obtain and the value actually obtained from the network. n represents the number of samples used in the test set.

The RMSE, however, is an absolute error. Therefore, there are also the Mean Absolute Percentage Error (MAPE) and the Mean Absolute Error (MAE). Those errors are given by the following formulae:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |Y_t - \hat{Y}_t| \tag{8}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|Y_t - \hat{Y}_t|}{|Y_t|} \tag{9}$$

where Y_t represents the real value, $\hat{Y}_t$ the predicted value and n represents the total number of samples.

4. Data Preparation

Data are key to developing efficient modeling and planning. However, to be valuable, data need to be processed and structured before being analyzed.

4.1. The Problem

The main goal of the present work is to predict potential failures in an industrial drying press before they happen. Data come from six sensors installed in the press. Those sensors monitor the operation of the press, with a sampling period of one minute. The monitored variables are: (1) electric current intensity; (2) oil level at the hydraulic unit; (3) VAT pressure; (4) rotation speed; (5) temperature in the hydraulic unit; and (6) torque. The dataset contains six time series, one for each sensor, with the values stored in the database from 2016 to August 2020.

Figure 2 shows a plot of the six time series, before any processing is applied. These data present some upper and lower extremes, which may be discrepant data. Those discrepant samples may be due to reading errors or periods when the equipment was off or in another atypical state.

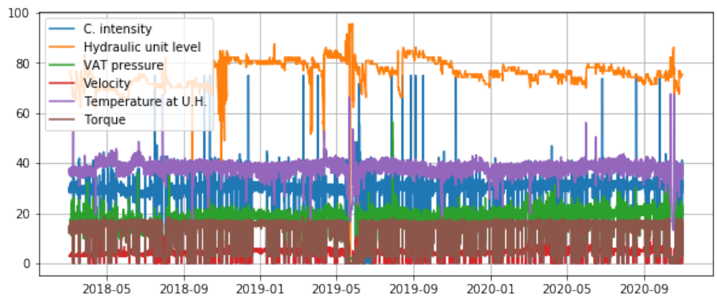


Figure 2. Plot of the original dataset values. The variables are electric current intensity, hydraulic unit oil level, VAT pressure, motor velocity, temperature at the hydraulic unit, and torque.

Some of the samples, such as those when the equipment was off but the sensors were still reading, can compromise the training of the machine learning models to be developed. Table 1 shows some statistical parameters such as mean, standard deviation (std), minimum, third quantiles, and maximum value.

Table 1. Statistical parameters of the dataset variables, before processing: C. intensity, hydraulic unit oil level, torque, VAT pressure, velocity, and temperature.

	C. Intensity	Hydraulic	Torque	VAT	Velocity	Temperature
mean	30.26	75.90	15.28	18.25	4.59	38.22
std	1.36	4.54	0.69	2.67	0.98	1.62
min	26.34	62.93	13.59	9.67	1.27	33.19
Q_1—25%	29.30	72.86	14.90	17.13	3.92	37.17
Q_2—50%	30.46	75.53	15.43	18.72	4.57	38.33
Q_3—75%	31.28	79.52	15.78	19.97	5.28	39.35
max	34.26	88.97	17.09	26.17	7.87	43.10

4.2. Cleaning Discrepant Data

In order to facilitate the training process, discrepant samples were identified and removed using the quantiles method. Samples which are beyond the $Q_1 - 3 \times std$ or $Q_3 + 3 \times std$ are replaced by the mean value. The extreme values were replaced with the average. Figure 3 shows the same variables after discrepant data samples were removed.

As the figure shows, the lines are now smoother and easier to read. Figure 4 shows that the samples are evenly distributed after the withdrawal of discrepant data.

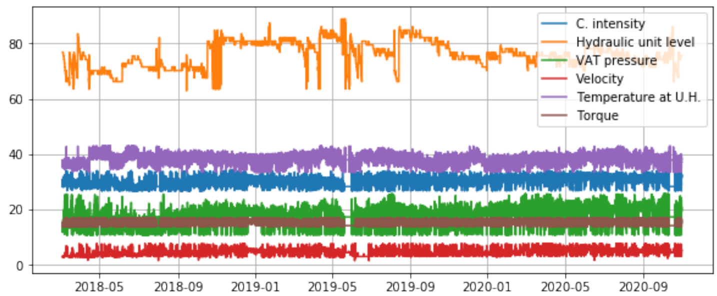


Figure 3. Plot of the dataset values after cleaning discrepant data. The variables are current intensity, hydraulic unit oil level, VAT pressure, velocity, temperature, and torque.

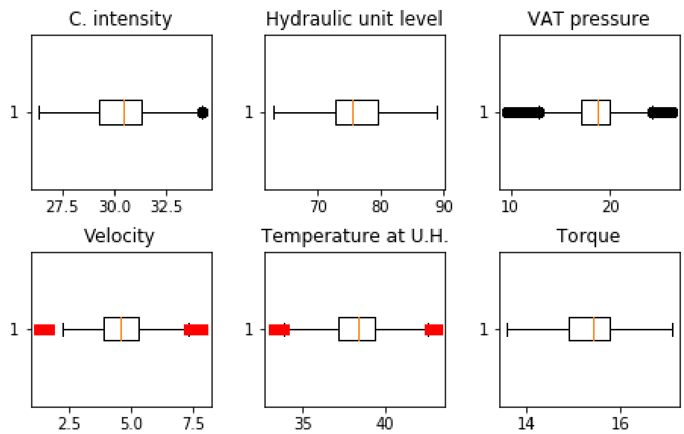


Figure 4. Distribution of data points of all the sensors, with lowly and highly discrepant data cleaned.

The predictive models to be used are robust and tolerant to noise. However, the cleaner data are expected to show better results. As an example, a provisional experiment to train a neural network LSTM model with a historical window of 70 samples and 40 LSTM unit cells showed higher and undetermined errors. The model was not able to learn or predict some variables, as shown in Table 2. With clean data, there were better and determinable results, as shown in Table 3. The tables show the MAPE and MAE for all input variables, as determined in the test set. They also show the RMSE, as calculated in the train and test sets, globally for all variables.

Table 2. Prediction results without cleaning discrepant data in the database, with a window of 70 samples and 40 LSTM units.

	Window 70 Days					
	C. Intensity	**Hydraulic**	**Torque**	**VAT**	**Velocity**	**Temperature**
MAPE	inf	8.46	inf	98.19	inf	11.59
MAE	3.52	6.57	24.73	10.53	14.88	4.21
	Train	**Test**				
RMSE	79.52	79.64				

Table 3. Forecast results with treatments in the database with 40 LSTM units.

	Window 70 Days					
	C. Intensity	**Hydraulic**	**Torque**	**VAT**	**Velocity**	**Temperature**
MAPE	2.52	3.02	2.44	13.10	inf	2.48
MAE	0.76	2.28	0.37	1.32	0.57	0.94
	Train	**Test**				
RMSE	1.71	1.97				

5. Experiments and Results

Experiments were performed with the aim of validating the model that has the best performance in predicting data from the industrial press. The tests are divided into two subsections, first with resampling of data to one sample per day and then with resampling for a sample each 12 h.

5.1. LSTM Models and Dataset Partition

After processing the data, experiments were performed with an LSTM model. The model included an encoder and decoder, with one hidden LSTM layer in the middle and a dense layer at the output. The model was used to train and predict, with six variables that represent data coming from the paper press sensors. The goal was to forecast the value of those variables with the highest possible level of confidence so that it brings added benefits in predictive maintenance.

Figure 5 describes the architecture of one of the network models used. The models were implemented in Python using the TensorFlow library and Keras.

The experiments were performed aiming to obtain a prediction for all variables one month in advance, from a window of a number of past samples.

The LSTM models received, as an input, a sequence consisting of the composition of a number of samples for each variable. The number of samples depended on the window size and the resampling rate used. The output sequence is composed of the values predicted for each of the variables.

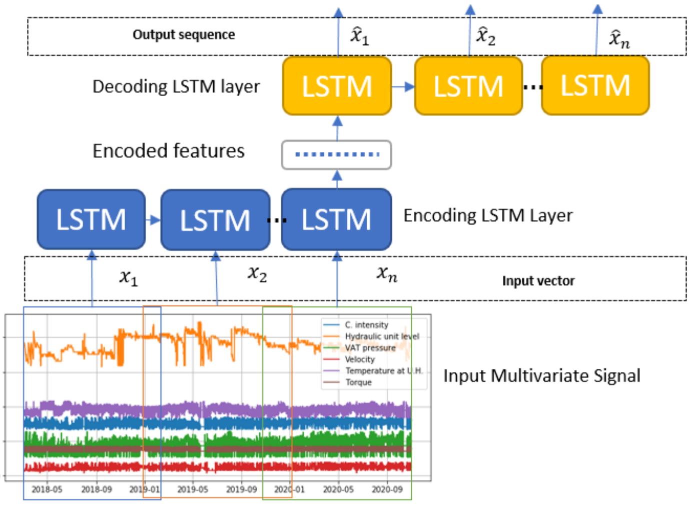


Figure 5. Model summary of one of the LSTM networks used. The model receives a window of n samples of each variable and predicts the value of those variables as predicted 30 days ahead.

To train and test the models, the dataset was divided into train and test subsets. Validation was performed using the test set, but those samples were not incorporated into the training set. The training set contained 85% of the samples and the test set the remaining 15% of samples. These values are adequate for convergence during learning. As an example, Figure 6 shows a learning curve for a model with 70 units in the middle layer and a window of 30 lag samples. The figure shows that learning converges and takes fewer than 10 epochs. The remainder experiments were performed using 100 epochs.

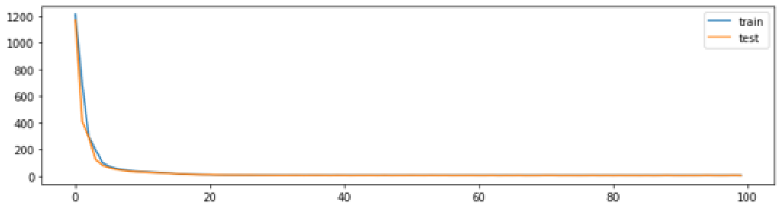


Figure 6. Example of learning curve, showing the loss measured during training of an LSTM model.

5.2. Experiments to Determine Historical Window Size and Number of LSTM Units Using One Sample per Day

The first experiments performed aimed to determine the best window size to use. The smaller the window, the smaller and faster the model that can be used. However, if the window is too small, it may be insufficient to make accurate predictions.

The original dataset had 1,445,760 data points, which is very large and would require a lot of memory and time to train and test. The experiments were performed after downsampling the data, so that there is only one sample per day. That sample is the average of 1004 original samples. The downsampled dataset is, therefore, less than the one thousand of the original dataset.

The results are measured in the test set. The figure above shows the MAPE and MAE measured for each variable. It also shows the global RMSE measured globally for the train and test sets.

As Figure 7 shows, models with windows of 40 and 50 samples allow better learning and produce smaller prediction errors.

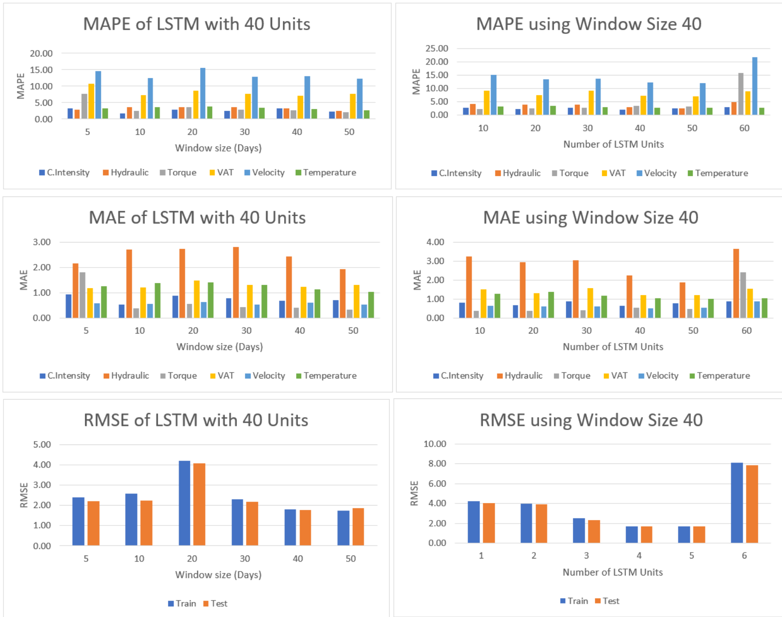


Figure 7. Results obtained with a different number of LSTM cells in the hidden layer, as well as different sliding window sizes, to predict values 30 days in advance with downsampling to one sample per day.

Additional experiments were performed to determine the best size for the number of cells in the hidden layer. For those experiments, a window of 40 historical samples was used, relying on the results of the previous experiments.

Figure 7 shows the results obtained for experiments with a window of 40 days and different numbers of hidden cells. As the results show, the model with the best performance is the one with 50 hidden cells.

After the results of the first experiments with one sample per day, additional experiments were conducted to determine if there was any considerable loss in downsampling from one sample per minute to one sample per day. A first experiment was performed, which consisted of halving the downsampling period from 24 to just 12 h. Therefore, the dataset doubled in size, since it contained two samples per day instead of just one.

5.3. Experiments to Determine Historical Window Size and Number of Unit LSTMs Using Two Samples per Day

According to the results shown in Figure 8, it is concluded that a window of 10 days (20 samples) shows the best performance. This shows that the model can exhibit approximately the same performance with even fewer input samples when compared to the models above. The models used for those experiments had 20 cells in the hidden layer.

Once the impact of the window size was determined, additional experiments were performed to gain a better insight into the impact of using more or less cells in the hidden layer. Figure 8 shows results of using different numbers of cells.

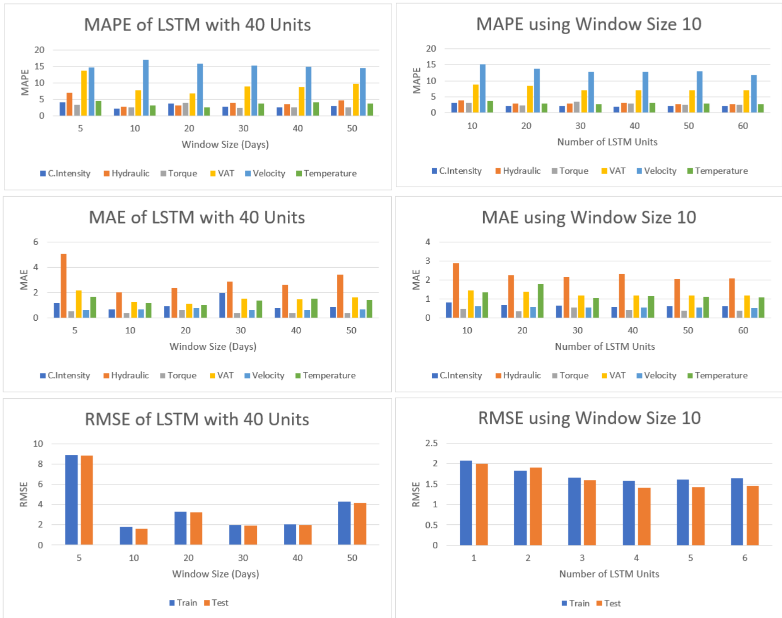


Figure 8. Results obtained with a different number of cells in the hidden layer, also using different window samples to predict values 30 days in advance with resampling for the two samples for a day.

5.4. *Plot of One Result*

Figure 9 shows plots of the results obtained using the model using 40 units in the hidden layer and a 10-day window of samples. As the figure shows, the forecasts in general follow the actual signals most of the time. However, there are still some areas where the actual signal diverges a small percentage from the prediction, namely for velocity and temperature. Most of the differences may be due to behaviors that are still difficult to capture due to the small dimension of the dataset. As more data will be collected, the neural models will probably be able to capture more patterns and offer more accurate predictions.

In addition to the graphs shown in Figure 9, in Tables 4 and 5, the magnitudes of the RMSE errors in the training set and test set are also presented. They were measured in the model validation dataset.

Table 4. The magnitude of RMSE errors in the test and training set, using one sample per day.

Window Size (Days)	Train	Test	Units	Train	Test
5	2.39	2.20	10	4.23	4.07
10	2.57	2.24	20	3.99	3.93
20	4.21	4.09	30	2.52	2.35
30	2.31	2.19	40	1.68	1.70
40	1.81	1.77	50	1.66	1.70
50	1.74	1.86	60	8.14	7.85

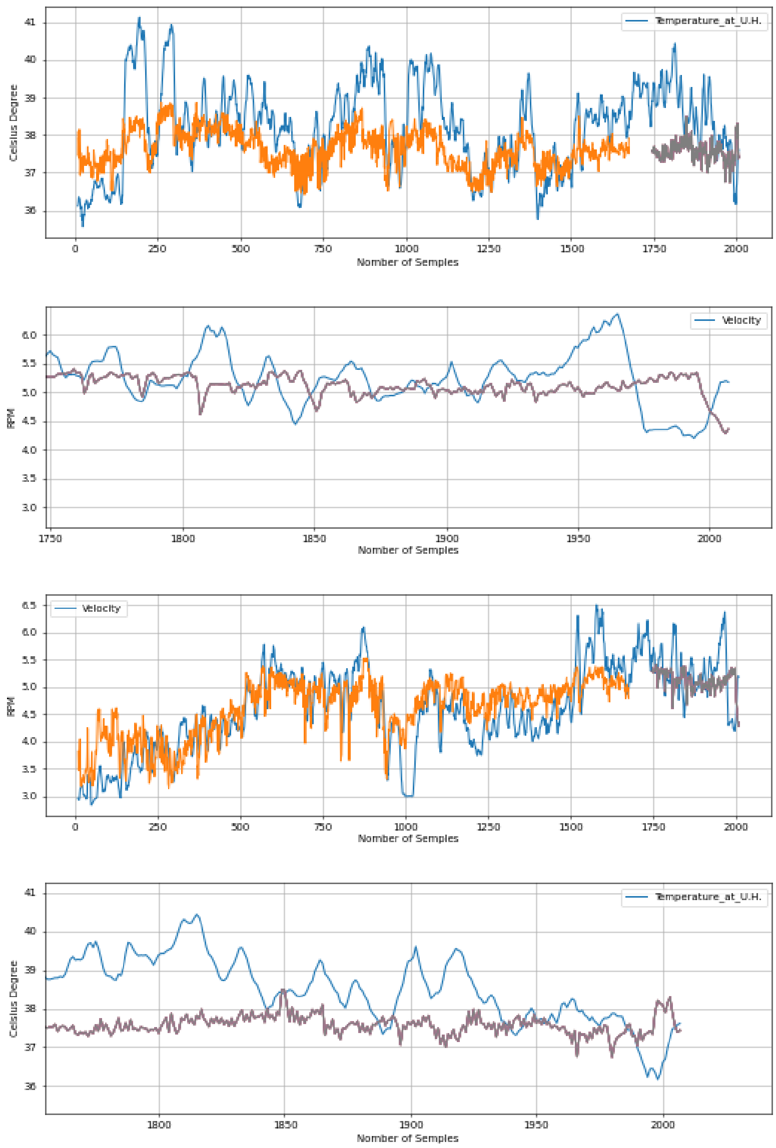


Figure 9. Variable forecast with a window of samples of 10 days, sampling rate two samples per day, and a network model with 50 units in the hidden layer. The blue lines show the actual value. The orange lines show the predictions during the training set and the gray lines show the predictions in the test set.

Table 5. The magnitude of RMSE errors in the test and training set, using two samples per day.

Window Size (Days)	Train	Test	Units	Train	Test
5	8.91	8.87	10	2.07	1.99
10	1.80	1.61	20	1.82	1.91
20	3.29	3.23	30	1.65	1.59
30	1.98	1.94	40	1.58	1.41
40	2.07	1.98	50	1.61	1.42
50	4.32	4.16	60	1.64	1.46

6. Discussion

Anticipating industrial equipment's future behavior is a goal that has been long sought after, for it allows predictive maintenance to perform the right actions at the right time. Therefore, the application of time series and other artificial intelligence models to forecast the equipment's state is a new and growing area of interest.

The present research uses a dataset of approximately 2.5 years of data of an industrial paper press. A procedure to clean the data is proposed and different experiments are described to use a deep neural model based on LSTM recurrent networks.

The method proposed is going to be applied in other industrial presses, aiming to improve predictive maintenance. Based on the state of the art and experiments, this architecture presents a good versatility, depending of course on the quality of data and hyperparameter settings.

The results show that it is possible to optimize neural models to forecast future values 30 days in advance. The model experimented uses as input a vector consisting of concatenation of a number of samples of all variables. The output is a vector with the predictions of all samples too. The performance of the models is generally better for some variables and worse for others. Those differences will be dealt with in future work.

An important conclusion is that the downsampling used might have been too aggressive. Experiments were performed using one sample per day and two samples per day. The models trained with two samples per day showed a better performance. Hence, more resolution is better for reducing errors and may allow for better learning. That is achieved at the cost of additional processing power. This is also another research question which will be dealt with in future work.

7. Conclusions

Predicting industrial machines' future behaviors is key for predictive maintenance success. The present research aims to find prediction models adequate for anticipating the future behavior of industrial equipment with good certainty.

The predictive model used was based on LSTM networks, with encoding and decoding layers as the input and output, respectively. In this study, different data pre-processing techniques, network architectures, and hyperparameters were tested, in order to determine the best models.

The predictive model used was based on LSTM network, with encoding and decoding layers as the input and output, respectively.

The results show that the model proposed is able to learn and forecast the behavior of the six variables studied: torque, pressure, current intensity, velocity, oil level and temperature. The best results were obtained using a window of samples of the last 10 days at two samples per day. The MAPE errors varied in the range of 2 to 17% for one of the best models for different variables.

Future work includes additional experiments to determine the optimal sampling rate and stabilize the results for optimal performance with all the variables. The predicted results will also be used to determine an expected probability of failure, using classification models. Other methods may also be used to deal with discrepant data. Later, the models developed will also be applied to other equipment.

Author Contributions: Conceptualization, J.T.F., A.M.C., M.M.; methodology, J.T.F. and M.M.; software, B.C.M. and M.M.; validation, J.T.F. and M.M.; formal analysis, J.T.F. and M.M.; investigation, B.C.M. and M.M.; resources, T.F., A.M.C. and M.M.; writing—original draft preparation, B.C.M.; writing—review and editing, J.T.F. and M.M.; project administration, J.T.F. and A.M.C.; funding acquisition, J.T.F. and A.M.C. All authors have read and agreed to the published version of the manuscript.

Funding: The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowvska-Curie grant agreement 871284 project SSHARE and the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under Project POCI-01-0145-FEDER-029494, and by National Funds through the FCT—Portuguese Foundation for Science and Technology, under Projects PTDC/EEI-EEE/29494/2017, UIDB/04131/2020, and UIDP/04131/2020.

Acknowledgments: This research is sponsored by FEDER funds through the program COMPETE—Programa Operacional Factores de Competitividade—and by national funds through FCT—Fundação para a Ciência e a Tecnologia—under the project UIDB/00285/2020. This work was produced with the support of INCD funded by FCT and FEDER under the project 01/SAICT/2016 nº 022153.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ARMA	Autoregressive Integrated Moving Average
CNN	Convolution Neural Networks
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
RNN	Recurrent Neural Networks

References

1. Sahal, R.; Breslin, J.; Ali, M. Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case. *J. Manuf. Syst.* **2020**, *54*, 138–151. [CrossRef]
2. Ahmed, M.S.; Cook, A.R. *Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques*; Transportation Research Board: Washington, DC, USA, 2020; ISBN 9780309029728.
3. Majeed, A.; Zhang, Y.; Ren, S.; Lv, J.; Peng, T.; Waqar, S.; Yin, E. A big data-driven framework for sustainable and smart additive manufacturing. *Robot. Comput. Integr. Manuf.* **2021**, *67*, 102026. [CrossRef]
4. Ferreiro, S.; Konde, E.; Fernández, S.; Prado, A. Industry 4.0: Predictive intelligent maintenance for production equipment. In *European Conference of the Prognostics and Health Management Society*; 2016; pp. 1–8. Available online: https://www.semanticscholar.org/paper/INDUSTRY-4-.-0-%3A-Predictive-Intelligent-Maintenance-Ferreiro-Konde/638c2b72a747ea4b82e098572be820083dca9c7a (accessed on 25 June 2021).
5. Wang, K. Intelligent predictive maintenance (IPdM) system—Industry 4.0 scenario. *WIT Trans. Eng. Sci.* **2016**, *113*, 259–268.
6. Yu, W.; Dillon, T.; Mostafa, F.; Rahayu, W.; Liu, Y. A global manufacturing big data ecosystem for fault detection in predictive maintenance. *IEEE Trans. Ind. Inform.* **2020**, *16*, 183–192. [CrossRef]
7. Pais, E.; Farinha, J.T.; Cardoso, A.J.M.; Raposo, H. Optimizing the Life Cycle of Physical Assets—A Review. *WSEAS Trans. Syst. Control* **2020**, *15*, 417–430. [CrossRef]
8. Martins, A.B.; Torres Farinha, J.; Marques Cardoso, A. Calibration and Certification of Industrial Sensors—A Global Review. *WSEAS Trans. Syst. Control* **2020**, *15*, 394–416. [CrossRef]
9. Rodrigues, J.; Cost, I.; Farinha, J.; Mendes, M.; Margalho, L. Predicting motor oil condition using artificial neural networks and principal component analysis. *Eksploat. Niezawodn.* **2020**, *22*, 440–448. [CrossRef]
10. Torres Farinha, J. *Asset Maintenance Engineering Methodologies*; CRC Press: Boca Raton, FL, USA, 2018.
11. Chen, M.; Tseng, H. An approach to design of maintenance float systems. *Integr. Manuf.* **2003**, *14*, 458–467.
12. Daniyan, I.; Mpofu, K.; Oyesola, M.; Ramatsetse, B.; Adeodu, A. Artificial intelligence for predictive maintenance in the railcar learning factories. *Procedia Manuf.* **2020**, *45*, 13–18. [CrossRef]
13. Hsu, Y.Y.; Tung, T.T.; Yeh, H.C.; Lu, C.N. Two-Stage Artificial Neural Network Model for Short-Term Load Forecasting. *IFAC-PapersOnLine* **2018**, *51*, 678–683. [CrossRef]
14. Balduíno, M.; Torres Farinha, J.; Marques Cardoso, A. Production Optimization versus Asset Availability—A Review. *WSEAS Trans. Syst. Control* **2020**, *15*, 320–332. [CrossRef]

15. Jimenez, V.J.; Bouhmala, N.; Gausdal, A.H. Developing a predictive maintenance model for vessel machinery. *J. Ocean. Eng. Sci.* **2020**, *5*, 358–386. [CrossRef]
16. Ayvaz, S.; Alpay, K. Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time. *Expert Syst. Appl.* **2021**, *173*, 114598. [CrossRef]
17. Yu, Z.; Moirangthem, D.S.; Lee, M. Continuous Timescale Long-Short Term Memory Neural Network for Human Intent Understanding. *Front. Neurorobot.* **2017**, *11*, 42. [CrossRef]
18. Aydin, O.; Guldamlasioglu, S. Using LSTM networks to predict engine condition on large scale data processing framework. In Proceedings of the 2017 4th International Conference on Electrical and Electronic Engineering (ICEEE), Ankara, Turkey, 8–10 April 2017; pp. 281–285. [CrossRef]
19. Wang, Q.; Bu, S.; He, Z. Achieving Predictive and Proactive Maintenance for High-Speed Railway Power Equipment with LSTM-RNN. *IEEE Trans. Ind. Inform.* **2020**, *16*, 6509–6517. [CrossRef]
20. Bruneo, D.; De Vita, F. On the Use of LSTM Networks for Predictive Maintenance in Smart Industries. In Proceedings of the 2019 IEEE International Conference on Smart Computing (SMARTCOMP), Washington, DC, USA, 12–15 June 2019; pp. 241–248. [CrossRef]
21. Dong, D.; Li, X.Y.; Sun, F.Q. Life prediction of jet engines based on LSTM-recurrent neural networks. In Proceedings of the 2017 Prognostics and System Health Management Conference (PHM-Harbin), Harbin, China, 9–12 July 2017; pp. 1–6. [CrossRef]
22. Bampoula, X.; Siaterlis, G.; Nikolakis, N.; Alexopoulos, K. A Deep Learning Model for Predictive Maintenance in Cyber-Physical Production Systems Using LSTM Autoencoders. *Sensors* **2021**, *21*, 972. [CrossRef]
23. Mathew, V.; Toby, T.; Singh, V.; Rao, B.M.; Kumar, M.G. Prediction of Remaining Useful Lifetime (RUL) of turbofan engine using machine learning. In Proceedings of the 2017 IEEE International Conference on Circuits and Systems (ICCS), Thiruvananthapuram, India, 20–21 December 2017; pp. 306–311. [CrossRef]
24. Düdükçü, H.V.; Taşkıran, M.; Kahraman, N. LSTM and WaveNet Implementation for Predictive Maintenance of Turbofan Engines. In Proceedings of the 2020 IEEE 20th International Symposium on Computational Intelligence and Informatics (CINTI), Budapest, Hungary, 5–7 November 2020; pp. 000151–000156. [CrossRef]
25. Essien, A.; Giannetti, C. A Deep Learning Model for Smart Manufacturing Using Convolutional LSTM Neural Network Autoencoders. *IEEE Trans. Ind. Inform.* **2020**, *16*, 6069–6078. [CrossRef]
26. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]
27. Beshr, A.; Zarzoura, F. Using artificial neural networks for GNSS observations analysis and displacement prediction of suspension highway bridge. *Innov. Infrastruct. Solut.* **2021**, *6*. [CrossRef]
28. Sak, H.; Senior, A.; Beaufays, F. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. *arXiv* **2014**, arXiv:1402.1128.
29. Chen, Z.; Liu, Y.; Liu, S. Mechanical state prediction based on LSTM neural netwok. In Proceedings of the 2017 36th Chinese Control Conference (CCC), Dalian, China, 26–28 July 2017; pp. 3876–3881. [CrossRef]
30. Ghosh, S.; Vinyals, O.; Strope, B.; Roy, S.; Dean, T.; Heck, L. Contextual LSTM (CLSTM) models for Large scale NLP tasks. *arXiv* **2016**, arXiv:1602.06291.
31. Shahid, F.; Zameer, A.; Muneeb, M. A novel genetic LSTM model for wind power forecast. *Energy* **2021**, *223*, 120069. [CrossRef]
32. Guo, J.; Xie, Z.; Qin, Y.; Jia, L.; Wang, Y. Short-Term Abnormal Passenger Flow Prediction Based on the Fusion of SVR and LSTM. *IEEE Access* **2019**, *7*, 42946–42955. [CrossRef]
33. Ren, L.; Dong, J.; Wang, X.; Meng, Z.; Zhao, L.; Deen, M. A Data-Driven Auto-CNN-LSTM Prediction Model for Lithium-Ion Battery Remaining Useful Life. *IEEE Trans. Ind. Inform.* **2021**, *17*, 3478–3487. [CrossRef]
34. Niu, K.; Zhang, H.; Zhou, T.; Cheng, C.; Wang, C. A Novel Spatio-Temporal Model for City-Scale Traffic Speed Prediction. *IEEE Access* **2019**, *7*, 30050–30057. [CrossRef]
35. Feng, M.; Zheng, J.; Ren, J.; Hussain, A.; Li, X.; Xi, Y.; Liu, Q. Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data. *IEEE Access* **2019**, *7*, 106111–106123. [CrossRef]
36. Palangi, H.; Deng, L.; Shen, Y.; Gao, J.; He, X.; Chen, J.; Song, X.; Ward, R. Deep Sentence Embedding Using Long Short-Term Memory Networks: Analysis and Application to Information Retrieval. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2016**, *24*, 694–707. [CrossRef]
37. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
38. Mao, Y.; Qin, G.; Ni, P.; Liu, Q. Analysis of road traffic speed in Kunming plateau mountains: A fusion PSO-LSTM algorithm. *Int. J. Urban Sci.* **2021**, 1–21. [CrossRef]
39. Hutter, F.; Lücke, J.; Schmidt-Thieme, L. Beyond Manual Tuning of Hyperparameters. *KI Künstliche Intell.* **2015**, *29*, 329–337. [CrossRef]
40. Khalid, R.; Javaid, N. A survey on hyperparameters optimization algorithms of forecasting models in smart grid. *Sustain. Cities Soc.* **2020**, *61*, 102275. [CrossRef]
41. Hutter, F.; Hoos, H.; Leyton-Brown, K. An Efficient Approach for Assessing Hyperparameter Importance. In Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 21–26 June 2014; pp. 754–762.

42. Park, S.H.; Kim, B.; Kang, C.M.; Chung, C.C.; Choi, J.W. Sequence-to-Sequence Prediction of Vehicle Trajectory via LSTM Encoder-Decoder Architecture. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; pp. 1672–1678. [CrossRef]
43. Cho, K.; van Merrienboer, B.; Gülçehre, Ç.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv* **2014**, arXiv:1406.1078.
44. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. *arXiv* **2014**, arXiv:1409.3215.
45. Wang, T.; Chen, P.; Amaral, K.; Qiang, J. An Experimental Study of LSTM Encoder-Decoder Model for Text Simplification. *arXiv* **2016**, arXiv:1609.03663.
46. Bengio, S.; Vinyals, O.; Jaitly, N.; Shazeer, N. Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks. *arXiv* **2015**, arXiv:1506.03099.
47. Du, S.; Li, T.; Yang, Y.; Horng, S.J. Multivariate time series forecasting via attention-based encoder–decoder framework. *Neurocomputing* **2020**, *388*, 269–279. [CrossRef]
48. Gers, F.A.; Eck, D.; Schmidhuber, J. Applying LSTM to Time Series Predictable Through Time-Window Approaches. In *Neural Nets WIRN Vietri-01*; Tagliaferri, R., Marinaro, M., Eds.; Perspectives in Neural Computing; Springer: Berlin/Heidelberg, Germany, 2002; pp. 193–200._20. [CrossRef]
49. Meng, Q.; Wang, H.; He, M.; Gu, J.; Qi, J.; Yang, L. Displacement prediction of water-induced landslides using a recurrent deep learning model. *Eur. J. Environ. Civ. Eng.* **2020**. [CrossRef]
50. Sarikaya, R.; Hinton, G.E.; Deoras, A. Application of Deep Belief Networks for Natural Language Understanding. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2014**, *22*, 778–784. [CrossRef]
51. Sundermeyer, M.; Schlüter, R.; Ney, H. LSTM neural networks for language modeling. In Proceedings of the Thirteenth Annual Conference of the International Speech Communication Association, Portland, OR, USA, 9–13 September 2012.
52. Hu, Y.; Sun, X.; Nie, X.; Li, Y.; Liu, L. An Enhanced LSTM for Trend Following of Time Series. *IEEE Access* **2019**, *7*, 34020–34030. [CrossRef]
53. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to Forget: Continual Prediction with LSTM. In Proceedings of the 1999 Ninth International Conference on Artificial Neural Networks ICANN 99 (Conf. Publ. No. 470), Edinburgh, UK, 7–10 September 1999; pp. 850–855. [CrossRef]
54. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2222–2232. [CrossRef] [PubMed]
55. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
56. Li, C.; Jiang, P.; Zhou, A. Rigorous solution of slope stability under seismic action. *Comput. Geotech.* **2019**, *109*, 99–107. [CrossRef]
57. Li, C.; Hu, B.C.; Hu, D.; Xu, X.F.; Zong, X.C.; Li, J.P.; Wu, M.C. Stereoselective ring-opening of styrene oxide at elevated concentration by Phaseolus vulgaris epoxide hydrolase, PvEH2, in the organic/aqueous biphasic system. *Catal. Commun.* **2019**, *123*, 1–5. [CrossRef]
58. Qin, Y.; Li, K.; Liang, Z.; Lee, B.; Zhang, F.; Gu, Y.; Zhang, L.; Wu, F.; Rodriguez, D. Hybrid forecasting model based on long short term memory network and deep learning neural network for wind signal. *Appl. Energy* **2019**, *236*, 262–272. [CrossRef]
59. Zhang, Z.; Ye, L.; Qin, H.; Liu, Y.; Wang, C.; Yu, X.; Yin, X.; Li, J. Wind speed prediction method using Shared Weight Long Short-Term Memory Network and Gaussian Process Regression. *Appl. Energy* **2019**, *247*, 270–284. [CrossRef]
60. Zhao, R.; Yin, Y.; Shi, Y.; Xue, Z. Intelligent intrusion detection based on federated learning aided long short-term memory. *Phys. Commun.***2020**, *42*, 101157. [CrossRef]
61. Kong, X.; Kong, D.; Bai, L.; Xiao, J. Online pricing of demand response based on long short-term memory and reinforcement learning. *Appl. Energy* **2020**, *271*, 114945. [CrossRef]
62. Li, Y.; Lu, Y. LSTM-BA: DDoS Detection Approach Combining LSTM and Bayes. In Proceedings of the 2019 Seventh International Conference on Advanced Cloud and Big Data (CBD), Suzhou, China, 21–22 September 2019; pp. 180–185. [CrossRef]
63. Alameer, Z.; Fathalla, A.; Li, K.; Ye, H.; Jianhua, Z. Multistep-ahead forecasting of coal prices using a hybrid deep learning model. *Resour. Policy* **2020**, *65*, 101588. [CrossRef]

Appendix D

energies

Article

Comparing LSTM and GRU Models to Predict the Condition of a Pulp Paper Press

Balduíno César Mateus [1,2,*], Mateus Mendes [3,4], José Torres Farinha [4,5], Rui Assis [1] and António Marques Cardoso [2]

[1] EIGeS—Research Centre in Industrial Engineering, Management and Sustainability, Lusófona University, Campo Grande, 376, 1749-024 Lisboa, Portugal; rassis46@gmail.com
[2] CISE—Electromechatronic Systems Research Centre, University of Beira Interior, Calçada Fonte do Lameiro, 62001-001 Covilhã, Portugal; ajmc@ubi.pt
[3] Instituto Superior de Engenharia de Coimbra, Polytechnic of Coimbra, 3045-093 Coimbra, Portugal; mmendes@isec.pt
[4] Institute of Systems and Robotics, University of Coimbra, 3004-531 Coimbra, Portugal; torresfarinha@dem.uc.pt
[5] Centre for Mechanical Engineering, Materials and Processes—CEMMPRE, University of Coimbra, 3030-788 Coimbra, Portugal
* Correspondence: balduino.mateus@ubi.pt

Abstract: The accuracy of a predictive system is critical for predictive maintenance and to support the right decisions at the right times. Statistical models, such as ARIMA and SARIMA, are unable to describe the stochastic nature of the data. Neural networks, such as long short-term memory (LSTM) and the gated recurrent unit (GRU), are good predictors for univariate and multivariate data. The present paper describes a case study where the performances of long short-term memory and gated recurrent units are compared, based on different hyperparameters. In general, gated recurrent units exhibit better performance, based on a case study on pulp paper presses. The final result demonstrates that, to maximize the equipment availability, gated recurrent units, as demonstrated in the paper, are the best options.

Keywords: LSTM; recurrent neural network; GRU; paper press; predictive maintenance

check for updates

Citation: Mateus, B.C.; Mendes, M.; Farinha, J.T.; Assis, R.; Cardoso, A.M. Comparing LSTM and GRU Models to Predict the Condition of a Pulp Paper Press. *Energies* **2021**, *14*, 6958. https://doi.org/10.3390/en14216958

Academic Editor: Jaroslaw Krzywanski

Received: 21 September 2021
Accepted: 16 October 2021
Published: 22 October 2021

1. Introduction

Modern algorithms, data storage, and computing power make it possible to not only analyze past behavior, but to anticipate future behavior of industrial equipment with reasonable confidence [1–3]. Anticipating future failures is, therefore, a topic that has been receiving increasingly more attention from researchers.

There are a few types of maintenance: curative, which solves problems after they occur; preventive, which can be done at regular intervals, aimed at preventing common problems; conditioning, namely in the predictive way, which attempts to predict problems that are going to happen and prevent them from happening at the optimal time [4].

Nowadays, predictive maintenance is the most common approach. It aims to optimize maintenance costs and increase equipment availability [5]. Maintenance procedures are performed when parts are supposed to be worn out, preventing failures and halting the production processes for more time than strictly necessary. Its main focus is to prevent future failures. However, in this case, some parts may be replaced before they are actually worn out, while others may wear out faster than expected and still fail [6]. Predictive maintenance aims to make the process more efficient, narrowing down the optimal time window for maintenance procedures. Using sensory data and adequate forecasting algorithms, the state of the equipment can be determined and the optimal time for maintenance interventions can be predicted some time in advance, avoiding unnecessary costs, as well as failures due to lack of maintenance.

Traditional forecasting algorithms have relied more on time series models, such as exponential smoothing [7] and seasonal autoregressive integrated moving average (SARIMA) [8–10].

More recently, however, artificial intelligence methods have become more popular. They impact societies, politics, economies, and industries [11], offering tools for data analysis, pattern recognition, and prediction, which could be beneficial in predictive maintenance and in production systems.

Modern machine learning methods offer superior performance and have become more popular [12]. They can work with high-dimensional data and multivariate data [13]. The most popular tools include artificial neural networks (ANNs), which have been proposed in many industrial applications, including soft sensing [14] and predictive control [15]. Random forest models are also good predictors, as shown in this study [16].

Traditional ANNs are simple and adequate for a wide range of problems. Bangalore et al. have studied the performance of neural networks for early detection of faults in gearbox bearings, to optimize the maintenance of wind turbines [17]. However, for prediction in sequential data, long shot-term memory (LSTM) and gated recurrent units (GRUs) have shown superior performance [18].

LSTM is very good at predicting in a time series [19,20]. It could extract patterns from sequential data and store these patterns in internal state variables. Each LSTM cell can retain important information for a longer period when it is used. This information property allows the LSTM to perform well in classifying, processing, or predicting complex dynamic sequences [21].

The present study aims to compare the performance of LSTM and GRU to solve the problem of predicting the future behavior of an industrial paper pulp press.

Section 2 presents a survey of related work. Section 3 describes the theory of the LSTM and GRU networks, as well as the formulae used to calculate the different errors. Section 4 describes the methods used for cleaning the dataset and also the behavior of some samples. Section 6 describes the tests, results, and validation of the predictive models. Section 7 discusses the results and compares them to the state-of-the-art. Section 8 draws some conclusions and suggestions for future work.

2. Literature Review

Monitoring physical assets has becoming a priority for predictive maintenance. Recent studies prove the importance of the topic [22,23]. Many statistical and machine learning tools have been used for prediction purposes, in monitoring and preventing equipment failures [24,25], quality control [26], and in other areas [27].

Artificial neural networks have received special attention in the area of electrical energy. Studies, such as [27,28], show their capacity and performance as good predictors, as long as a dataset with sufficient quality and quantity of data is available and the right parameters are found.

2.1. Predictive Maintenance

The creation of a predictive maintenance program is a strategic decision that, until now, has lacked analysis of issues related to its installation, management, and control. Carnero [29] suggests that predictive maintenance can provide an increase in safety, quality, and availability in industrial plants.

Bansal et al. [30] present a new real-time predictive maintenance system for machine systems based on neural networks. Other studies, such as [31,32], indicate the feasibility of artificial neural networks for predictive maintenance.

2.2. MLP and Recurrent Networks

Multilayer Perceptron (MLP) neural networks have been used with success for predicting and diagnosing pump failures, showing promising results with different types of failures [33–36]. According to Ni and Wang [37] Partovi and Anandarajan [38], neural networks have high prediction accuracies and aid in decision-making [39].

In the context of recurrent neural networks, LSTM-based models presented good performance in time series classification tasks and prediction tasks [40]. The LSTM network is useful in solving non-linear problems due to its non-linear processing capacity [41].

Sakalle et al. [42] used an LSTM network to recognize a number of emotions in brain waves. The results obtained with the LSTM were superior when compared to the other models mentioned in the study. The same approach was used in predictive and proactive maintenance for high-speed rail power equipment [43]. Some architectures have good ability in predicting univariate or multivariate temporal series with LSTM and GRU networks [44–46].

Models that use RNN are usually suitable for time-series information. Hochreiter and Schmidhuber [47] proposed an LSTM, which showed an extraordinary execution power in several sequence-centric tasks, such as handwriting recognition [48,49], auditory speech demonstration [50,51], dialect modeling [52], and dialect translation. Besides these areas, networks have also been used in predicting heart failure [53].

2.3. Deep Learning

Recently, deep learning strategies have been used, with success, in a variety of areas [54]. Vincent et al. [55] show that deep neural networks can outperform other methods in voice recognition tasks. A similar approach was used in audio processing [56].

Yasaka et al. [35] used deep learning with a convolutional neural network (CNN), obtaining a high performance in image recognition. The images themselves can be used in a learning process with this technique, and feature extraction prior to the learning process is not necessary. Other studies in the field of computer vision include [57,58].

Krizhevsky et al. [36] showed good results in image processing, employing a layered pre-training technique. The analysis shows that a large deep convolutional neural network can achieve record-breaking results in a challenging data collection using supervised learning. This same study demonstrates how important the amount of convolutional layers is to achieve good results. In order to learn the types of difficult functions that can represent high-level abstractions, it is necessary to have deep architectures. There is a need for an exhaustive exploration of the types of layers, sizes, transfer functions, and other hyperparameters [59].

3. Theoretical Background

3.1. Long Short-Term Memory

Figure 1 shows the inner design of an LSTM unit cell, according to Li and Lu [60]. Formally, the LSTM cell model is characterized as follows:

$$f_t = \sigma(x_t W_f + h_{t-1} U_f + b_f) \tag{1}$$

$$i_t = \sigma(x_t W_i + h_{t-1} U_i + b_i) \tag{2}$$

$$o_t = \sigma(x_t W_o + h_{t-1} U_o + b_o) \tag{3}$$

$$\tilde{C}_t = tan[(x_t W_C + h_{t-1} U_c + b_c] \tag{4}$$

$$C_t = \sigma(f_t \times C_{t-1} + i_t \times \tilde{C}_t) \tag{5}$$

$$h_t = tanh(C_t) \times o_t \tag{6}$$

Matrices W_q and U_q contain the weights of the input and recurrent connections, where the index can be the input gate i, output gate o, the forgetting gate f or the memory cell c, depending on the activation being calculated. $c_t \in \mathbb{R}^h$ is not just a cell of an LSTM unit, but contains h cells of the LSTM units, while i_t, o_t and f_t represent the activations of, respectively, the input, output and forget gates, at time step t, where:

- $x_t \in \mathbb{R}^d$: input vector to the LSTM unit;
- $f_t \in (0,1)^h$ forget gate's activation vector;
- $i_t \in (0,1)^h$ input/update gate's activation vector;

- $o_t \in (0,1)^h$ output gate's activation vector;
- $h_t \in (-1,1)^h$ hidden state vector, also known as the output vector of the LSTM unit;
- $\tilde{c_t} \in (-1,1)^h$ cell input activation vector;
- $c_t \in \mathbb{R}^d$: cell state vector.

$W \in \mathbb{R}^{h \times d}$, $U \in \mathbb{R}^{h \times h}$ and $b \in \mathbb{R}^h$ are weight matrices and bias vector parameters, which need to be learned during training. The indices d and h refer to the number of input features and number of hidden units.

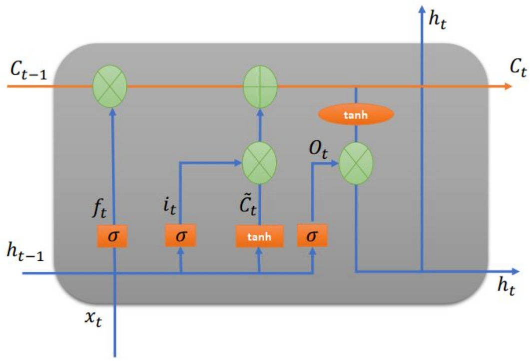


Figure 1. The cell structure of a long short-term memory unit.

3.2. Gated Recurrent Unit

The gated recurrent unit is a special type of optimized LSTM-based recurrent neural network [61]. The GRU internal unit is similar to the LSTM internal unit [62], except that the GRU combines the incoming port and the forgetting port in LSTM into a single update port. In [63], a new system called the multi-GRU prediction system was developed based on GRU models for the planning and operation of electricity generation.

The GRU was introduced by Cho et al. [64]. Although it was inspired by the LSTM unit, it is considered simpler to calculate and implement. It retains the LSTM immunity to the vanishing gradient problem. Its internal structure is simpler and, therefore, it is also easier to train, as less calculation is required to upgrade the internal states. The update port controls the extent to which the state information from the previous moment is retained in the current state, while the reset port determines whether the current state should be combined with the previous information [64].

Figure 2 shows the internal architecture of a GRU unit cell. These are the mathematical functions used to control the locking mechanism in the GRU cell:

$$z_t = \sigma(x_t W^z + h_{t-1} U^z + b_z) \tag{7}$$

$$r_t = \sigma(x_t W^r + h_{t-1} U^r + b_r) \tag{8}$$

$$\tilde{h}_t = tan(r_t \times h_{t-1} U + x_t W + b) \tag{9}$$

$$h_t = (1 - z_t) \times \tilde{h}_t + z_t \times h_{t-1} \tag{10}$$

where W^z, W^r, W denote the weight matrices for the corresponding connected input vector. U^z, U^r, U represent the weight matrices of the previous time step, and b_r, b_z and b are bias. The σ denotes the logistic sigmoid function, r_t denotes the reset gate, z_t denotes the update gate, and $\tilde{h}_t$ denotes the candidate hidden layer [65].

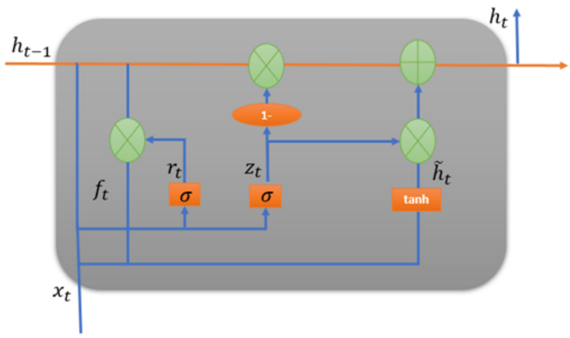


Figure 2. The cell structure of a gated recurrent unit.

It shows that the GRU has an updated port and a reset port similar to forget and input ports on the LSTM unit. The refresh port defines how much old memory to keep, and the reset port defines how to combine the new entry with the old memory. The main difference is that the GRU fully exposes its memory content using just integration (but with an adaptive time constant controlled by the update port).

Deep learning networks are very sensitive to hyperparameters. When the hyperparameters are incorrectly set, the predicted output will produce high-frequency oscillation [66]. Important hyperparameters for GRU network models are the number of hidden units in the recurrent layers, the dropout value, and the learning rate value.

Individually, these hyperparameters can significantly influence the performance of the LSTM or GRU neural models. Studies, such as [67,68], demonstrate how important the adjustment of hyperparameters is, as it optimizes the learning process and can present good results against more complex neural network structures.

3.3. Model Evaluation

In the present experiments, LSTM and GRU neural network models are compared. To evaluate the model prediction performance, the models used were root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), and mean absolute error (*MAE*). They are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y})^2} \tag{11}$$

where Y_t is the actual data value and $\hat{Y}$ is the forecast obtained from the model value. The prediction error is calculated as the difference between Y and $\hat{Y}$, i.e., the difference between the output desired and the output obtained. n is the number of samples used in the test set.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|Y_t - \hat{Y}_t| \tag{12}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|Y_t - \hat{Y}_t|}{|Y_t|} \tag{13}$$

4. Data Preparation

The present work is a continuation of previous work, where the data from the industrial press were already studied and analyzed using LSTM models [59]. The industrial presses are monitored by six sensors, with a sampling period of 1 min. The dataset contains data samples from 1 February, 2018 to December, 2020, for a total of 1004 days. The variables monitored are (1) electric current intensity (C. intensity); (2) hydraulic unit oil level (hydraulic unit level); (3) VAT pressure; (4) motor velocity (velocity); (5) temperature at the unit hydraulic (temperature at U.H.); and (6) torque.

Figure 3 shows the plot of the raw data. As the graph shows, there are zones of typical operation and spikes of discrepant data. Figure 4 is a Q–Q plot, showing the normality of the data. As the figure shows, the data are not homogeneous. There are many discrepant samples in the extreme quantiles and the distribution of data is not linear.

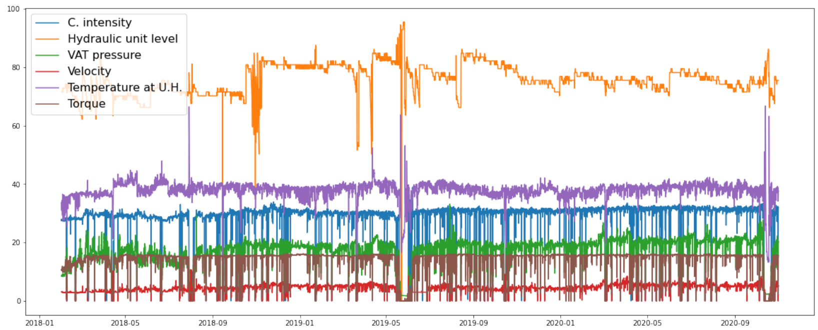


Figure 3. Plot of the sensor variables before applying data cleaning treatment. Many extreme values are visible for many variables, namely the hydraulic oil level and temperature.

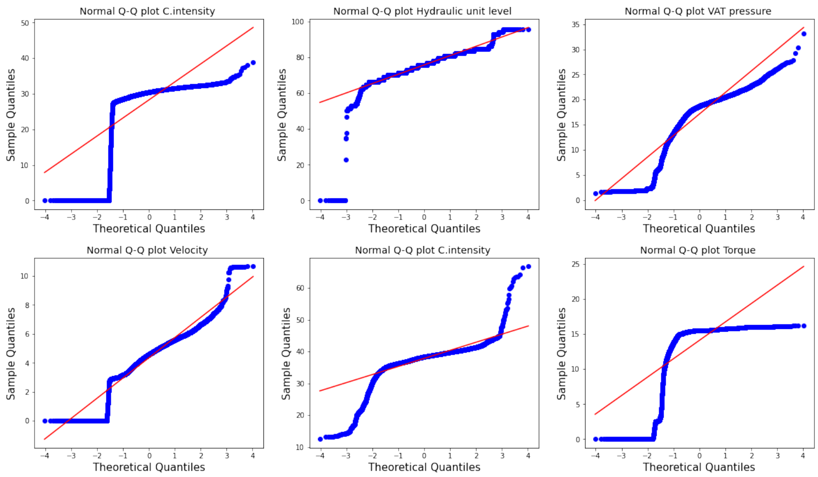


Figure 4. Q–Q plots of the sensor values before data cleaning treatment being applied.

Data quality is essential for developing effective modeling and planning. Data with discrepant values, as those shown in the charts, can pose difficulties to machine learning models. Therefore, data need to be processed and structured prior to analysis.

There are several treatment methods designed for this purpose, but a careful selection is needed so that information is not impaired. In the present work, the approach followed was the quantile method [59]. The quantile method removes extreme values, which are often due to sensor reading errors, stops, or other abnormal situations. After those samples are removed, it is possible to see more normal data distributions, such as those shown in Figures 5 and 6.

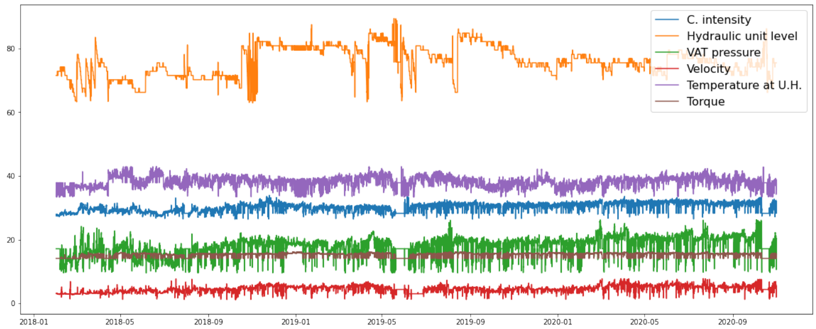


Figure 5. Sensor variables after applying data cleaning treatment. Many extreme values were removed.

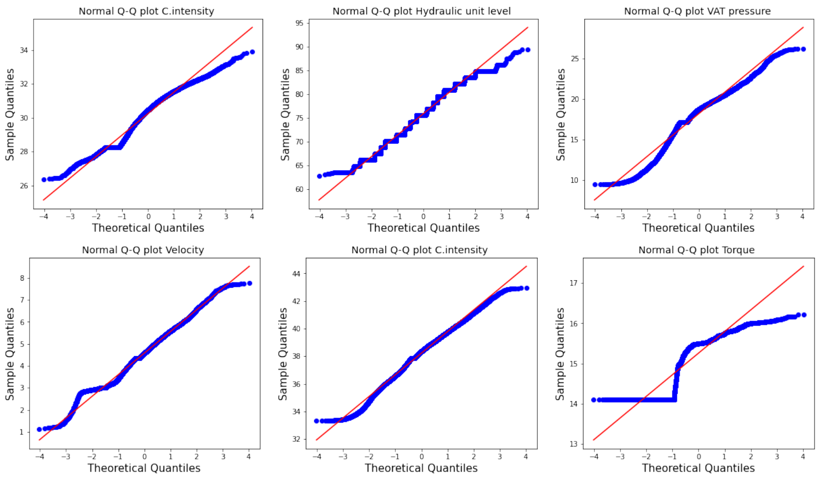


Figure 6. Q–Q plots of the sensor values after data cleaning treatment applied.

Since the present study relies on information that exists in the samples, this gives rise to the idea of presenting the correlation that exists between the variables. That information was condensed in the correlation matrix shown in Figure 7. As the figure shows, some of the correlations are interesting, such as those observed among the current, torque, and pressure. Other correlations are very low, such as those between oil level and temperature.

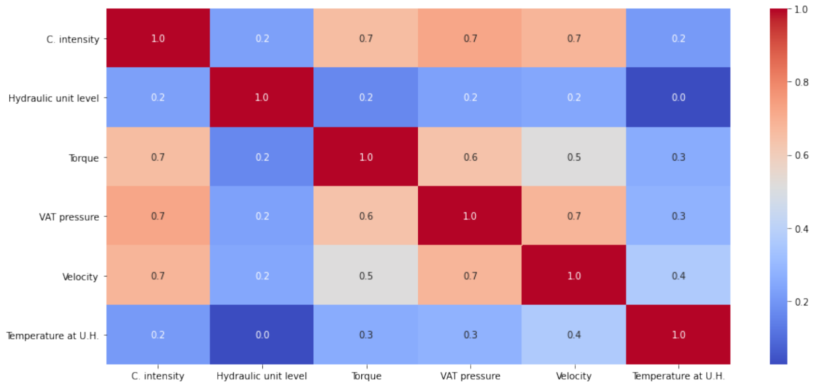


Figure 7. Correlation matrix, showing the correlation between all variables.

5. Methods

The present study aims to compare the performance of the LSTM model and the GRU model to predict future sensor values with 30 days advance, based on a window of past values. Experiments were performed using a computer with a third generation i5 processor, with 8 GB RAM. Previous work [59] shows that LSTM can make predictions with MAPE errors down to 2.17% for current intensity, 2.71% for hydraulic unit oil level, 2.50% for torque, 7.65% for VAT pressure, 16.88% for velocity, and 3.06% for temperature, using a window of 10 days and a sampling rate of two samples per day per sensor.

In the present work, different network architectures and hyperparameters were tested, for LSTM and GRU. In both cases, the networks rely on an encoding layer, a hidden layer of variable lengths, and an output layer. The internal architecture of the LSTM and GRU units are as shown in Figures 8 and 9.

The models were programmed in python, using the frameworks TensorFlow and Keras. For training, a batch size of 16 was used. Other hyperparameters, such as the activation function, when not indicated otherwise, are the TensorFlow and Keras default values.

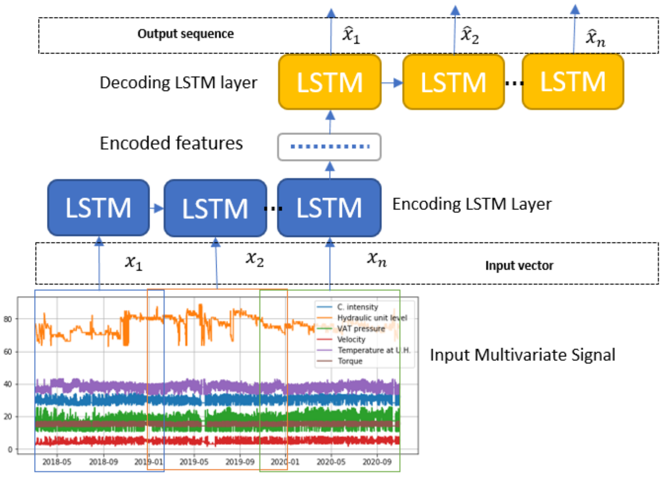


Figure 8. Base architecture of the LSTM model used.

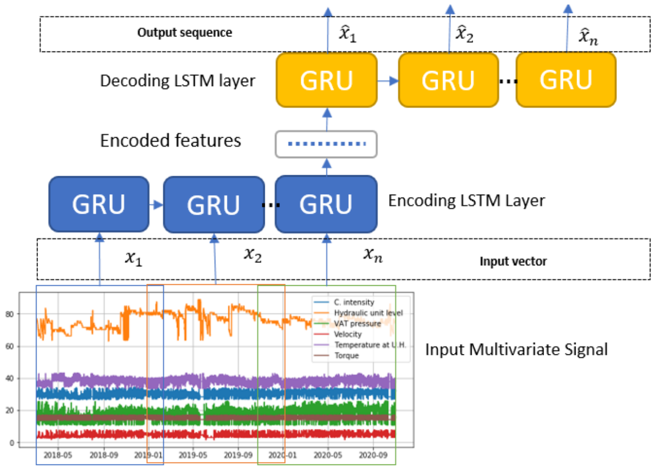


Figure 9. Base architecture of the GRU model used.

For the experiments, the dataset was divided into train and test subsets. The test set was used for validation during the training process and for final evaluation. The samples were not included in the training set. The training set consisted of 70% of the samples and the test set contained the remaining 30% of the samples.

Experiments were performed with different resampling rates. Using aggressive resampling, the size of the dataset is greatly reduced, which increases speed and decreases the influence of outliers in the data. However, for more precision, lower resampling rates must be used.

To determine the best size for the sliding window, experiments were performed, resampling to just one sample per day, which gave a total of 1004 samples, 70% of which were used for train and 30% for test. Experiments were also performed to determine the best resample rate, showing that using one sample per hour was a good compromise between the computation required and the performance of the model, as explained in Section 6.

Different experiments were performed to compare the performance of the LSTM and the GRU, with different sets of hyperparameters. The parameters were varied and tested one-by-one. Dense search methods, such as grid-search, were not used because of the processing time required.

6. Experiments and Results

Experimental work was performed to confirm the ability of the models to learn, and then to determine the optimal hyperparameters of the LSTM and GRU.

6.1. Testing the Convergence of the Learning Process

Figure 10 shows the learning curve of a GRU model, with 40 units in the hidden layer and window of 12 samples. The graph shows the loss measured in the train and in the test set. The learning process converges and takes less than 10 epochs to reach a small loss. This is similar to previous results obtained for the LSTM [59].

Although the learning curve shows that the model learns very quickly, in less than 10 epochs, in the following experiments, the number of epochs was limited to 15.

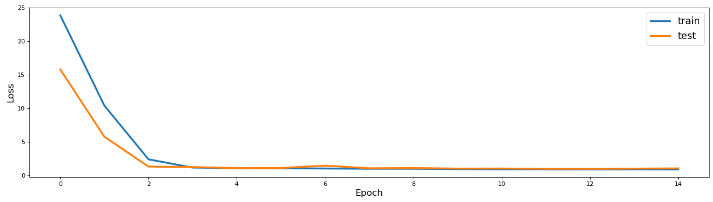
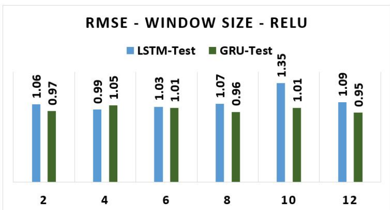


Figure 10. Learning curve of a GRU model, showing the loss measured in the train and test set during the first 14 epochs.

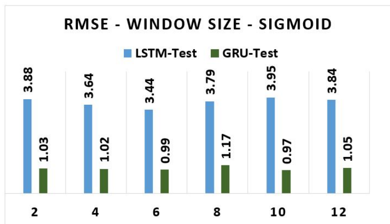
6.2. Experiments to Determine Model Performance with Different Window Sizes

The first experiment carried out, aimed to find the optimal window for the LSTM model and for the GRU model. The experiments were performed using one sample per day. Thus, the dataset had a total of 1004 samples. The models used had 40 units in the hidden layer.

Figure 11 shows the results of the two models, using different window sizes and two different activation functions in the output layer. The RMSE is the average of all the variables. As the charts show, the GRU is always better than the LSTM, regardless of the window size or activation function used. The window size only has a small impact on the performance of the model, being the differences minimal from two to 12 days. On the other hand, the results are better when the ReLU is used at the output layer. When the sigmoid function is used, the difference in performance between the GRU and the LSTM is larger than when the ReLU function is used.



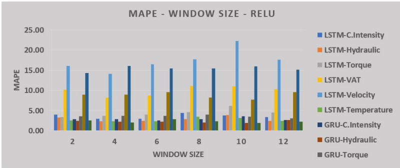
(**a**) Results with ReLU activation



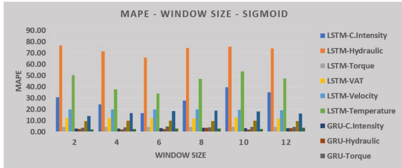
(**b**) Results with sigmoid activation

Figure 11. RMSE values for LSTM and GRU models, with different window sizes and activation functions for the output layer.

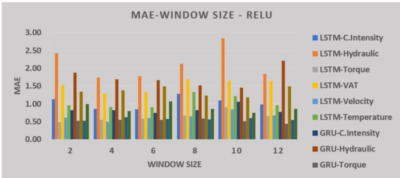
Figure 12 shows the MAPE and MAE associated with the 30 day forecast, for past windows of 2 to 12 days. The charts demonstrate that the LSTM architecture that uses a ReLU activation function in the output layer has lower errors. Using the sigmoid function, the LSTM errors are much larger. The GRU, however, in general performs better than the LSTM for all variables and activation functions. The prediction error results are much more stable for the GRU than they are for the LSTM. Table 1 shows exact error values for the best window sizes for the LSTM model. Table 2 shows the best window sizes for the GRU model.



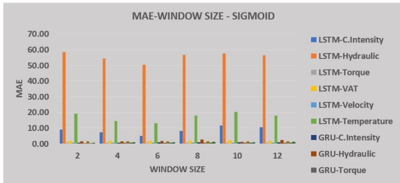
(**a**) MAPE using ReLU activation



(**b**) MAPE using sigmoid activation



(**c**) MAE using ReLU activation



(**d**) MAE using sigmoid activation

Figure 12. MAPE and MAE errors, for each variable, using ReLU and sigmoid activation functions, for window sizes of 2, 4, 6, 8, 10, and 12 days, using one sample per day. Exact values are shown in Tables 1 and 2 for the best window sizes.

Table 1. Summary of the best prediction errors obtained with the LSTM models. Window is the historical window size in days. AF is the output activation function.

			MAPE			
Window-AF	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
4-ReLU	2.95	2.32	3.68	8.28	14.06	2.38
6-Sigmoid	16.48	65.98	4.24	12.09	19.70	34.00
			MAE			
Window-AF	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
4-ReLU	0.86	1.74	0.54	1.29	0.50	0.91
6-Sigmoid	4.91	50.34	0.61	1.83	0.72	13.02

Table 2. Summary of the best prediction errors obtained with the GRU models. Window is the historical window size in days. AF is the output activation function.

			MAPE			
Window-AF	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
12-ReLU	3.63	1.95	3.53	7.74	15.99	1.92
10-Sigmoid	2.57	2.21	3.74	9.53	15.41	2.82
			MAE			
Window-AF	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
12-ReLU	0.77	2.20	0.44	1.49	0.55	0.86
10-Sigmoid	0.93	1.32	0.61	1.42	0.64	0.91

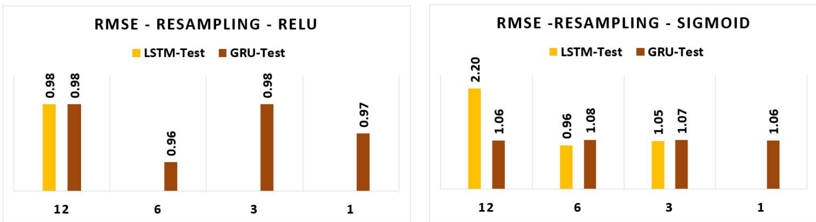
6.3. Experiments to Determine Model Performance with Different Resample Rates

In a second experiment, the models were tested with different resampling rates. Resampling is often used as a preprocessing method. Different techniques are used. Some of them are undersampling, in which the dataset size is reduced. This speeds up the data processing. In other cases, oversampling methods (such as data augmentation) are used in order to increase the number of samples.

In the present experiments, the dataset contains a large number of samples, so only undersampling techniques are necessary in order to reduce the number of data points. The method used was to average a number of samples, depending on the size of the dataset desired. Experiments were performed undersampling to obtain one sample per 12 h (two per day), one per six hours (four samples per day), one per each three hours, and finally one sample per hour. So the dataset size was greatly reduced.

The window sizes were the best of the previous experiments: a window size of 4 days for the LSTM and 12 days for the GRU, with the ReLU. A window size of 6 days for the LSTM and 10 days for the GRU, with the sigmoid.

Figure 13 shows the average RMSE errors for both models. As the results show, sometimes the LSTM overperformed the GRU, namely when using the sigmoid function with periods of six and three hours. However, the difference was not statistically significant. On the other hand, the GRU was able to learn in all the situations and the RMSE error was always approximately 1. So, the GRU is robust and accepts larger periods with minimal impact on the performance, while the LSTM model is much more unstable.



(**a**) Using ReLU function at the output layer (**b**) Using sigmoid at the output layer

Figure 13. RMSE value for LSTM and GRU model with ReLU and sigmoid at the output layer, for different undersampling rates: using one data point per 12 h, one per six hours, one per 3 h, and one per hour.

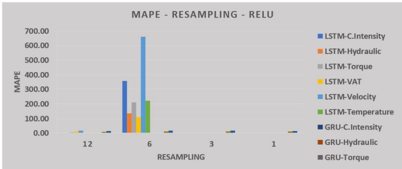
Figure 14 shows the MAE and MAPE errors calculated for each variable. It is possible to verify that, in general, the errors are much smaller with the sigmoid function. The LSTM model with the ReLU function is able to learn when a period of 12 h is used. When the sampling period is six hours, it seems the error gradient explodes for all variables and the errors become extremely large. For lower sampling periods, the LSTM does not learn. The GRU model continues to learn with acceptable errors. Table 3 shows the best results for the LSTM model with different resampling rates. Table 4 shows the best results for the GRU model with different resampling rates.

Table 3. Summary of the best prediction errors obtained with the LSTM models, using different resampling rates.

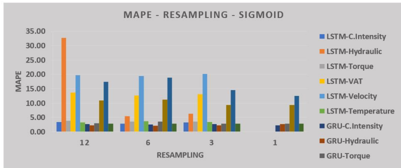
			MAPE			
Resampling-AF	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
12-ReLU	2.42	2.92	3.72	10.36	17.19	2.30
			MAE			
Resampling-AF	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
12-ReLU	0.71	2.22	0.55	1.57	0.64	0.88

Table 4. Summary of the best prediction errors obtained with the GRU models, using different resampling rates.

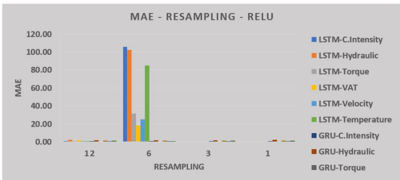
			MAPE			
Resampling-AF	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
1-ReLU	2.52	2.94	3.03	9.91	15.05	2.84
1-Sigmoid	2.22	2.72	2.88	9.29	12.42	2.74
			MAE			
Resampling-AF	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
1-ReLU	0.70	2.62	0.43	1.58	0.50	1.21
1-Sigmoid	0.65	1.99	0.43	1.41	0.48	1.03



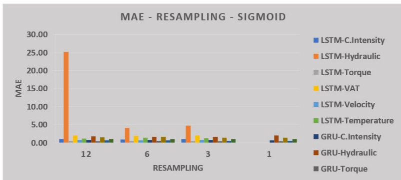
(**a**) MAPE using ReLU activation function



(**b**) MAPE using sigmoid activation function



(**c**) MAE using ReLU activation function



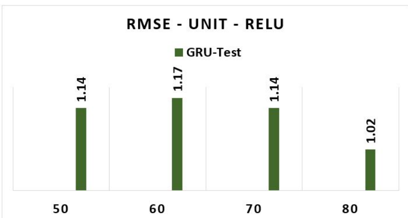
(**d**) MAE using sigmoid activation function

Figure 14. Results of the errors MAPE and MAE obtained with different undersampling rates, forecasting 30 days in advance.

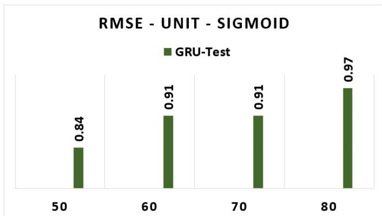
6.4. Experiments with Different Layer Sizes

An additional experiment was performed, to compare the performance of the models with different numbers of units in the hidden layer.

Using the GRU model, it is possible to learn with a larger number of samples, and with different variations of the model units, as shown in Figures 15 and 16. The LSTM was unable to learn with the resampling rate period of 1 h; therefore, results are missing. The window used in the experiments was 10 days for the sigmoid and 12 days for the ReLU, which were the optimal windows for the GRU using the ReLU and sigmoid functions, respectively.

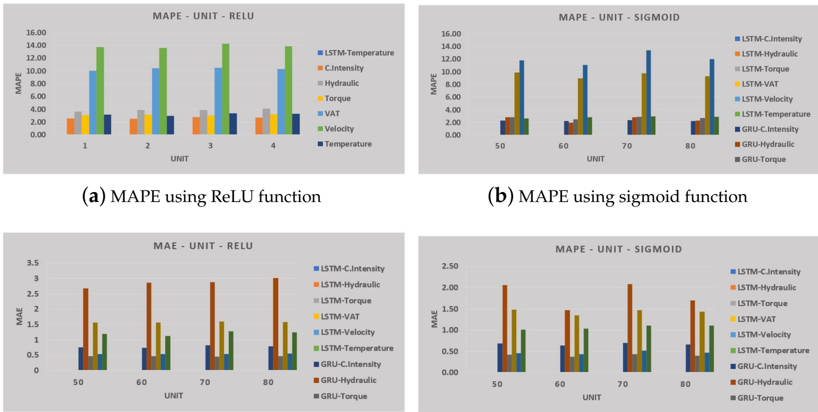


(**a**) Using ReLU function at the output layer



(**b**) Using Sigmoid at the output layer

Figure 15. RMSE errors measured, with different numbers of cells in the hidden layer.



(**a**) MAPE using ReLU function (**b**) MAPE using sigmoid function

(**c**) MAE using ReLU function (**d**) MAE using sigmoid function

Figure 16. MAPE and MAE obtained with different numbers of units in the hidden layer, measured when predicting future values 30 days in advance, with a resampling period of one hour. The LSTM was not able to learn, so the results are just for the GRU.

As the charts show, the GRU, using the sigmoid activation function, achieves the lowest RMSE error with 50 units in the hidden layer. Experiments described in Section 6.3 showed that the GRU with the same parameters, with 40 units in the hidden layer, had an RMSE error of 1.06. Table 5 shows the best results for the GRU model, after the tests with different numbers of cells in the hidden layer.

Table 5. Summary of the best results obtained with different numbers of units in the hidden layer.

			MAPE			
Unit	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
80-ReLU	2.66	4.09	3.19	10.31	13.83	3.29
50-Sigmoid	2.30	2.80	2.85	9.87	11.80	2.66
			MAE			
Unit	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
80-ReLU	0.78	3.02	0.47	1.58	0.55	1.25
50-Sigmoid	0.68	2.05	0.42	1.48	0.46	1.01

6.5. Comparing Many-to-Many and One-to-Many Architectures

An additional experiment was performed, in order to determine if the models are better trained to predict all the variables at the same time (one model, six outputs—many-to-many variables) or trained to predict just one variable (six models, one output each—many-to-one variable).

This experiment was just performed for the GRU, which presented the best results in the previous experiments.

According to the graphs presented in Figure 17, it is clear that architecture 'many-to-many' presents slightly better results. Therefore, there is no advantage in training one model to predict each variable.

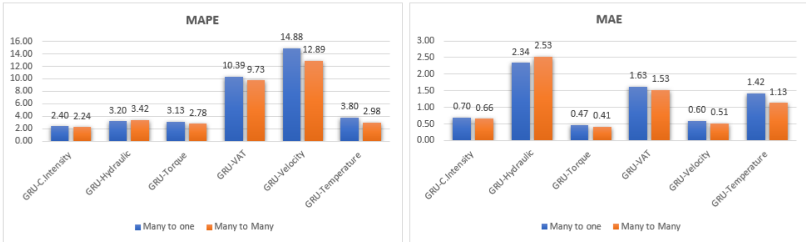


Figure 17. Comparison of the performance of the GRU models, trained to predict many-to-many and many-to-one variables.

6.6. Tests with Different Activation Functions in the Hidden Layer

An additional step was to test combinations of different activation functions, for the hidden and output layers of the GRU. The activation functions tested were sigmoid, hyperbolic tangent (tanh), and ReLU. Figure 18 shows a chart with the average RMSE of the models. Globally, ReLU in the hidden layer and tanh for the output are the best models, even though ReLU–sigmoid and ReLU–ReLU are closely behind.

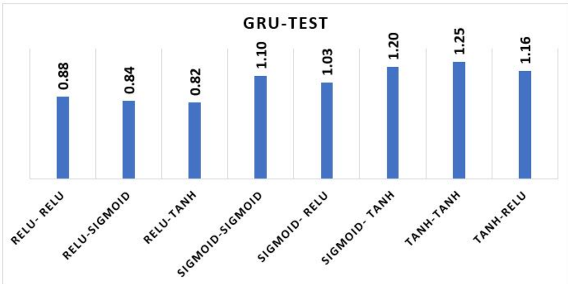


Figure 18. Average RMSE values, different types of activation functions.

Table 6 shows the RMSE error for the different combinations of activation functions, for each variable. As the table shows, different variables may benefit from different functions, although, in general, a first layer of ReLU and a second layer of ReLU, sigmoid, or tanh are good choices.

The values shown in Table 6 are calculated for the raw output predicted. However, the raw output values have some sharp variations, which are undesirable for a predictive system. Therefore, the values were filtered and smoothed using a median filter. Figure 19 shows plots of selected results, where the signals and predictions were filtered with a rolling median filter, with a rolling window of 48 h. Table 7 shows the MSE errors calculated after smoothing. As the table shows, after smoothing, the prediction errors decrease.

Table 6. Average RMSE obtained for the six variables, with different activation functions, calculated after the values were smoothed with a median filter.

	RMSE					
Function	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
ReLU–ReLU	0.96	3.48	0.42	1.90	0.84	1.90
ReLU–Sigmoid	0.93	1.72	0.53	1.60	0.79	1.19
ReLU–Tanh	0.83	2.47	0.48	1.70	0.76	1.25
Sigmoid–Sigmoid	0.98	6.40	0.45	2.14	0.89	1.35
Sigmoid–ReLU	1.22	4.87	0.43	1.86	0.74	1.31
Sigmoid–Tanh	1.19	7.38	0.45	2.03	0.78	1.35
Tanh–Tanh	1.36	7.84	0.44	2.24	0.91	1.35
Tanh–ReLU	0.86	7.3	0.42	1.91	0.76	1.41

Table 7. Average RMSE obtained for the six variables after the average clean method, with different activation functions using the GRU model.

	RMSE					
Function	**C. Intensity**	**Hydraulic**	**Torque**	**Pressure**	**Velocity**	**Temperature**
ReLU–ReLU	0.71	3.33	0.28	1.36	0.66	0.80
ReLU–Sigmoid	0.61	1.58	0.39	1.08	0.61	0.78
ReLU–Tanh	0.54	2.33	0.35	1.13	0.54	0.82
Sigmoid–Sigmoid	0.73	6.36	0.30	1.70	0.68	0.94
Sigmoid–ReLU	1.03	4.80	0.28	1.32	0.50	0.89
Sigmoid–Tanh	0.98	7.35	0.29	1.53	0.54	0.94
Tanh–Tanh	1.18	7.81	0.29	1.80	0.70	0.96

Figure 19 shows examples of plots of different prediction lines in part of the test set. As the results show, in some cases the ReLU–tanh combination is the best, while in other cases, the ReLU–sigmoid offers better performance. The ReLU–tanh combination is better, in general, but in the case of temperature, the sigmoid output shows the best performance.

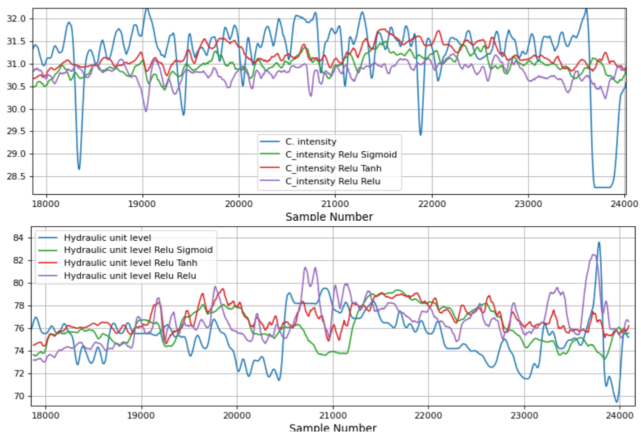


Figure 19. *Cont.*

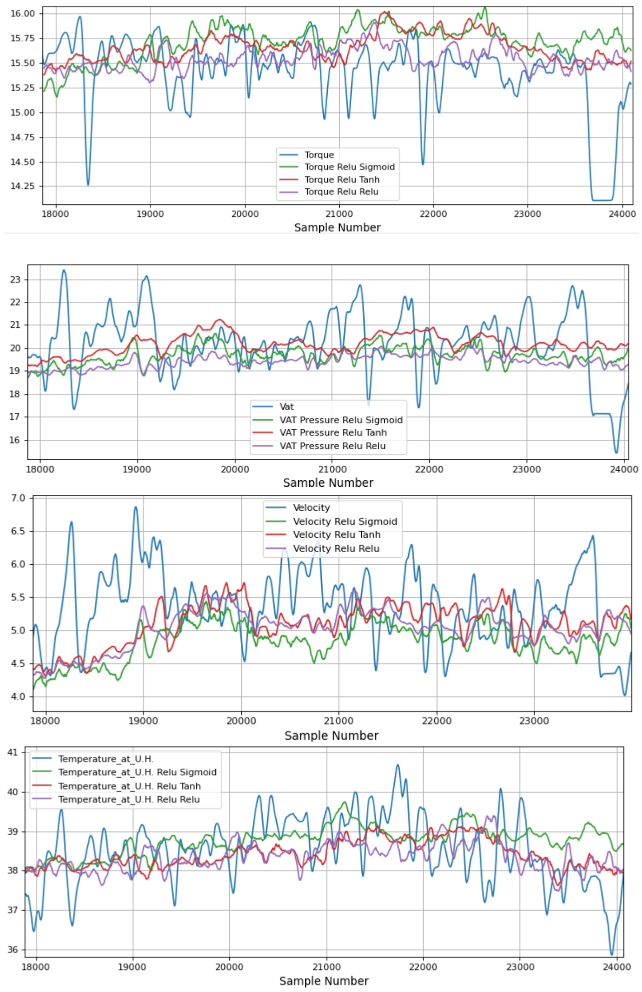


Figure 19. Plot of the predictions with different combinations of activation functions.

7. Discussion

Based on studies presented in the state of the art, it is possible to verify the usefulness of deep networks for prediction in time series variables. The area of prediction using deep neural networks has grown fast, due to the development of new models and the evolution of calculation power. LSTM and GRU models are two of the best forecast models. They have gained popularity recently, even though most of the state-of-the-art models are more traditional architectures.

The GRU network is simpler than the LSTM, supports higher resampling rates, and it can work on smaller and larger datasets. The experiments performed showed that the best results are based on the GRU neural network: it is easier and faster to train and achieve good results. A GRU network, with encoding and decoding layers, is able to forecast future behavior of an industrial paper press, 30 days in advance, with MAPE in general less than 10%.

An optimized GRU model offers better results with a 12-day sampling sliding window, with a sampling period of 1 h, and 50 units in the hidden layer. The best activation functions depend on the model. However, the ReLU–tanh is perhaps one of the best models, on average.

The results also demonstrate that training the models using just one output variable, thus optimizing a model for each variable separately, is not advantageous when compared to training one model to predict all six variables at the same time.

The present work shows that a GRU network, with encoding and decoding layers, can be used to anticipate future behavior of an industrial paper press. It shows better overall performance, with less processing requirements, when compared to an equivalent LSTM model. To the best of the authors' knowledge, this is the first time such a study has been made. The prediction errors are smaller than those presented by the LSTM neural network and the GRU is more immune to exploding or vanishing gradient problems, so it learns in a wider range of configurations.

Compared to the literature, previous research has shown that the GRU is often the best predictor [69–71]. However, those studies were performed for univariate data only. The present work uses six variables in a time series and compares the multivariate and the univariate models. In [72], the model that presents the lowest RMSE is the ARIMA. However, that is just for a small dataset and forecast with 6 samples advance. In [44,73], forecasting models with LSTM, including encoding and decoding, are proposed, although not compared to GRU.

8. Conclusions

In the industrial world, it is important to minimize downtime. Equipment downtime, due to failure or curative maintenance, represents hours of production lost. To solve this problem, predictive maintenance is, nowadays, the best solution. Artificial intelligence models have been employed, aimed at anticipating the future behavior of machines and, therefore, avoiding potential failures.

The study presented in this paper compares the performance of LSTM and GRU models, predicting future values of six sensors, installed at an industrial paper press 30 days in advance.

The GRU models, in general, operate with less data and offer better results, with a wider range of parameters, as demonstrated in the case study based on pulp presses.

Future work will include testing the performance of the GRU with different time gaps, in order to determine the best performance for different time gaps.

Author Contributions: Conceptualization, J.T.F., A.M.C., M.M.; methodology, J.T.F. and M.M.; software, B.C.M. and M.M.; validation, J.T.F., M.M., R.A.; formal analysis, J.T.F. and M.M.; investigation, B.C.M. and M.M.; resources, J.T.F., A.M.C. and M.M.; writing—original draft preparation, B.C.M.; writing—review and editing, J.T.F., R.A. and M.M.; project administration, J.T.F. and A.M.C.; funding acquisition, J.T.F. and A.M.C. All authors have read and agreed to the published version of the manuscript.

Funding: Our research, leading to these results, has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowvska-Curie grant agreement 871284 project SSHARE and the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under project POCI-01-0145-FEDER-029494, and by national funds through the FCT—Portuguese Foundation for Science and Technology, under projects PTDC/EEI-EEE/29494/2017, UIDB/04131/2020, and UIDP/04131/2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AF	activation function
ARIMA	autoregressive integrated moving average
NN	neural network
GRU	gated recurrent unit
LSTM	long short-term memory
MAE	mean absolute error
MAPE	mean absolute percentage error
RMSE	root mean square error
RNN	recurrent neural network

References

1. Bousdekis, A.; Lepenioti, K.; Apostolou, D.; Mentzas, G. A review of data-driven decision-making methods for industry 4.0 maintenance applications. *Electronics* **2021**, *10*, 828. [CrossRef]
2. Pech, M.; Vrchota, J.; Bednář, J. Predictive maintenance and intelligent sensors in smart factory: Review. *Sensors* **2021**, *21*, 1470. [CrossRef]
3. Martins, A.; Fonseca, I.; Farinha, J.T.; Reis, J.; Cardoso, A.J.M. Maintenance Prediction through Sensing Using Hidden Markov Models—A Case Study. *Appl. Sci.* **2021**, *11*, 7685.10.3390/app11167685. [CrossRef]
4. Hao, Q.; Xue, Y.; Shen, W.; Jones, B.; Zhu, J. *A Decision Support System for Integrating Corrective Maintenance, Preventive Maintenance, and Condition-Based Maintenance*; Construction Research Congress: Banff, AB, Canada, 2012; pp. 470–479.10.1061/41109(373)47. [CrossRef]
5. Chen, C.; Liu, Y.; Wang, S.; Sun, X.; Di Cairano-Gilfedder, C.; Titmus, S.; Syntetos, A.A. Predictive maintenance using cox proportional hazard deep learning. *Adv. Eng. Inform.* **2020**, *44*, 101054. [CrossRef]
6. Sherwin, D.J. Age-based opportunity maintenance. *J. Qual. Maint. Eng.* **1999**, *5*, 221–235. [CrossRef]
7. Bianchi, F.M.; De Santis, E.; Rizzi, A.; Sadeghian, A. Short-Term Electric Load Forecasting Using Echo State Networks and PCA Decomposition. *IEEE Access* **2015**, *3*, 1931–1943. [CrossRef]
8. Pati, J.; Kumar, B.; Manjhi, D.; Shukla, K.K. A Comparison Among ARIMA, BP-NN, and MOGA-NN for Software Clone Evolution Prediction. *IEEE Access* **2017**, *5*, 11841–11851. [CrossRef]
9. Akaike, H. Autoregressive Model Fitting for Control. In *Selected Papers of Hirotugu Akaike*; Parzen, E., Tanabe, K., Kitagawa, G., Eds.; Springer Series in Statistics; Springer: Berlin/Heidelberg, Germany, 1998; pp. 153–170. [CrossRef]
10. Ray, S.; Das, S.S.; Mishra, P.; Al Khatib, A.M.G. Time Series SARIMA Modelling and Forecasting of Monthly Rainfall and Temperature in the South Asian Countries. *Earth Syst. Environ.* **2021**, *5*, 531–546. [CrossRef]
11. Wang, K.; Wang, Y. How AI Affects the Future Predictive Maintenance: A Primer of Deep Learning. In *Advanced Manufacturing and Automation VII*; Wang, K., Wang, Y., Strandhagen, J.O., Yu, T., Eds.; Notas de aula sobre engenharia elétrica; Springer: Berlin/Heidelberg, Germany, 2018; pp. 1–9. [CrossRef]
12. Carvalho, T.P.; Soares, F.A.A.M.N.; Vita, R.; da Francisco, P.R.; Basto, J.P.; Alcalá, S.G.S. A systematic literature review of machine learning methods applied to predictive maintenance. *Comput. Ind. Eng.* **2019**, *137*, 106024. [CrossRef]
13. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.D. Machine learning in manufacturing: Advantages, challenges, and applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45. [CrossRef]
14. Soares, S.G. Ensemble Learning Methodologies for Soft Sensor Development in Industrial Processes. Ph.D. Thesis, Faculty of Science and Technology of the University of Coimbra, Coimbra, Portugal, 2015.
15. Shin, J.H.; Jun, H.B.; Kim, J.G. Dynamic control of intelligent parking guidance using neural network predictive control. *Comput. Ind. Eng.* **2018**, *120*, 15–30. [CrossRef]
16. Paolanti, M.; Romeo, L.; Felicetti, A.; Mancini, A.; Frontoni, E.; Loncarski, J. Machine Learning approach for Predictive Maintenance in Industry 4.0. In Proceedings of the 2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA), Oulu, Finland, 2–4 July 2018; pp. 1–6. [CrossRef]
17. Bangalore, P.; Tjernberg, L.B. An Artificial Neural Network Approach for Early Fault Detection of Gearbox Bearings. *IEEE Trans. Smart Grid* **2015**, *6*, 980–987. [CrossRef]
18. Sugiyarto, A.W.; Abadi, A.M. Prediction of Indonesian Palm Oil Production Using Long Short-Term Memory Recurrent Neural Network (LSTM-RNN). In Proceedings of the 2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS), Ipoh, Malaysia, 19 September 2019; pp. 53–57. [CrossRef]
19. Lara-Benítez, P.; Carranza-García, M.; Luna-Romera, J.M.; Riquelme, J.C. Temporal convolutional networks applied to energy-related time series forecasting. *Appl. Sci.* **2020**, *10*, 2322. [CrossRef]
20. Yeomans, J.; Thwaites, S.; Robertson, W.S.; Booth, D.; Ng, B.; Thewlis, D. Simulating Time-Series Data for Improved Deep Neural Network Performance. *IEEE Access* **2019**, *7*, 131248–131255. [CrossRef]
21. Yu, Z.; Moirangthem, D.S.; Lee, M. Continuous Timescale Long-Short Term Memory Neural Network for Human Intent Understanding. *Front. Neurorobot.* **2017**, *11*, 42. [CrossRef]

22. Aydin, O.; Guldamlasioglu, S. Using LSTM Networks to Predict Engine Condition on Large Scale Data Processing Framework. In Proceedings of the 4th International Conference on Electrical and Electronic Engineering (ICEEE), Ankara, Turkey, 8–10 April 2017; pp. 281–285. [CrossRef]
23. Dong, D.; Li, X.Y.; Sun, F.Q. Life prediction of jet engines based on LSTM-recurrent neural networks. In Proceedings of the 2017 Prognostics and System Health Management Conference (PHM-Harbin), Harbin, China, 9–12 July 2017; pp. 1–6. [CrossRef]
24. Baptista, M.; Sankararaman, S.; de Medeiros, I.P.; Nascimento, C.; Prendinger, H.; Henriques, E.M.P. Forecasting fault events for predictive maintenance using data-driven techniques and ARIMA modeling. *Comput. Ind. Eng.* **2018**, *115*, 41–53. [CrossRef]
25. Wang, J.; Zhang, T. Degradation prediction method by use of autoregressive algorithm. In Proceedings of the 2008 IEEE International Conference on Industrial Technology, Chengdu, China, 21–24 April 2008; pp. 1–6. [CrossRef]
26. Cruz, S.; Paulino, A.; Duraes, J.; Mendes, M. Real-Time Quality Control of Heat Sealed Bottles Using Thermal Images and Artificial Neural Network. *J. Imaging* **2021**, *7*, 24. [CrossRef]
27. Su, C.T.; Yang, T.; Ke, C.M. A neural-network approach for semiconductor wafer post-sawing inspection. *IEEE Trans. Semicond. Manuf.* **2002**, *15*, 260–266. [CrossRef]
28. Zhang, J.T.; Xiao, S. A note on the modified two-way MANOVA tests. *Stat. Probab. Lett.* **2012**, *82*, 519–527. [CrossRef]
29. Carnero, M. An evaluation system of the setting up of predictive maintenance programmes. *Reliab. Eng. Syst. Saf.* **2006**, *91*, 945–963. [CrossRef]
30. Bansal, D.; Evans, D.J.; Jones, B. A real-time predictive maintenance system for machine systems. *Int. J. Mach. Tools Manuf.* **2004**, *44*, 759–766. [CrossRef]
31. Ghaboussi, J.; Joghataie, A. Active Control of Structures Using Neural Networks. *J. Eng. Mech.* **1995**, *121*, 555–567. [CrossRef]
32. Bruneo, D.; De Vita, F. On the Use of LSTM Networks for Predictive Maintenance in Smart Industries. In Proceedings of the 2019 IEEE International Conference on Smart Computing (SMARTCOMP), Washington, DC, USA, 12–15 June 2019; pp. 241–248. [CrossRef]
33. Wang, L.; Hope, A.D. Fault diagnosis: Bearing fault diagnosis using multi-layer neural networks. *Insight-Non-Destr. Test. Cond. Monit.* **2004**, *46*, 451–455. [CrossRef]
34. Kittisupakorn, P.; Thitiyasook, P.; Hussain, M.; Daosud, W. Neural network based model predictive control for a steel pickling process. *J. Process Control* **2009**, *19*, 579–590. [CrossRef]
35. Yasaka, K.; Akai, H.; Kunimatsu, A.; Kiryu, S.; Abe, O. Deep learning with convolutional neural network in radiology. *Jpn. J. Radiol.* **2018**, *36*, 257–272. [CrossRef] [PubMed]
36. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
37. Ni, H.G.; Wang, J.Z. Prediction of compressive strength of concrete by neural networks. *Cem. Concr. Res.* **2000**, *30*, 1245–1250. [CrossRef]
38. Partovi, F.Y.; Anandarajan, M. Classifying inventory using an artificial neural network approach. *Comput. Ind. Eng.* **2002**, *41*, 389–404. [CrossRef]
39. Fonseca, D.; Navaresse, D.; Moynihan, G. Simulation metamodeling through artificial neural networks. *Eng. Appl. Artif. Intell.* **2003**, *16*, 177–183. [CrossRef]
40. Guo, Y.; Wu, Z.; Ji, Y. A Hybrid Deep Representation Learning Model for Time Series Classification and Prediction. In Proceedings of the 2017 3rd International Conference on Big Data Computing and Communications (BIGCOM), Chengdu, China, 10–11 August 2017; pp. 226–231. [CrossRef]
41. Liu, Y.; Duan, W.; Huang, L.; Duan, S.; Ma, X. The input vector space optimization for LSTM deep learning model in real-time prediction of ship motions. *Ocean Eng.* **2020**, *213*, 107681. [CrossRef]
42. Sakalle, A.; Tomar, P.; Bhardwaj, H.; Acharya, D.; Bhardwaj, A. A LSTM based deep learning network for recognizing emotions using wireless brainwave driven system. *Expert Syst. Appl.* **2021**, *173*, 114516. [CrossRef]
43. Wang, Q.; Bu, S.; He, Z. Achieving Predictive and Proactive Maintenance for High-Speed Railway Power Equipment with LSTM-RNN. *IEEE Trans. Ind. Inform.* **2020**, *16*, 6509–6517. [CrossRef]
44. Park, S.H.; Kim, B.; Kang, C.M.; Chung, C.C.; Choi, J.W. Sequence-to-Sequence Prediction of Vehicle Trajectory via LSTM Encoder-Decoder Architecture. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; pp. 1672–1678. [CrossRef]
45. Essien, A.; Giannetti, C. A Deep Learning Model for Smart Manufacturing Using Convolutional LSTM Neural Network Autoencoders. *IEEE Trans. Ind. Inform.* **2020**, *16*, 6069–6078. [CrossRef]
46. Soloway, D.; Haley, P.J. Neural generalized predictive control. In Proceedings of the 1996 IEEE International Symposium on Intelligent Control, Dearborn, MI, USA, 15–18 September 1996; pp. 277–282. [CrossRef]
47. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
48. Sak, H.; Senior, A.W.; Beaufays, F. Long sHort-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling. 2014. Available online: https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/43905.pdf (accessed on 20 September 2021).
49. Dahl, G.; Yu, D.; Deng, L.; Acero, A. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Trans. Audio Speech, Lang. Process.* **2012**, *20*, 30–42. [CrossRef]

Appendix E

IMPROVED GRU PREDICTION OF PAPER PULP PRESS VARIABLES USING DIFFERENT PRE-PROCESSING METHODS

Taylor & Francis LaTeX template for authors (Interact layout + American Psychological Association reference style)

Balduno Csar Mateus[1,2], Mateus Mendes[3,4], Jos Torres Farinha[3,5], Antnio Marques Cardoso[2], Rui Assis[1] and Hamzeh Soltanali [6]

[1]EIGeS—Research Centre in Industrial Engineering, Management and Sustainability, Lusfona University, Campo Grande, 376, 1749-024 Lisboa, Portugal;
[2]CISE—Electromechatronic Systems Research Centre, University of Beira Interior, Calada Fonte do Lameiro, 62001-001 Covilh, Portugal;
[3]Instituto Superior de Engenharia de Coimbra, Polytechnic of Coimbra, 3045-093 Coimbra, Portugal;
[4] Institute of Systems and Robotics, University of Coimbra, 3004-531 Coimbra, Portugal;
[5] Centre for Mechanical Engineering, Materials and Processes—CEMMPRE, University of Coimbra, 3030-788 Coimbra, Portugal; [6] Department of Biosystems Engineering, Ferdowsi University of Mashhad.

ARTICLE HISTORY
Compiled September 30, 2022

ABSTRACT
Predictive maintenance strategies are becoming increasingly more important with the increased needs for automation and digitalization within pulp and paper manufacturing sector. To that end, artificial intelligence-based prediction algorithms are more useful for supporting predictive mechanisms in data pre-processing applications of diagnosis and prognosis. Hence, this study contributes to examine the most efficient pre-processing approaches for predicting sensory data trends based on Gated Recurrent Unit (GRU) neural networks. To validate the model, the data from two paper pulp presses with several pre-processing methods are utilized for predicting the units conditions. The results of validation criteria show that pre-processing data using a LOWESS in combination with the Elimination of discrepant data filter achieves more stable results, the prediction error decreases, and the predicted values are easier to interpret. The model can anticipate future values wit MAPE, RMSE and MAE of 1.2, 0.27 and 0.30 respectively. The errors are below the significance level. Moreover, it is identified that the best hyperparameters found for each paper pulp press must be different. The proposed approach may be useful in resolving major future prediction problems, thereby increasing the availability of paper processing systems.

KEYWORDS
Deep Learning, LOWESS, Forecasting Failures, Industrial Press, Recurrent Neural Network, Predictive Maintenance.

balduino.mateus@ubi.pt
mmendes@isec.pt
torresfarinha@dem.uc.pt
ajmc@ubi.pt
ap1028@ulusofona.pt
ha.soltanali@mail.um.ac.ir

1. Introduction

1.1. *The new paradigm of predictive maintenance*

A good maintenance strategy aims to provide the best reliability, availability, safety and performance, with the lowest possible maintenance cost (Almeida Pais et al., 2021; Cline et al., 2017). In recent years, maintenance has gained more and more attention due to increasing demand for system safety and reliability, while at the same time the systems become increasingly more complex and commodities and labor become more expensive (Sherif and Smith, 1981). In the UK manufacturing industry, maintenance costs account for 12–23 % of the total plant operating costs (Cross, 1988).

The concept of Maintenance has been evolving from the corrective to the preventive maintenance, and from scheduled, to on-condition (condition monitoring), until the most recent concept of predictive. The predictive maintenance started with stochastic models. From that, evolved to algorithms based on Artificial Intelligence, namely with traditional Machine Learning and also Deep learning approaches.

The potential of artificial intelligence tools, especially machine learning, enables to improve system availability, reduce maintenance costs, improve operational performance and safety. It also supports decision making regarding the optimal time and action to perform maintenance interventions (Lv et al., 2021; Yam et al., 2001; Zhikun et al., 2013).

Maintenance activities play an important role in almost all areas of industry. Preventive maintenance has proven to be a great support when it comes to maximizing asset availability. It is fundamental for example to guarantee good availability of wind farms (Asgarpour and Sørensen, 2018; Canizo et al., 2017; Florea et al., 2012; Lei et al., 2015; Turnbull and Carroll, 2021; Udo and Muhammad, 2021), and also to improve, manufacturing capabilities in industry (Edwards et al., 1998; Lee et al., 2006; Spendla et al., 2017).

More recently, developments in hardware computational power and artificial intelligence algorithms make predictive maintenance possible. This has been achieved through some advances at the level of predictive maintenance tools, which aim to predict the variations that may occur in each period. Using those tools, the probability of failure can be estimated and many failures can be prevented through maintenance interventions, therefore increasing equipment availability and maintaining the production flow. Predictive maintenance has demonstrated its great effectiveness in anticipating problems of malfunction that could otherwise occur in the future. (Zhikun et al., 2013) use stochastic models for predictive maintenance of power transformers. (Rodrigues et al., 2021) use feed forward neural networks to predict future behavior of a paper press. (Mateus et al., 2021) do the same using LSTM and GRU networks.

As more sensors and data are available, prediction algorithms have become increasingly more popular in recent years. The connection with Big Data data storage technology is a relevant topic for possibly all industrial sectors. Machine learning shows good results in prediction with Big Data (L'Heureux et al., 2017; Qiu et al., 2016; Zhou et al., 2017). For the entertainment industry, for example, modern techniques are applied to get a good approximation and knowledge of their customers to propose more specific products, possibly customized to each customer.

1.2. *Industry 4.0 and IoT*

Industry 4.0, which is based mostly on the digitization of information, documents, and even assets, is facilitating the use of predictive maintenance because it is easier to acquire, store and share information, which in turn brings great benefits in developing strategies for dealing with anomalies that occur during the production process (Glistau and Coello Machado, 2018; Kalsoom et al., 2020).

Big data analytics, Autonomous Robots, Simulation, The Internet of Things (IoT), Cloud Computing, Additive Manufacturing, Augmented Reality and Cyber Security are the most important pillars in industry 4.0 (Erboz, 2017). Big data analysis can be used in different fields such as fault prediction to reduce the probability of error (Ji and Wang, 2017). In the case of maintenance, it is boosted due to the large amounts of data which are now possible to collect using network sensors.

The Internet of Things (IoT) is considered the future of the Internet, which allows machine-to-machine communication and learning (Balevi et al., 2018; Huang and Li, 2010).

It is on the basis of the modern sensor networks, which allow real time monitoring of modern industries. The IoT is presented as possibly the most important pillar of the fourth industrial revolution (Drath and Horch, 2014)

Machines can exchange data, perform data analysis, make decisions and perform operations without human intervention (Husain et al., 2014).

The Internet of Things (IoT) is presented as the most important pillar of the fourth industrial revolution (Drath and Horch, 2014).

The benefits of predictive maintenance include increased productivity, reduction of system errors (Dalzochio et al., 2020; Li et al., 2014) and minimization of unplanned downtime (Jezzini et al., 2013).

Maintenance 4.0 is about predicting future asset failures and ultimately determining the most effective preventive measures by applying advanced analytics techniques to Big Data about the technical condition, usage, environment, maintenance history and similar assets elsewhere and, in fact, anything that might correlate with an asset's performance.

1.3. *Data pre-processing and fault detection*

When data are collected, most of the times they come with discrepant data. That can be due to failure of the sensors themselves, events that happen in the environment or communication problems. The problem of dealing with discrepant data has been subject to heavy research and different treatment methods have been proposed, including different types of filters (Kim et al., 2017; Martins et al., 2020; Narendra et al., 2015).

Fault detection through machine learning techniques has provided additional benefits beyond improvements in risk mitigation and maximising system up time (Cline et al., 2017).

There are many machine learning techniques which can be used to detect failure patterns (for example, (Lykourentzou et al., 2009; Zibar et al., 2016), where the regression approach is used to predict numbers that can represent possible failures in the future state of the machine.), as well as predict future trends of the variables monitored, as in the present work.

1.4. *Research method*

Modern Artificial Intelligence (AI) methods are efficient in predicting machine failure, using different types of data (Jabeur et al., 2021; Yam et al., 2001). Therefore, predictive maintenance has attracted the attention of several scientific areas.

Predictive maintenance through artificial intelligence is a great way to overcome problems of unexpected machine breakdowns (Liu et al., 2018).

The literature search was conducted using the publications searched in Scopus, Web of Science, and ScienceDirect, as shown in Table 1.

Table 1. Total articles searched

Documents searched)	**Search (Scoup)**	**Search (WOS)**	**Search (ScienceDirect)**
Keywords		”Predictive Maintenance”	
Total of documents	2,587	3,308	2,730
Keywords		”Predictive Maintenance” ”Recurrent Neural Network”	
Total of documents	66	94	337
Keywords		”Predictive Maintenance”, ”Recurrent Neural Network”,”GRU”	
Total of documents	14	17	90
Keywords		”Predictive Maintenance”, ”Recurrent Neural Network”,”GRU”, ”Pre-Processing Methods”	
Total of documents	0	0	3
Keywords		”Predictive Maintenance”, ”Recurrent Neural Network”,”GRU”,” ”Pre-Processing Methods”, ”LOWESS	
Total of documents	0	0	0

The total number of articles associated with the keyword ”Predictive Maintenance” in the search engines presented above is 8625 articles, this number decreases to 497 when the keyword ”Recurrent Neural Network” is added. Adding the keyword ”GRU” decreases the total number of articles to 121, and adding the keyword ”Pre-Processing Methods” decreases the total number of articles to 3, and none of them uses the LOWESS method proposed in our research.

Table 2 lists the research results of the articles that use the techniques presented in the present work. Although the articles in the table have used similar techniques, they use a low sample rate, except one of the three, which also demonstrates the importance of the LOWESS technique. Additionally, the studies present limitations at the level of long-term prediction. They do not compare the performance of neural network architecture for different types of samples.

Table 2 lists the research findings of articles that use the same techniques presented in this article. Although the articles in the table use the techniques, they have a low sample rate, but one of the three is that it demonstrates the importance of the LOWESS technique. In addition, the studies have limitations at the long-term prediction level, i.e. they are not comparing the same neural network architecture for different types of samples.

Machine learning methods are useful for predictive maintenance, namely managing machine operations based on data collected by sensors. Those data contain patterns

Table 2. Most Relevant Article

Author	Focus	Concept Theoretical Model	Method	Sample	Findings
Dai et al. (2022)	Impact of data fluctuations on forecast accuracy	Data processing GRU and Random Forest	LOWESS Smoothing	Photovoltaic power generation	LOWESS smoothing can generate the smallest prediction error. Optimize the prediction performance of GRU model.
He et al. (2022)	Voltage Prediction	Auto-encoder based health indicator and LSTM network	LOWESS Smoothing	Voltage	GRU model is more suitable for the prediction of photovoltaic power generation. The method is more suitable for the hort-term forecasting han the medium and slong-term forecasting. Good prediction between different load profiles.
Wang et al. (2022)	Online useful life batteries prediction	Bi-LSTM	LOWESS Smoothing	Capacity(Ah)	The proposed online RUL prediction method proves to achieve better prediction results than LSTM, LSTM-AT, and Bi-LSTM models.

and information on phenomena that occur during the production process (Gorski et al., 2021; Zfle et al., 2021). The machine learning algorithms are able to discover those patterns using computational power, rather than human work, with minimal human intervention.

In the field of prediction, there are some typical machine learning algorithms, such as neural network models (Wang, 2003), deep random forest (Miller et al., 2017), genetic algorithms (Zhou et al., 2018), fuzzy logic (Couso et al., 2019), Bayesian algorithms (Tipping, 2003) and hidden Markov model algorithms (Martins et al., 2021), which have been applied in the diagnosis of dynamic device failures. Each of these models has its advantages with respect to the problems presented. For example, although multilayer neural networks and decision trees are two very different techniques for classification purposes, some researchers have conducted some empirical comparative studies (Eklund et al., 1998; Lim et al., 2000). Some general conclusions drawn in this work are:

(1) Neural networks are generally better at incremental learning than decision trees;
(2) The training time for a neural network is generally much longer than the training time for decision trees;
(3) Neural networks generally perform as well as decision trees, but rarely better.

The third point can be refuted by recent studies that report good performance of neural networks, even with optimized architecture (Schwenk and Bengio, 2000). Studies such as (Chong et al., 2004) use a combination of the two approaches to exploit their strengths.

The present work focuses on a supervised learning method, namely GRU neural network, to anticipate future trends of a number of variables. The GRU is in general accepted as one of the best models for prediction using multivariate data. The experiments were performed using sensor data acquired at an industrial paper pulp press. The main goal is to develop a model that can predict future sensor values, and therefore the state of the equipment, with at least 30 days advance, so that maintenance interventions can be planned and failures can be prevented. In previous work, the best prediction results were already obtained with the GRU model (Mateus et al., 2021). The encoder and decoder architecture with GRU unit to data from same press, called press number 2, and another press, called press number 4. Data pre-processing is done, both eliminating discrepant data and smoothing using the LOWESS filter to achieve more stable results.

The focus of this section is to present the contributions and objectives of this paper. Based on the literature, the current preprocessing approaches, although they are well known, are rarely used for this purpose, as well as the Gated Recurrent Unit (GRU) neural network. To validate the proposed model, the sensory data, from two paper pulp presses, are used. The data is composed of six variables: Current Intensity; Hydraulic Unit Oil Level; Torque; VAT Pressure; Rotation Velocity; Temperature at Hydraulic Unit. The results of this research contribute to adapt appropriate predictive policies to upgrade the operational reliability of paper processing systems. Therefore, the main objectives of this research are as follows: Review and survey of current AI-based predictive maintenance algorithms in processing industries; Develop a novel Gated Recurrent Unit (GRU) neural network for future predictive failure applications by comparing various pre-processing approaches; Validate the proposed model with sensory data from paper presses 2 and 4; Realization of the results to predict future failures as well as maintenance tasks in pulp industries.

Section 2 describes the theory of GRU recurrent networks, as well as the formulae used to calculate the different errors. Section 3 describes the method used to clean the dataset, prepare data and properties of some samples. Section 4 describes tests performed using the GRU neural network, results, and validation of the predictive models. Section 5 discusses the results and compares them to work already done. Section 6 draws some conclusions and highlights suggestions for future work.

2. Background and Methods

2.1. *LSTM and GRU neural networks*

Recurrent Neural Networks (RNN) are relatively popular for predictive maintenance tasks. They are one of the most efficient methods of prediction. They present a good performance at fault prediction based on data time series (Koprinkova-Hristova et al., 2011; Markiewicz et al., 2019; Nascimento and Viana, 2019; Rivas et al., 2019).

Wang et al. (2020) used a RNN for achieving predictive and proactive maintenance for high-speed railway power equipment. They also used a similar approach for IoT based predictive maintenance based on a Long Short-Term Memory (LSTM) RNN estimator. Chui et al. (2021) also used an RNN model for predicting remaining useful life of turbofan engines. According to the authors, the Root Mean Squared Error (RMSE) improved 12.95–39.32 % compared to existing works.

LSTM networks have also been used to predict the failure of air compressor motors (Tsibulnikova et al., 2019), induction furnaces (Choi et al., 2020), oil and gas equipment (Abbasi et al., 2019), and machine components such as bearings (Wu et al., 2020).

The studies conducted so far mostly refer to the type of encoder and decoder architecture using the recurrent neural network LSTM. The LSTM model is good and versatile for working with sequences. Nonetheless, it has many parameters and therefore it is hard to fine tune. The GRU is a simpler model, with less parameters and therefore easier to fine tune. According to Santra and Lin (2019), the GRU neural network can be called an LSTM optimized neural network. There is less research on using GRU models, although the GRU often produces better results than the LSTM in experimental work; In (Mateus et al., 2021), this alternative is proposed and its good long-term prediction capability is shown.

Introduced by (Cho et al., 2014), GRU aims to solve the vanishing gradient problem that comes with standard recurrent neural networks.These are the mathematical functions used to control the locking mechanism in the GRU cell:

$$z_t = \sigma(x_t W^z + h_{t-1} U^z + b_z) \tag{1}$$

$$r_t = \sigma(x_t W^r + h_{t-1} U^r + b_r) \tag{2}$$

$$\tilde{\boldsymbol{h}}_{\boldsymbol{t}} = tan(r_t \times h_{t-1} U + x_t W + b_h) \tag{3}$$

$$h_t = (1 - z_t) \times \tilde{h}_t + z_t \times h_{t-1} \tag{4}$$

Where,

- W^z, W^r, W are the weight matrices for the corresponding connected input vector;
- U^z, U^r, U the weight matrices of the previous time step;
- b_r, b_z and b_h are bias;
- x_t is the input vector;
- h_t is the output vector;
- $\tilde{h}_t$ is the candidate activation vector;
- z_t is the update gate vector;
- r_t is the reset gate vector.

Figure 1 shows a diagram of a GRU unit. The activation function is usually tanh or a sigmoid function. The GRU was developed as a solution for short-term memory. It has built-in mechanisms called gates that regulate the flow of information (Li et al., 2018; Zhang et al., 2021).

Figure 2 shows the scheme of the proposed method, with the function of extracting the data treatment by means of the two proposed methods, in order to have a predictive model with good predictive capacity. It is possible to predict patterns of failures in the variables of the presses.

2.2. *Model evaluation*

The Mean Absolute Percentage Error (MAPE) was used as a model performance measure. It is calculated according to Equation 5. It is a metric commonly used to estimate AI models' error and works best when there are no extremes in the data,

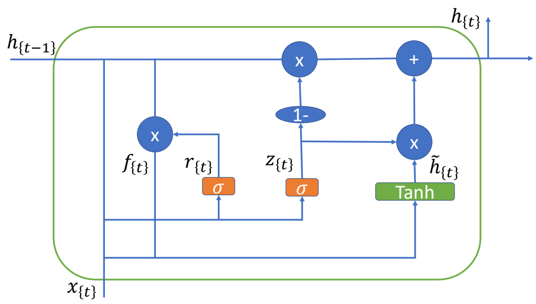


Figure 1. The cell structure of a Gated recurrent unit. (Mateus et al., 2021)

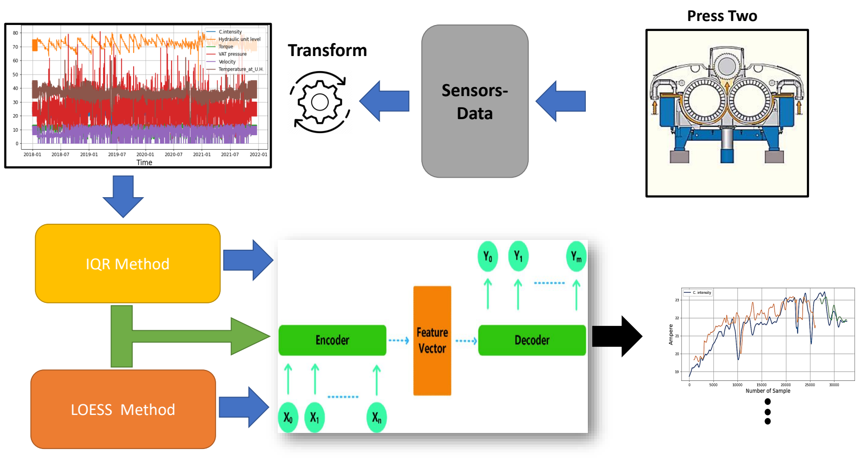


Figure 2. Diagram showing the data flow.

namely, zeros cannot exist in the actual output, so that the value of the fraction can be calculated.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|Y_t - \hat{Y}_t|}{|Y_t|} \tag{5}$$

Where:

- n is total number of observations;
- Y_t is the actual value;
- $\hat{Y}_t$ is the value predicted by the model.

Root Mean Square Error (RMSE) was also used to validate the results, which is given by the mathematical formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2} \tag{6}$$

The Mean Average Error (MAE), which evaluates the magnitude of the average error in a set of predictions without considering their direction, has also been used.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|Y_t - \hat{Y}_t| \tag{7}$$

3. Data Pre-processing

In order to ensure quality of data fed to the machine learning models, one of the first steps of the present study was the analysis and elimination of discrepant data which could interfere with the convergence of the learning algorithms. Two methods were used: the first was the Elimination of lower and upper extreme values, the second was based on smoothing using linear regression.

3.1. *Eliminating discrepant data*

The method of eliminating discrepant values is based on the idea that extreme values are most probably data reading failures. They often happen due to sensor failures, communication interference or other type of problems during data acquisition. As a result, the dataset sometimes contains invalid samples such as readings outside of the expected sensor ranges, or zero when the machine was stopped. Those samples can be eliminated, so that they do not negatively affect the machine learning process.

In the present work, limits were calculated for each variable and the samples out of the allowed range were replaced by the average. The limits were calculated using the following equations:

$$Q_{\frac{1}{4}} = \frac{1}{4}(n+1) \tag{8}$$

$$Q_{\frac{3}{4}} = \frac{3}{4}(n+1) \tag{9}$$

$$IQR = Q_{\frac{1}{4}} - Q_{\frac{3}{4}} \tag{10}$$

$$Down_{limit} = Q_{\frac{1}{4}} - K \times IQR \tag{11}$$

$$Up_{limit} = Q_{\frac{3}{4}} + K \times IQR \tag{12}$$

$Down_{limit}$ is the lower limit accepted for the variable, calculated by subtracting the constant k multiplied by IQR to $Q_{\frac{1}{4}}$. Up_{limit} is the upper limit accepted for the variable, calculated by adding the constant k multiplied by IQR to $Q_{\frac{3}{4}}$, where k is the constant of variation of the limits. The limits are calculated for each variable. Sample data points that contain values that are out of the interval $[Down_{limit}, Up_{limit}]$ are replaced by the average.

3.2. *Data smoothing*

LOWESS/LOESS (locally weighted/estimated scatterplot smoothing) is a non-parametric regression technique developed by Cleveland (Cleveland, 1981). Robust locally weighted regression is a method for smoothing variables, $(x_i, y_i), i = 1, \cdots, n$, in which the fitted value at z_k is the value of a polynomial fit to the data using weighted least squares, where the weight for (x_i, y_i) is large if x_i is close to x_k and small if it is not. The number of samples (n) used for each local approximation (z_k) is a parameter of the model. The degree of the polynomial function is also a parameter of the model. Often the polynomial degree is 1, which means a linear regression is performed.

Recent research has used the LOWESS smoothing technique in order to optimize the process of training and testing deep neural networks (Bury et al., 2021; Kulkarni et al., 2021). According to Phyo et al. (2019), LOWESS/LOESS procedure is used to overcome the problem of discrepant values. The study by (Jeenanunta et al., 2019) presents the influence that the LOWESS smoothing processing method has on the forecast errors of time series. According to Dai et al. (2022) all five different smoothing methods used in the study can improve the prediction performance of the GRU model. Among them, LOWESS smoothing can produce the smallest prediction error.

3.3. *Data before and after pre-processing*

The data set used in the present research contains samples from two paper pulp presses. The samples were collected through several sensors that are installed in the two presses, in a large industrial plant. The sensors read the following variables: i) Current Intensity: current absorbed by the press motor, in Ampere; ii) Hydraulic Unit Oil Level (in percentage); iii) Torque of the motor (in N.m); iv) VAT Pressure: Pressure inside the cuba (in KPa); v) Rotation Velocity: velocity of rotation of the press' rolls, in rotations per minute; vi) Temperature at Hydraulic Unit, in degree Celsius. There are nominal values for each of those variables, from the press manufacturer. Deviations from the expected intervals, which are related among them, may cause equipment failure.

A plot of the original data is shown in Figure 3. The samples were registered with sampling period of 1 min for press number 2 and 5 min for press number 4. For most many of the experiments the dataset was downsampled, in order to reduce processing time. The downsampling rate varied, although most of the time the 12 or 60 samples of each hour are averaged, which is equivalent to using a sampling period of 1 hour.

The original data contain many discrepant samples, shown as extremes values in Figure 3. There are spikes and sudden variations, which are mostly noise for the machine learning algorithms. Using the methods described in the previous subsections, most of the extremes are removed, specially the zeroes which were abundant and may be caused by reading errors or production line stops.

The discrepant data cleaning eliminates many extreme values. Nonetheless, the amplitude and frequency of variations still make the readings very unstable. Testing

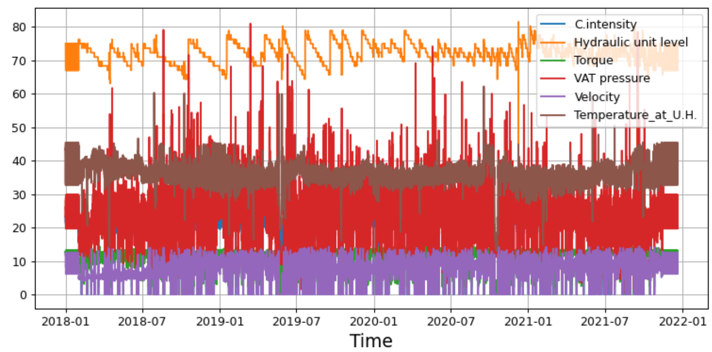


Figure 3. Plot of the variables for press number 4, before any data pre-processing. The variables contain a large amount of noise.

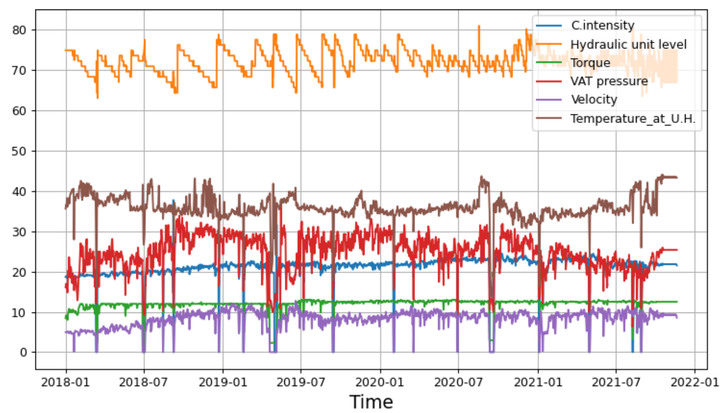


Figure 4. Plot of the variables for press number 4, after data pre-processing. The variables contain a low amount of noise.

the LOWESS method with a window size of 3 days it is possible to verify that in Figure 4 there is a significant reduction of the extreme values which were present in Figure 3, without affecting the trends that the data was showing. The trends are maintained and the variables are smoothed.

4. Case Study

4.1. *Analysis of correlations before and after pre-processing*

In order to have a better understanding on the impact of filtering the data using the LOWESS filter, an analysis of variable autocorrelation was performed. Figure 5 shows autocorrelations of the six variables before cleaning and applying the LOWESS filter. As the charts show, the correlations decay at a fast pace. The current intensity and torque, which are two very important variables, show autocorrelations of almost zero at 400 lags, which corresponds to 17 days. As for the variables VAT pressure, Hydraulic unit oil level, and Temperature, the correlation reaches almost zero at 500 lags, corresponding to 21 days. For velocity the decay happens at a slower pace, where the correlation is still about 0.1 at 1000 lags, corresponding to 42 days.

This shows that prediction with 30 days in advance is an ambitious goal, although not impossible, specially combining all variables into a multivariate model as done before (Mateus et al., 2021).

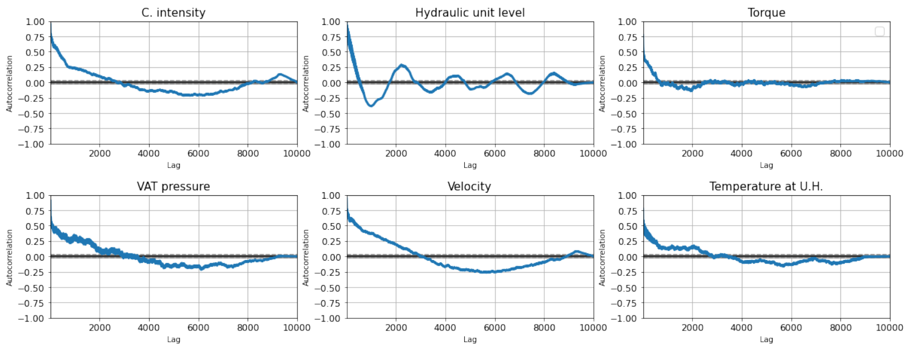


Figure 5. Variable autocorrelations, before cleaning and filtering the data.

Figure 6 shows the autocorrelations of the variables after data cleaning and filtering using the LOWESS method with 36 days window size. As the figure shows, the correlations for all variables have become larger than shown in Figure 5. The hydraulic unit oil level is the one with faster autocorrelation decay. The other variables show a good improvement, indicating better chances of small prediction errors.

4.2. *Prediction and comparison of the results*

For model validation the data were divided into two subsets. The training subset uses the first 80% of the total data and the test subset contains the remainder 20% of the data samples.

The purpose of the experiments is to find the best data preprocessing methods, neu-

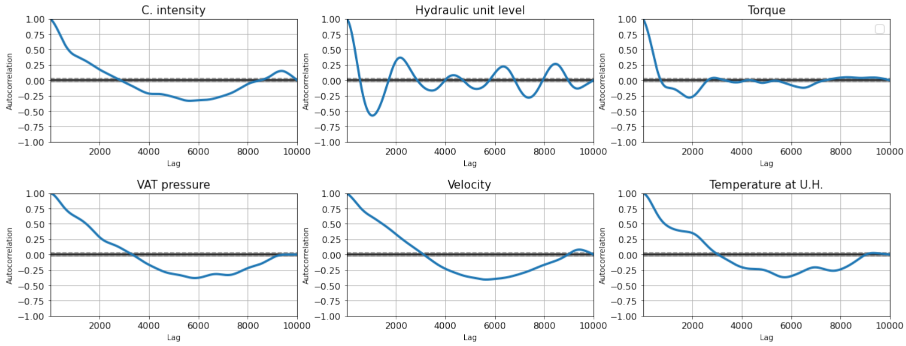
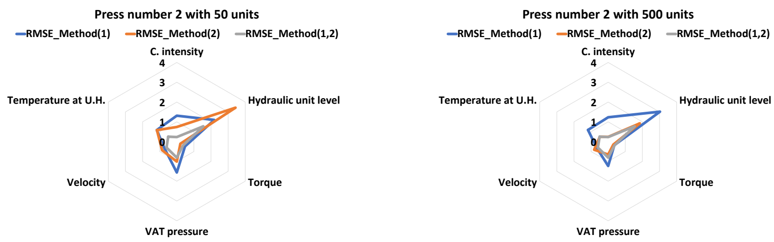


Figure 6. Autocorrelation for the all variables, obtained after cleaning and smoothing the data using LOWESS with 36 days window.

ral model architectures and hyperparameters that produce the best results predicting future behaviour of the paper pulp presses. The tests were performed using a GRU neural network with data encoder and decoder architecture, for it was the architecture that showed best results in previous work (Mateus et al., 2021).

Compared to LSTM models, GRU models have fewer parameters and simpler structures. (Gao et al., 2020) show that GRU models perform as well as LSTM models. (Mateus et al., 2021) show that GRU has a higher capacity in terms of the sampling rate.



(a) 50 GRU units

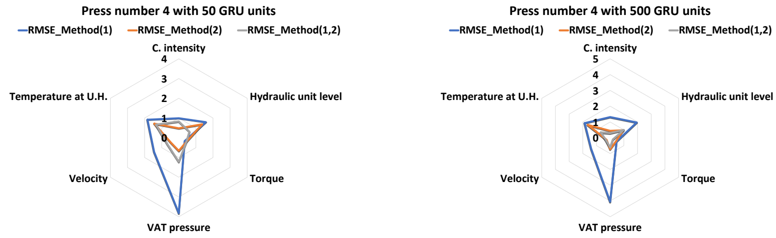
(b) 500 GRU units

Figure 7. RMSE of the best models for press 2, using the two different methods for pre-processing data, for the smaller and larger GRU networks. Method 1 only removes discrepant data. Method 2 smoothes the data using a LOWESS filter. Method (1,2) is the application of both. (a) prediction test with 50 GRU units with the two data processing methods, (b) prediction test with 500 GRU units with the two data processing methods.

The experiments aim at testing different pre-processing methods. Elimination of discrepant values is Method 1. Data smoothing using the LOWESS filter is Method 2. The combination of both—first the elimination of discrepant data, then smoothing—, is called Method (1, 2). The architecture of the neural network was the same for all the experiments, and it is the same that showed best results in previous work. Nonetheless, experiments were still performed with a smaller and faster GRU, with just 50 units, and a larger and slower network, with 500 units.

For press number 2, LOWESS method presented better results using a window of 5 days. The window size was halved because the number of data samples available from press 2 was too small for using larger windows. The dataset for press 4 contains 34800 hours of data, while the dataset for press 2 contains just 24096 hours of data.

Figure 7 shows the RMSE values of predictions for press 2, with the smaller and the larger GRU neural networks, with and without LOWESS filtering. As the figure shows, the prediction errors are much smaller when data are filtered. The difference is even more notorious in the larger network. For the same press and the same architecture, increasing the GRU units of the neural network to 500, it is verified that the combination of the methods leads to the same result, but with much smaller errors. The hydraulic variable in particular shows a larger error for both network structures.



(a) 50 GRU units

(b) 500 GRU units

Figure 8. RMSE for predictions of press 4 using the different data pre-processing methods. LOWESS filtering and 500 GRU units result in smaller RMSE errors. (a) prediction test with 50 GRU units with the two data processing methods, (b) prediction test with 500 GRU units with the two data processing methods.

For data originary from press number 4, the LOWESS filter presented better results using a window of 36 days. From the RMSE diagram in Figure 8, it can be seen that the results for press 4 also show much lower errors when the LOWESS filter is applied. The smaller model, with 50 GRU neural units, shows errors slightly larger than the larger model. For the same press using 500 GRU neural units, the RMSE errors are smaller, as demonstrated by the smaller area of the chart polygons.

Table 3. Prediction error results for 30 days advance forecast, using the two data preprocessing methods, removal of discrepant data and smoothing, for the 500 unit GRU and LOWESS with 5 days window, for press 2.

Prediction errors for press 2						
	C. Intensity	**Hydraulic**	**Torque**	**VAT**	**Velocity**	**Temperature**
MAPE	0.62	1.85	2.24	3.91	10.27	0.96
MAE	0.2	1.39	0.35	0.82	0.57	0.38
RMSE	0.23	1.55	0.37	0.95	0.6	0.5

Applying the two methods to press 2 data, it can be seen that while the errors in Table 3 are small, the important information are omitted from the graph in Figure 9, which is not good for possible press failure analysis.

Figure 10 shows the result of predicting the model with the better method of data processing for the press 4, which in this case falls on the intersection of the two methods. From the Table 4 it can be seen that the error is smaller.

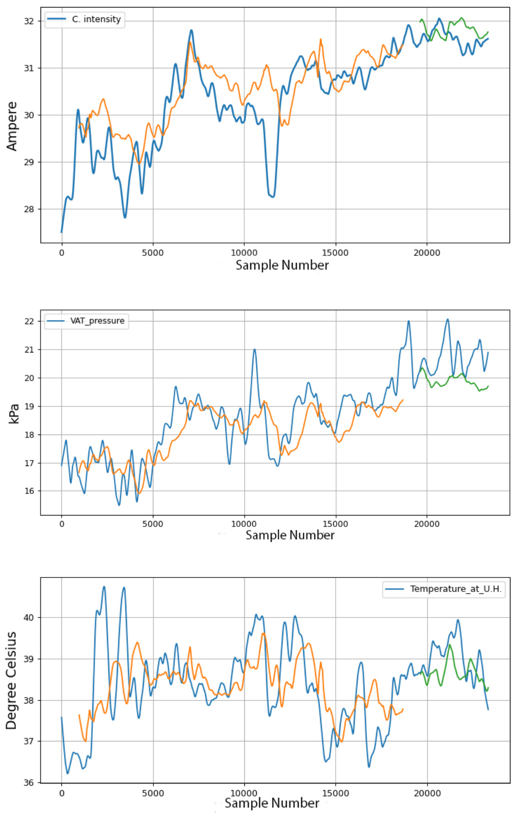


Figure 9. Signals and forecast results for press 2, with 30 day advance, using the two data processing methods, both removal of discrepant data and data smoothing using LOWESS filtering with 5 days window. The blue lines represent the actual value. The orange and green lines are predictions, respectively, in the train and test subsets.

Table 4. Prediction error results for 30 days advance forecast, using the two data preprocessing methods, removal of discrepant data and smoothing, for the 500 unit GRU and LOWESS with 36 days window, for press 4.

Prediction errors for press 4						
	C. Intensity	**Hydraulic**	**Torque**	**VAT**	**Velocity**	**Temperature**
MAPE	1.2	1.12	2.32	1.6	2.77	1.36
MAE	0.27	0.8	0.18	0.51	0.26	0.5
RMSE	0.30	1.00	0.20	0.61	0.30	0.69

5. Discussion

Data processing removing discrepant data simplifies the learning process of the RNN model and also leads to an improvement in the prediction results. The results obtained showed an improvement with data from both presses when discrepant data samples were replaced by the average. An analysis of autocorrelations shows that the use of data processing methods results in higher correlations for larger periods of time, when compared to untreated data as shown in Figure 9, and Figure 10.

In the literature review, no other studies were found to deal with forecast for industrial paper pulp presses using encoder-decoder architectures and recurrent neural units. The present work and comparative analysis of the results obtained for two industrial presses show that the architecture proposed is versatile and the same network architecture can be applied to both datasets, forecasting with acceptable errors after training. The larger architecture, using 500 GRU units, is slower and produces lower errors. The smaller architecture, with just 50 units, is faster and is still able to learn, although produces larger errors. Using data smoothed with the LOWESS filter, the learning process is highly facilitated. The prediction errors obtained in a 30 days advance forecast are smaller, with MAPE in general less than 10 %.

Compared to previous results (Mateus et al., 2021), the MAPE for the Current Intensity for press 2 decreased from 2.30% to 0.62%. For the Hydraulic oil level the MAPE decreased from 2.8% to 1.85%. For the Torque, the MAPE decreased from 2.85% to 2.24%. For the VAT pressure, the MAPE comes from 9.87% to 3.91%. For the Velocity, MAPE decreased from 11.8% to 10.27%. Finally, for the Temperature the MAPE decreased from 2.66% to 0.96%.

The quality of the results is confirmed visually in the charts, where the charts are in general easy to read and show the main trends of the variables.

In summary, we demonstrate that the approach done innovates, namely the following one: -The conjugation of Elimination of lower and upper discrepant values and LOWESS to data processing before inserting them in the NN, what proved to have better results than the other approaches described in the literature.

Additionally, the approach proposed can be adapted to other types of equipment, helping to solve prediction problems and contributing to increasing their availability.

6. Conclusions

In modern industries, prediction algorithms can anticipate future trends and contribute for better management decisions, namely in predictive maintenance. The results obtained in the present work demonstrate the applicability of recurrent neural networks (i.e., GRUs) in predicting future behavior in the paper press industry. The encoder and decoder architecture with GRU unit showed good results learning data from two different industrial pulp presses, and by applying the LOWESS technique the prediction errors decrease considerably, as described in Section 5.

Data pre-processing can play a very important role in improving the predictions. In the present work, filtering out discrepant data and smoothing using a LOWESS filter reduced the MAPE errors for all variables.

The results show that it is possible to forecast future behavior of industrial paper pulp presses up to 30 days in advance with good degree of certainty. That can be a good opportunity for optimizing maintenance decisions, reducing downtime and costs.

As limitations of the present approach, it must be referred that the method requires near real time operation, demanding high-speed networks and high power computation for monitoring the equipment and producing forecasts in advance. Additionally, the approach being based on machine learning algorithms produces only estimates with a degree of uncertainty.

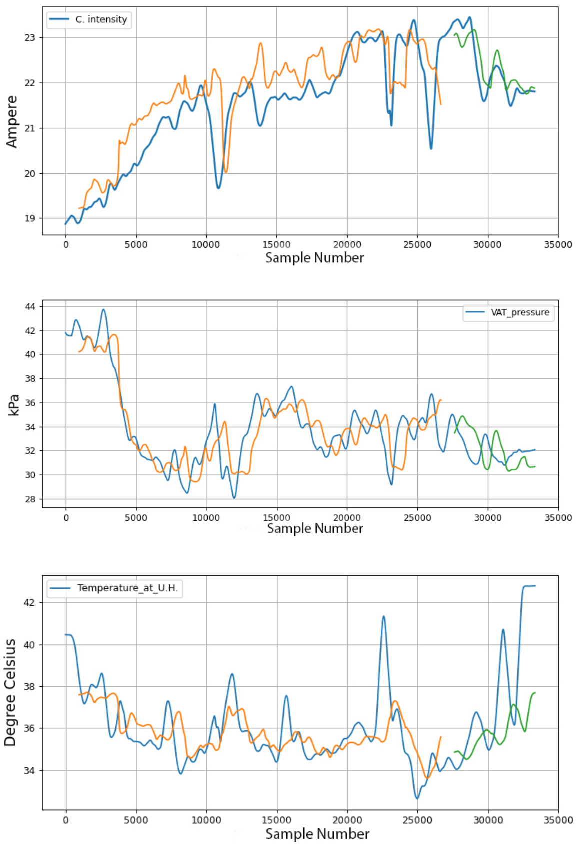


Figure 10. Signals and forecast results for press 4, with 30 day advance, using the two data processing methods, both removal of discrepant data and data smoothing using LOWESS filtering with 36 days window. The blue lines represent the actual value. The orange and green lines are predictions, respectively, in the train and test subsets.

In future work, other variables can be included in the study, namely through the inclusion of stock market variables in the model. These variables will aim to improve the predictive model, exploring the link between the stock market and the need for the production of the machines and their corresponding availability.

Abbreviations:

AI	Artificial Intelligence
ANN	Artificial Neural Networks
IoT	Internet of Things
IQR	Interquartile Range
NN	Neural Networks
GRU	Gated Recurrent Units
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
RNN	Recurrent Neural Networks
AF	Activation Function

Acknowledgments

The research leading to these results has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Sklodowvska-Curie grant agreement 871284 project SSHARE and the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under Project POCI-01-0145-FEDER-029494, and by National Funds through the FCTPortuguese Foundation for Science and Technology, under Projects PTDC/EEI-EEE/29494/2017, UIDB/04131/2020, and UIDP/04131/2020.

Conflict of interest

The authors declare that do not have any conflict of interest.

References

Abbasi, T., Lim, K. H., and Yam, K. S. (2019). Predictive maintenance of oil and gas equipment using recurrent neural network. *IOP Conference Series: Materials Science and Engineering*, 495:012067.

Almeida Pais, J. E. d., Raposo, H. D. N., Farinha, J. T., Cardoso, A. J. M., and Marques, P. A. (2021). Optimizing the life cycle of physical assets through an integrated life cycle assessment method. *Energies*, 14(19).

Asgarpour, M. and Sørensen, J. (2018). Bayesian based prognostic model for predictive maintenance of offshore wind farms. *International Journal of Prognostics and Health Management*, 9(1).

Balevi, E., Rabee, F. T. A., and Gitlin, R. D. (2018). Aloha-noma for massive machine-to-machine iot communication. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–5.

Bury, T. M., Sujith, R., Pavithran, I., Scheffer, M., Lenton, T. M., Anand, M., and Bauch, C. T. (2021). Deep learning for early warning signals of tipping points. *Proceedings of the National Academy of Sciences*, 118(39):e2106140118.

Canizo, M., Onieva, E., Conde, A., Charramendieta, S., and Trujillo, S. (2017). Real-time predictive maintenance for wind turbines using big data frameworks. *2017 IEEE International Conference on Prognostics and Health Management, ICPHM 2017*, pages 70–77.

Cho, K., van Merrienboer, B., Bahdanau, D., and Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *CoRR*, abs/1409.1259.

Choi, Y., Kwun, H., Kim, D., Lee, E., and Bae, H. (2020). Method of predictive maintenance for induction furnace based on neural network. *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 609–612.

Chong, M. M., Abraham, A., and Paprzycki, M. (2004). Traffic accident analysis using decision trees and neural networks. *International Journal of Information Technology and Computer Science*, 6:22–28.

Chui, K. T., Gupta, B. B., and Vasant, P. (2021). A genetic algorithm optimized rnn-lstm model for remaining useful life prediction of turbofan engine. *Electronics*, 10(3).

Cleveland, W. S. (1981). Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, 35(1):54–54.

Cline, B., Niculescu, R. S., Huffman, D., and Deckel, B. (2017). Predictive maintenance applications for machine learning. *Proceedings - Annual Reliability and Maintainability Symposium*.

Couso, I., Borgelt, C., Hullermeier, E., and Kruse, R. (2019). Fuzzy sets in data analysis: From statistical foundations to machine learning. *IEEE Computational Intelligence Magazine*, 14:31–44.

Cross, M. (1988). Raising the value of maintenance in the corporate environment. *Management Research News*, 11:8–11.

Dai, Y., Wang, Y., Leng, M., Yang, X., and Zhou, Q. (2022). Lowess smoothing and random forest based gru model: A short-term photovoltaic power generation forecasting method. *Energy*, 256:124661.

Dalzochio, J., Kunst, R., Pignaton, E., Binotto, A., Sanyal, S., Favilla, J., and Barbosa, J. (2020). Machine learning and reasoning for predictive maintenance in industry 4.0: Current status and challenges. *Computers in Industry*, 123:103298.

Drath, R. and Horch, A. (2014). Industrie 4.0: Hit or hype? [industry forum]. *IEEE Industrial Electronics Magazine*, 8(2):56–58.

Edwards, D. J., Holt, G. D., and Harris, F. (1998). Predictive maintenance techniques and their relevance to construction plant. *Journal of Quality in Maintenance Engineering*.

Eklund, P., of Intelligent Information Systems. v9 i1, A. H., and undefined 2002 (1998). A performance survey of public domain supervised machine learning algorithms. *researchgate.net*.

Erboz, G. (2017). How to define industry 4.0: main pillars of industry 4.0. *Managerial trends in the development of enterprises in globalization era*, 761:767.

Florea, G., Paraschiv, A., and Cimpoesu, E. (2012). Wind farm noise monitoring used for predictive maintenance. *IFAC Proceedings Volumes*, 45:1822–1827.

Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., and Lin, Q. (2020). Short-term runoff prediction with gru and lstm networks without requiring time step optimization during sample generation. *Journal of Hydrology*, 589:125188.

Glistau, E. and Coello Machado, N. I. (2018). Industry 4.0, logistics 4.0 and materials - chances and solutions. *Materials Science Forum*, 919:307–314.

Gorski, E. G., de Freitas Rocha Loures, E., Santos, E. A. P., Kondo, R. E., and Martins, G. R. D. N. (2021). Towards a smart workflow in cmms/eam systems: An approach based on ml and mcdm. *Journal of Industrial Information Integration*, page 100278.

He, K., Liu, Z., Sun, Y., Mao, L., and Lu, S. (2022). Degradation prediction of proton exchange membrane fuel cell using auto-encoder based health indicator and long short-term memory network. *International Journal of Hydrogen Energy*.

Huang, Y. and Li, G. (2010). Descriptive models for internet of things. In *2010 International Conference on Intelligent Control and Information Processing*, pages 483–486.

Husain, S., Prasad, A., Kunz, A., Papageorgiou, A., and Song, J. (2014). Recent trends in standards related to the internet of things and machine-to-machine communications. *Journal* *of information and communication convergence engineering*, 12(4):228–236.

Jabeur, S. B., Gharib, C., Mefteh-Wali, S., and Arfi, W. B. (2021). Catboost model and artificial intelligence techniques for corporate failure prediction. *Technological Forecasting and Social Change*, 166:120658.

Jeenanunta, C., Abeyrathna, K. D., Dilhani, M. H. M. R. S., Hnin, S. W., and Phyo, P. P. (2019). Time series outlier detection for short-term electricity load demand forecasting. *INTERNATIONAL SCIENTIFIC JOURNAL OF ENGINEERING AND TECHNOLOGY (ISJET)*, 2(1):3750.

Jezzini, A., Ayache, M., Elkhansa, L., Makki, B., and Zein, M. (2013). Effects of predictive maintenance(pdm), proactive maintenace(pom) & preventive maintenance(pm) on minimizing the faults in medical instruments. In *2013 2nd International Conference on Advances in Biomedical Engineering*, pages 53–56.

Ji, W. and Wang, L. (2017). Big data analytics based fault prediction for shop floor scheduling. *Journal of Manufacturing Systems*, 43:187–194.

Kalsoom, T., Ramzan, N., Ahmed, S., and Ur-Rehman, M. (2020). Advances in sensor technologies in the era of smart factory and industry 4.0. *Sensors*, 20(23).

Kim, D. Y., Jeong, Y. S., and Kim, S. (2017). Data-filtering system to avoid total data distortion in iot networking. *Symmetry 2017, Vol. 9, Page 16*, 9:16.

Koprinkova-Hristova, P. D., Hadjiski, M. B., Doukovska, L. A., and Beloreshki, S. V. (2011). Recurrent neural networks for predictive maintenance of mill fan systems. *International Journal of Electronics and Telecommunications*, 57:401–406.

Kulkarni, H., Thangam, M., and Amin, A. P. (2021). Artificial neural network-based prediction of prolonged length of stay and need for post-acute care in acute coronary syndrome patients undergoing percutaneous coronary intervention. *European Journal of Clinical Investigation*, 51(3):e13406.

Lee, J., Ni, J., Djurdjanovic, D., Qiu, H., and Liao, H. (2006). Intelligent prognostics tools and e-maintenance. *Computers in Industry*, 57(6):476–489. E-maintenance Special Issue.

Lei, X., Sandborn, P., Bakhshi, R., Kashani-Pour, A., and Goudarzi, N. (2015). Phm based predictive maintenance optimization for offshore wind farms. In *2015 IEEE Conference on Prognostics and Health Management (PHM)*, pages 1–8.

L'Heureux, A., Grolinger, K., Elyamany, H. F., and Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. *IEEE Access*, 5:7776–7797.

Li, C., Xie, C., Zhang, B., Chen, C., and Han, J. (2018). Deep fisher discriminant learning for mobile hand gesture recognition. *Pattern Recognition*, 77:276–288.

Li, H., Parikh, D., He, Q., Qian, B., Li, Z., Fang, D., and Hampapur, A. (2014). Improving rail network velocity: A machine learning approach to predictive maintenance. *Transportation Research Part C: Emerging Technologies*, 45:17–26. Advances in Computing and Communications and their Impact on Transportation Science and Technologies.

Lim, T. S., Loh, W. Y., and Shih, Y. S. (2000). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning 2000 40:3*, 40:203–228.

Liu, R., Yang, B., Zio, E., and Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 108:33–47.

Lv, Y., Zhou, Q., Li, Y., and Li, W. (2021). A predictive maintenance system for multi-granularity faults based on adabelief-bp neural network and fuzzy decision making. *Advanced Engineering Informatics*, 49:101318.

Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., and Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education*, 53(3):950–965.

Markiewicz, M., Wielgosz, M., Bochenski, M., Tabaczynski, W., Konieczny, T., and Kowalczyk, L. (2019). Predictive maintenance of induction motors using ultra-low power wireless sensors and compressed recurrent neural networks. *IEEE Access*, 7:178891–178902.

Martins, A., Fonseca, I., Farinha, J. T., Reis, J., and Cardoso, A. M. (2021). Maintenance prediction through sensing using hidden markov modelsa case study. *Applied Sciences 2021,* *Vol. 11, Page 7685*, 11:7685.

Martins, A. B., Farinha, J. T., and Cardoso, A. M. (2020). Calibration and certification of industrial sensors a global review. *WSEAS Transactions on Systems and Control*, 15:394–416.

Mateus, B. C., Mendes, M., Farinha, J. T., Assis, R., and Cardoso, A. M. (2021). Comparing lstm and gru models to predict the condition of a pulp paper press. *Energies*, 14(21).

Miller, K., Hettinger, C., Humpherys, J., Jarvis, T., and Kartchner, D. (2017). Forward thinking: Building deep random forests.

Narendra, N., Ponnalagu, K., Ghose, A., and Tamilselvam, S. (2015). Goal-driven context-aware data filtering in iot-based systems. *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pages 2172–2179.

Nascimento, R. G. and Viana, F. A. (2019). Fleet prognosis with physics-informed recurrent neural networks. *Structural Health Monitoring 2019: Enabling Intelligent Life-Cycle Health Management for Industry Internet of Things (IIOT) - Proceedings of the 12th International Workshop on Structural Health Monitoring*, 2:1740–1747.

Phyo, P. P., Jeenanunta, C., and Hashimoto, K. (2019). Electricity load forecasting in thailand using deep learning models. *International Journal of Electrical and Electronic Engineering & Telecommunications*, 8(4):221–225.

Qiu, J., Wu, Q., Ding, G., Xu, Y., and Feng, S. (2016). A survey of machine learning for big data processing. *Eurasip Journal on Advances in Signal Processing*, 2016:1–16.

Rivas, A., Fraile, J. M., Chamoso, P., Gonzlez-Briones, A., Sittn, I., and Corchado, J. M. (2019). A predictive maintenance model using recurrent neural networks. *Advances in Intelligent Systems and Computing*, 950:261–270.

Rodrigues, J. A., Farinha, J. T., Mendes, M., Mateus, R., and Cardoso, A. (2021). Short and long forecast to implement predictive maintenance in a pulp industry. *Eksploatacja i Niezawodnosc - Maintenance and Reliability*, 24:33–41.

Santra, A. S. and Lin, J.-L. (2019-01). Integrating long short-term memory and genetic algorithm for short-term load forecasting. *Energies*, 12(11):2040. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

Schwenk, H. and Bengio, Y. (2000). Boosting neural networks. *Neural Computation*, 12:1869–1887.

Sherif, Y. S. and Smith, M. L. (1981). Optimal maintenance models for systems subject to failurea review. *Naval Research Logistics Quarterly*, 28:47–74.

Spendla, L., Kebisek, M., Tanuska, P., and Hrcka, L. (2017). Concept of predictive maintenance of production systems in accordance with industry 4.0. In *2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pages 000405–000410.

Tipping, M. E. (2003). Bayesian inference: An introduction to principles and practice in machine learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3176:41–62.

Tsibulnikova, M. R., Pham, V. A., Aikina, T. Y., Xue-feng, L., Xiao-ben, L., Jan-ding, H., al, Abbasi, T., Lim, K. H., and Yam, K. S. (2019). Predictive maintenance of oil and gas equipment using recurrent neural network. *IOP Conference Series: Materials Science and Engineering*, 495:012067.

Turnbull, A. and Carroll, J. (2021). Cost benefit of implementing advanced monitoring and predictive maintenance strategies for offshore wind farms. *Energies*, 14(16).

Udo, W. and Muhammad, Y. (2021). Data-driven predictive maintenance of wind turbine based on scada data. *IEEE Access*, 9:162370–162388.

Wang, F.-K., Amogne, Z. E., Chou, J.-H., and Tseng, C. (2022). Online remaining useful life prediction of lithium-ion batteries using bidirectional long short-term memory with attention mechanism. *Energy*, 254:124344.

Wang, Q., Bu, S., and He, Z. (2020). Achieving predictive and proactive maintenance for high-speed railway power equipment with lstm-rnn. *IEEE Transactions on Industrial Informatics*, 16(10):6509–6517.

Wang, S.-C. (2003). Artificial neural network. *Interdisciplinary Computing in Java Program-*

ming, pages 81–100.

Wu, H., Huang, A., and Sutherland, J. W. (2020). Avoiding environmental consequences of equipment failure via an lstm-based model for predictive maintenance. *Procedia Manufacturing*, 43:666–673. Sustainable Manufacturing - Hand in Hand to Sustainability on Globe: Proceedings of the 17th Global Conference on Sustainable Manufacturing.

Yam, R. C., Tse, P. W., Li, L., and Tu, P. (2001). Intelligent predictive decision support system for condition-based maintenance. *The International Journal of Advanced Manufacturing Technology 2001 17:5*, 17:383–391.

Zhang, J., Zeng, Y., and Starly, B. (2021). Recurrent neural networks with long term temporal dependencies in machine tool wear diagnosis and prognosis. *SN Applied Sciences*, 3(4):1–13.

Zhikun, H., Bin, J., Linzi, Y., and Xiaolong, C. (2013). Predictive maintenance strategy of variable period of power transformer based on reliability and cost. *2013 25th Chinese Control and Decision Conference, CCDC 2013*, pages 4803–4807.

Zhou, C., Liu, X., Chen, W., Xu, F., and Cao, B. (2018). Optimal sliding mode control for an active suspension system based on a genetic algorithm. *Algorithms*, 11(12).

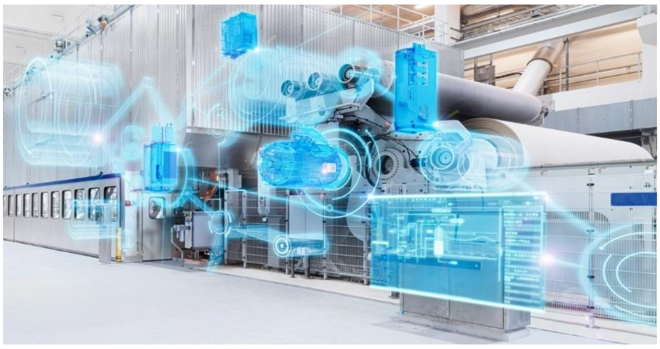
Zhou, L., Pan, S., Wang, J., and Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, 237:350–361.

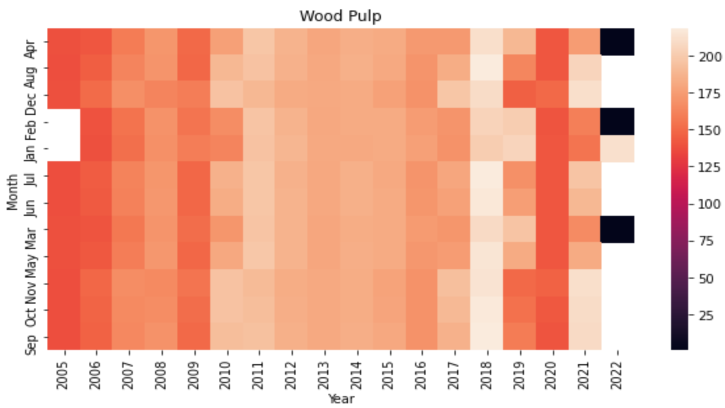
Zibar, D., Piels, M., Jones, R., and Scheffer, C. G. (2016). Machine learning techniques in optical communication. *Journal of Lightwave Technology*, 34(6):1442–1452.

Zfle, M., Moog, F., Lesch, V., Krupitzer, C., and Kounev, S. (2021). A machine learning-based workflow for automatic detection of anomalies in machine tools. *ISA Transactions*.

Appendix F

Otimização e Previsão da Produção Usando Redes Neuronais Recorrentes





Autor: Balduíno Mateus

31 de março de 2022

Resumo

O entendimento das variáveis que constituem o mercado das bolsas apresenta inúmeras vantagens pouco exploradas no âmbito industrial. Estas variáveis podem influenciar direta ou indiretamente o valor dos nossos produtos. Ao longo do tempo têm surgido novos desenvolvimentos no âmbito da gestão das operações, designadamente as que permitiram compreender e otimizar as dinâmicas de fabrico por via da redução do desperdício do tempo que se perde na manutenção das linhas de produção. A manutenção preditiva é fundamental para as indústrias modernas, a fim de melhorar a disponibilidade dos ativos físicos, a tomada de decisões e a racionalização dos custos. Isso exige a implementação de redes de sensores, armazenamento de dados e desenvolvimento de métodos de tratamento de dados que possam satisfazer a qualidade exigida nos modelos de previsão. O presente projeto tem como finalidade apresentar uma nova abordagem de monitorização da cadeia de produção por via da previsão da produção total das prensas de polpa de papel e também das previsões das falhas que possam ocorrer. Os dados foram obtidos por via de sensores instalados nas prensas de pasta de papel industrial para monitorizar o estado da mesma e prever a sua evolução com até 30 dias de antecedência, usando redes neuronais recorrentes *Gated Recurrent Unit* (GRU). O GRU é um dos modelos de inteligência artificial que tem produzido melhores resultados nos problemas de previsão, nomeadamente com séries temporais. No caso presente consegue antecipar valores futuros com o erro quadrático médio de seis variáveis entre 0.39 e 1.58.

Índice

1. Introdução

1.1. Enquadramento

A gestão da produção pretende otimizar recursos de produção, necessários para produzir produtos e serviços da qualidade desejada no mais curto espaço de tempo e ao menor custo. A implementação operacional desta função depende da cooperação de diferentes áreas técnicas com marketing, finanças, gestão e funções similares. Os recursos humanos e as ferramentas tecnológicas são um pré-requisito para assegurar esta coordenação e sinergias.

Os promotores do crescimento económico afirmam que a tecnologia acabará por conduzir a métodos de produção mais eficientes e a menos danos ambientais. Contudo, a maioria dos estudiosos da sustentabilidade afirmam que este pensamento está errado e que precisamos de mudar a nossa definição de prosperidade e ignorar a busca do crescimento económico sustentável (Jackson, 2016). Outos estudos analisam a forma como a reciclagem de papel na Europa pode ser expandida através de várias melhorias ao longo da cadeia de valor do papel (Blanco *et al*., 2013).

A tomada de decisão em tempo certo apresenta vantagens para a indústria, uma vez que uma decisão fora do tempo certo pode acarretar consigo custos indesejados. Por exemplo, a paragem não planeada do equipamento pode aumentar os custos da produção. Se antes deste acontecimento houvesse uma manutenção planeada, poderia ser evitada esta paragem indesejada. Para além dos custos, a tomada de decisão também pode afetar na eficiência e flexibilidade da produção dos bens.

Não há dúvida de que estamos a enfrentar a pior crise de sustentabilidade da história. Mudanças no clima, pobreza, poluição, escassez de água, e sobre consumo são algumas das preocupações de investigação, incluindo áreas como a ecologia, biologia, química, economia, sociologia, gestão, criatividade e inovação (Brem & Puente-Díaz, 2020); (Syren *et al*., 2021). Embora o papel seja um bom parceiro na luta contra o uso excessivo do plástico, não está isento de potenciais problemas que podem trazer para o ambiente, como estes estudos demonstram (Bloemhof-Ruwaard *et al*., 1996; Odada *et al*., 2004; Vaccari *et al*., 2005).

Uma produção mais equilibrada e tecnologicamente sustentável, baseada na gestão da disponibilidade face às aquisições no mercado pode ser uma das possíveis soluções. E, mais importante ainda, a gestão das instalações que produzem estes bens, porque a sua má gestão pode conduzir ao abate da instalação, e muitas destas instalações têm uma pegada ambiental significativa.

Durante o ciclo de vida de um bem tangível, ocorrem uma variedade de mudanças internas e externas. Por conseguinte, é importante desenvolver estratégias para apoiar a tomada de decisão. Enquanto algumas mudanças estão para além das previsões, outras, tais como alterações na legislação, impactos ambientais, e requisitos de produção, devem ser antecipadas. No entanto, a gestão de ativos baseia-se numa visão holística que oferece a possibilidade de evitar os acontecimentos mais imprevisíveis (Almeida Pais *et al*., 2021).

A monitorização do equipamento industrial é essencial para antecipar e evitar potenciais falhas, que podem pôr em perigo pessoas e bens. A manutenção preditiva visa tornar o processo mais eficiente, reduzindo a janela de tempo ideal para os procedimentos de manutenção. Usando dados sensoriais e algoritmos de previsão adequados, o estado do equipamento pode ser determinado e o tempo ideal para intervenções de manutenção pode ser previsto com alguma antecedência, evitando custos desnecessários e falhas por falta de manutenção.

Tal como discutido em (Sullivan *et al*., 2010), a manutenção preditiva pode reduzir os custos de manutenção em 25 %-35 %, eliminar falhas em 70 %-75 %, reduzir o tempo de paragem em 35 %-45 %, e aumentar a produção em 25 %-35 %. O estudo refere que estas percentagens não têm em conta aspetos importantes como a segurança do sistema e a imagem corporativa.

1.1. Objetivos do presente projeto

Algoritmos modernos, capacidade de armazenamento de dados e poder de computação, tornam possível não só analisar o comportamento passado, mas também antecipar o comportamento futuro de equipamentos industriais com razoável confiança (Bousdekis et al., 2021; Martins et al., 2020; Pech et al., 2021). Antecipar falhas futuras é, portanto, um tema que tem recebido cada vez mais atenção dos investigadores.

O presente projeto tem como finalidade demonstrar a correlação que existe entre as variáveis da prensa da produção de papel com fluidez de produção, também apresentar a inovação do monitorização dos valores cotados em bolsas em relação ao fabrico do papel, consumo do papel no mundo, com principal intuito de tornar a produção mais estável, ecológica, evitando grandes quantidades de stocks que por vezes acabam por se deteriorar, tornar a produção fiável, de modo a acompanhar as tendências do consumo do mercado, mas sem desperdiçar os recursos.

1.2. Estrutura do Documento

O presente projeto está organizado pelas seguintes secções:

A secção 2 objetivos, expetativa de contribuição para o ramo da engenharia em questão; A secção 3 conceitos teóricos da inteligência artificial nomeadamente as redes neuronais; A secção 4 redes neuronais recorrentes LSTM e GRU; A secção 5 desenvolvimento do projeto. A secção 6 aplicabilidade prática e perspetivas de futuro; A secção 7 conclusões.

2. Objetivos, expetativa de contribuição para o ramo da engenharia em questão

No processo de produção, a otimização não está apenas na primeira interação, como é apresentado no estudo (Monostori *et al*., 2016). O objetivo é otimizar ciclicamente o seu sistema de produção de modo a atingir os objetivos definidos no período de tempo pré-definido.

A abordagem de planeamento e programação de processos em circuito fechado na Figura 1 utiliza feedback dinâmico do calendário de produção e informação sobre a atual disponibilidade de recursos para gerar planos de processo. Em suma, a fase de planeamento comunica ao processo de programação a atual disponibilidade de máquinas no chão de fábrica.

Cada vez que uma operação é concluída na oficina, é gerada uma carga de trabalho baseada em recursos para determinar a próxima operação e atribuir os recursos necessários. Em geral, os departamentos de planeamento e programação do processo devem ser completamente reestruturados devido à necessidade de comunicação bidirecional em tempo real (K. Iwata and Y. Fukuda, 1989)

Figura 1 - Planeamento e Programação do Processo ClosedLoop (Zijm et al., 2019).

Tsang (2002) apresentou uma forma de visualizar o sistema de manutenção com base no estado do equipamento, carga operacional, ações de manutenção (estratégias), e objetivos comerciais. De acordo com isto, o estado do equipamento é afetado tanto pela carga operacional como pelas ações de manutenção. A Figura 2 apresenta a estrutura de produção, e o seu ambiente. Durante anos, a relação entre a produção e a manutenção foi considerada um antagonismo na tomada de decisões de gestão. Esta situação não se alterou porque os requisitos de escala de cada papel não estão alinhados (Weinstein & Chung, 1999).

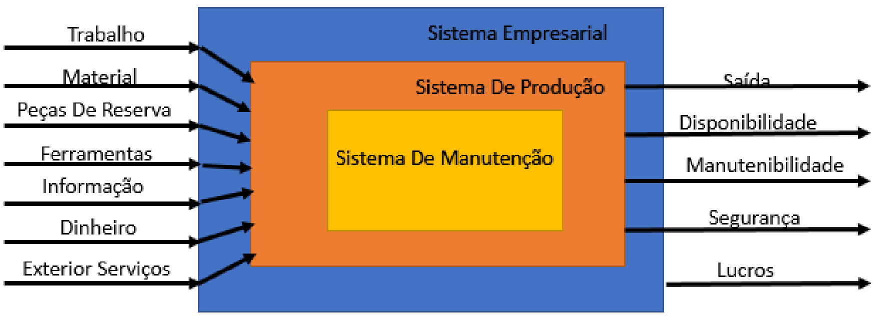


Figura 2 - Estrutura de produção, e seu entorno.

Um estudo mais detalhado da natureza das máquinas de papel pode ser encontrado em estudos (Holmberg *et al*., 2013) (Zvolinschi *et al*., 2006) (Stewart *et al*., 2003), nos quais mostram as perdas e danos que podem ocorrer nestas suas relações e a praticidade do problema torna-o desafiante e interessante (Xiao *et al*., 2016).

2.1. Manutenção preditiva

A manutenção preditiva pode ser descrita como uma estratégia de manutenção que visa determinar o momento exato em que a ação de manutenção efetiva deve ser desencadeada (Montero-Jiménez & Vingerhoeds, 2018).

A criação de um Programa de Manutenção Preditiva é uma decisão estratégica que até agora careceu de análise das questões relacionadas à sua instalação, gestão e controle. Carnero (2006) refere que a Manutenção Preditiva pode proporcionar um aumento na segurança, qualidade e disponibilidade nas indústrias. Bansal *et al*. (2004) apresentam um novo sistema de manutenção preditiva em tempo real para sistemas de máquinas baseado em redes neuronais. Outros estudos como (Bruneo & De Vita, 2019; Ghaboussi & Joghataie, 1995) indicam a viabilidade de redes neuronais artificiais para manutenção preditiva.

Em (Carvalho *et al*., 2019), é possível ver que cada abordagem proposta deste trabalho aborda um equipamento específico, pelo que se torna mais difícil compará-lo com outras técnicas. Ao mesmo tempo, é de notar que a própria manutenção preditiva está a tornar-se uma nova ferramenta para a gestão de eventos de serviço. A monitorização de todos os fluxos com estratégias ótimas de qualidade e manutenção como resultado de um sistema regulado pode permitir às empresas aumentar a sua rentabilidade e nível de serviço ao cliente (Gejo-García *et al*., 2022).

A monitorização de ativos físicos tornou-se uma prioridade para a manutenção preditiva. Estudos recentes comprovam a importância do tema (Aydin & Guldamlasioglu, 2017; Dong et al., 2017). (Sana, 2012) fala sobre um modelo baseado no inventário para gerir a produção imperfeita através de manutenção preventiva, de trabalho e garantia. Neste modelo de inventário, a produtividade é constante e a procura é determinista.

Muitas ferramentas estatísticas e de aprendizagem computacional têm sido usadas para fins de previsão, na monitorização e prevenção de falhas de equipamentos *(*Baptista *et al*., 2018) (J. Wang & Zhang, 2008), controle de qualidade (Cruz *et al*., 2021), bem como em outras áreas também (Chao-Ton Su *et al*., 2002).

Os métodos modernos de Inteligência Artificial (IA) são eficientes na previsão de falhas de máquinas, utilizando diferentes tipos de dados (Yam [i]*et al*., 2001) (Jabeur *et al*., 2021; Liu *et al*., 2018). Por conseguinte, a manutenção preditiva tem atraído a atenção de várias áreas científicas. A Figura 3 mostra a distribuição de documentos SCOPUS por área científica, ao procurar "manutenção preditiva" (https://www.scopus.com, verificado 2022-03-12).

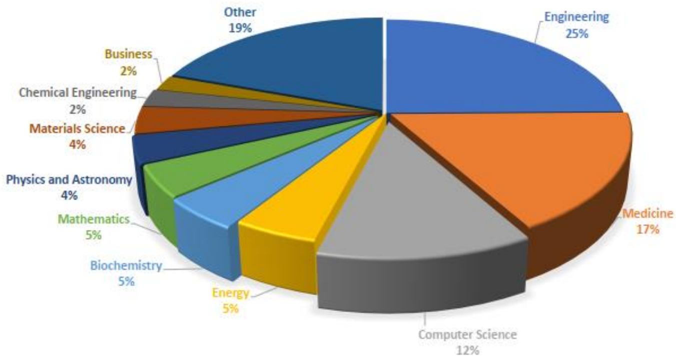


Figura 3 - Distribuição de documentos SCOPUS na área "preditivos manutenção".

Os métodos de aprendizagem computacional são úteis para a manutenção preditiva, nomeadamente a gestão de operações de máquinas com base em modelos que usam dados recolhidos por sensores. Esses dados contêm padrões e informações sobre fenómenos que ocorrem durante o processo de produção (Züfle *et al*., 2021)(Gorski *et al*., 2022). Os algoritmos de aprendizagem de máquinas são capazes de descobrir esses padrões utilizando poder computacional, reduzindo o esforço da mão de obra.

3. Conceitos teóricos

3.1. Inteligência artificial

Mais recentemente, entretanto, os métodos de Inteligência Artificial tornaram-se mais populares. Eles estão a ter impacto na sociedade, na política, na economia e nas indústrias (K. Wang & Wang, 2018), oferecendo ferramentas para análise de dados, reconhecimento de padrões e previsão. Esse impacto pode ser benéfico na manutenção preditiva e nos sistemas de produção.

Chen *et al*., (2003) aplica redes neuronais a um mercado financeiro emergente: previsão e negociação do índice de ações de Taiwan. Resultados empíricos mostram que as estratégias de investimento baseadas em redes neuronais probabilísticas (PNN) obtêm retornos mais elevados do que outras estratégias de investimento examinadas neste estudo. São também consideradas as influências da duração do horizonte de investimento e da taxa de comissão.(Freitas *et al*., 2009), (Reynolds *et al*., 2019) utilizada as redes neuronais e um algoritmo genético para a otimização operacional da oferta e da procura.

Métodos modernos de aprendizagem de máquina provaram oferecer desempenho superior e tornaram-se mais populares (Carvalho *et al*., 2019). Eles podem trabalhar com dados de elevada dimensionalidade e dados multivariados (Wuest *et al*., 2016). As ferramentas mais populares incluem Redes Neuronais Artificiais (ANN), que foram propostas em muitas aplicações industriais, incluindo *soft sensing* (Soares, 2015) e controle preditivo (Shin *et al*., 2018). Modelos de floresta aleatória (*Random Forest*) também são bons preditores, como mostrado neste estudo (Paolanti *et al*., 2018).

3.2. Redes Neuronais Artificiais

As Redes Neuronais Artificiais têm recebido atenção especial, na área de energia elétrica. Estudos de (Chao-Ton Su *et al*., 2002; J.-T. Zhang & Xiao, 2012) mostram a sua capacidade e desempenho como bons preditores, desde que um conjunto de dados com qualidade e quantidade de dados suficientes esteja disponível e os parâmetros corretos sejam encontrados.

As RNA tradicionais são simples e adequadas para uma ampla gama de problemas. Bangalore & Tjernberg, (2015) mostraram o desempenho de redes neuronais para detecção precoce de falhas em rolamentos de caixa de engrenagens, para otimizar a manutenção de turbinas eólicas. As Redes Neuronais Recorrentes (RNN) são relativamente populares para tarefas de manutenção preditiva. São um dos métodos mais eficientes de previsão. Apresentam um bom desempenho na previsão de falhas com base em séries temporais de dados (Koprinkova-Hristova *et al*., 2011) (Rivas *et al*., 2019).

Wang et al. (2020) utilizaram um RNN para manutenção preditiva e proactiva para equipamentos de energia ferroviária de alta velocidade. Também utilizaram uma abordagem semelhante para a manutenção preditiva baseada em tecnologias *Internet of Things* (IoT) utilizando um estimador RNN de *Long Shot-Term Memory* (LSTM). No entanto, para previsão em dados sequenciais, LSTM e *Gated Recurrent Units* (GRU) mostraram desempenho superior (Sugiyarto & Abadi, 2019) (Mateus *et al*., 2021; Mateus, Mendes, Farinha, & Cardoso, 2021).

Chui *et al*. (2021) também utilizaram um modelo RNN para prever a vida útil remanescente dos motores *turbofan*. De acordo com os autores, o erro médio quadrático (RMSE) melhorou 12,95-39,32 % em comparação com trabalhos existentes. As redes LSTM também foram utilizadas para prever falhas de motores de compressores de ar (Tsibulnikova *et al*., 2019), fornos de indução (Choi *et al*., 2020), equipamento de petróleo e gás (Tsibulnikova *et al*., 2019), e componentes de máquinas, tais como rolamentos (Wu *et al*., 2018).

4. Unidades de LSTM e GRU

A Figura *4* mostra o desenho interior de uma célula de unidade LSTM, de acordo com (Li & Lu, 2019). Formalmente, o modelo de célula LSTM é caracterizado da seguinte forma:

$$f_t = \sigma\left(x_t W_f + h_{t-1} U_f + b_f\right) \tag{1}$$

$$i_t = \sigma(x_t W_i + h_{t-1} U_i + b_i) \tag{2}$$

$$o_t = \sigma(x_t W_o + h_{t-1} U_o + b_o) \tag{3}$$

$$\tilde{C}_t = tan \tag{4}$$

$$C_t = \sigma(f_t \times C_{t-1} + i_t \times \tilde{C}_t) \tag{5}$$

$$h_t = tanh(C_t) \times o_t \tag{6}$$

As matrizes W_q e U_q contêm os pesos da entrada e das ligações recorrentes, onde o índice pode ser a porta de entrada i, porta de saída o, a porta de esquecimento f ou a célula de memória c, dependendo da activação a ser calculada. $C_t \in R^h$ não é apenas uma célula de um LSTM, mas contém células h das unidades LSTM, enquanto i_t, $o_t e f_t$ representam as activações da unidade, respectivamente, os portões de entrada, saída e esquecimento, no espaço de tempo t, onde:

- $x_t \in R^h$: vector de entrada para a unidade LSTM;
- $f_t \in (0{,}1)^h$ esquecer o vetor de ativação do portão;
- $i_t \in (0{,}1)^h$ vector de activação do portão de entrada/actualização;
- $o_t \in (0{,}1)^h$ vetor de ativação do portão de saída;
- $h_t \in (-1{,}1)^h$ vetor de estado oculto, também conhecido como o vetor de saída da unidade LSTM;
- $\tilde{c}_t \in (-1{,}1)^h$ vetor de ativação de entrada da célula;
- $c_t \in R^d$: vector de estado das células.

$W \in R^{h \times d}$×d, $U \in R^{h \times h}$ e $b \in R^h$são matrizes de peso e parâmetros vetoriais de polarização, que precisam de ser aprendidas durante a formação. Os índices d e h referem-se ao número de entradas características e número de unidades escondidas.

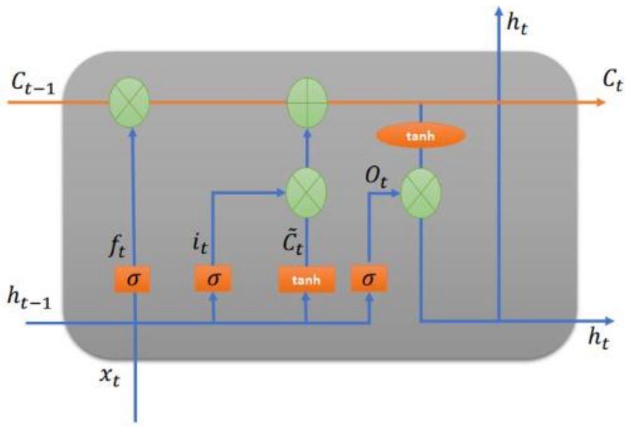


Figura 4- Estrutura detalhado de uma unidade de memória LSTM.

O GRU foi introduzido por (Chung et al., 2014). Embora inspirado na unidade LSTM, é considerado mais simples de calcular e implementar. Ele retém a imunidade do LSTM ao problema do *vanishing gradient*, um problema que dificulta o treino. A sua estrutura interna é mais simples e, portanto, também é mais fácil de treinar, pois menos cálculos são necessários para atualizar os estados internos. A porta de atualização controla até que ponto as informações de estado do momento anterior são retidas no estado atual, enquanto a porta de redefinição determina se o estado atual deve ser combinado com as informações anteriores (Cho *et al*., 2014). A Figura *5* mostra a arquitetura interna de uma célula unitária GRU.

Para avaliar o desempenho de predição do modelo, utilizou-se *Root Mean Squared Error* (RMSE), Erro Médio Percentual Absoluto (MAPE) e Erro Médio Absoluto (MAE), que são definidos da seguinte forma:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}\right)^2} \tag{7}$$

onde Y_t representa o valor desejado (real) e $\hat{Y}$é o valor predito (obtido do modelo). A diferença entre Y e $\hat{Y}$ é o erro entre o valor que se espera obter e o valor realmente obtido da rede. n representa o número de amostras usadas no conjunto de teste.

$$MAE = \frac{1}{n}\sum_{t=1}^{n} Y_t - \hat{Y} \vee \tag{8}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|Y_t - \hat{Y}|}{Y_t} \tag{9}$$

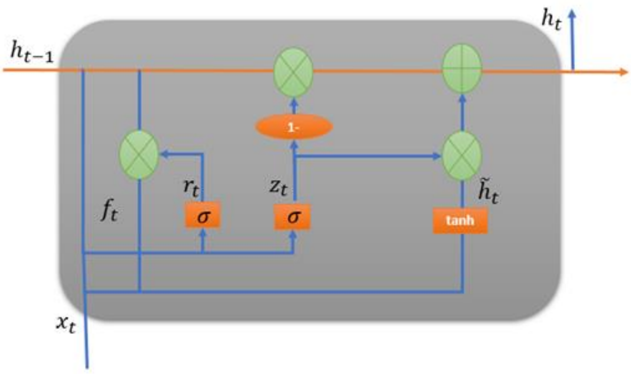


Figura 5- Estrutura detalhado de uma unidade de memória GRU.

$$z_t = \sigma(x_t W^z + h_{t-1}U^z + b_z) \tag{10}$$

$$r_t = \sigma(x_t W^r + h_{t-1}U^r + b_r) \tag{11}$$

$$\tilde{h}_t = tan(r_t \times h_{t-1}U + x_t W + b) \tag{12}$$

$$h_t = (1 - z_t) \times \tilde{h}_t + z_t \times h_{t-1} \tag{13}$$

Onde Wz, Wr, W denotam as matrizes de peso para o correspondente vector de entrada ligado. Uz, Ur, U representam as matrizes de peso do passo de tempo anterior, e b_r, b_z e b são o enviesamento. O σ denota a função sigmoide logística, r_t denota a porta de reinicialização, z_t denota a porta de atualização, e $\hbar_t$ denota a camada oculta candidata (Lynn *et al*., 2019).

5. Desenvolvimento

O presente projeto debruça-se sobre a otimização do funcionamento de 5 prensas indústrias monitoradas por seis sensores (por prensa), com período de amostragem a cada 5 minuto. O conjunto de dados contém amostras de dados desde 1 de fevereiro de 2018 a 20 de novembro de 2021, para um total de 1388 dias.

As variáveis monitoradas para todas as prensas são 1) Intensidade de corrente elétrica (C. intensity); 2) Nível de óleo da unidade hidráulica (Hydraulic unit level); 3) pressão do VAT; 4) Velocidade do motor (Velocity); 5) Temperatura na unidade hidráulica (Temperature at U.H.); e 6) Binário (Torque). Na Figura 6 está presente todas as variáveis em formato de uma série temporal.

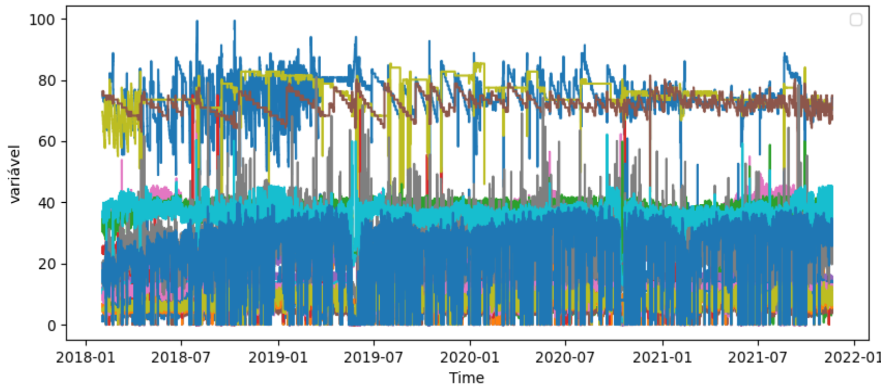


Figura 6- Série temporal de todas as variáveis das prensas de polpa de papel.

Na Figura 7 verifica-se que existe uma correlação elevada sobre os valores de pasta de papel produzida, em relação à corrente e ao torque. Estes valores da correlação chegam a ser de 0.9, sendo que o valor máximo da correlação 1. Verificando estas correlações entre o valor da produção e as variáveis Intensidade da Corrente, Torque e a Velocidade, pode-se validar a exequibilidade das previsões destas variáveis em conjunto.

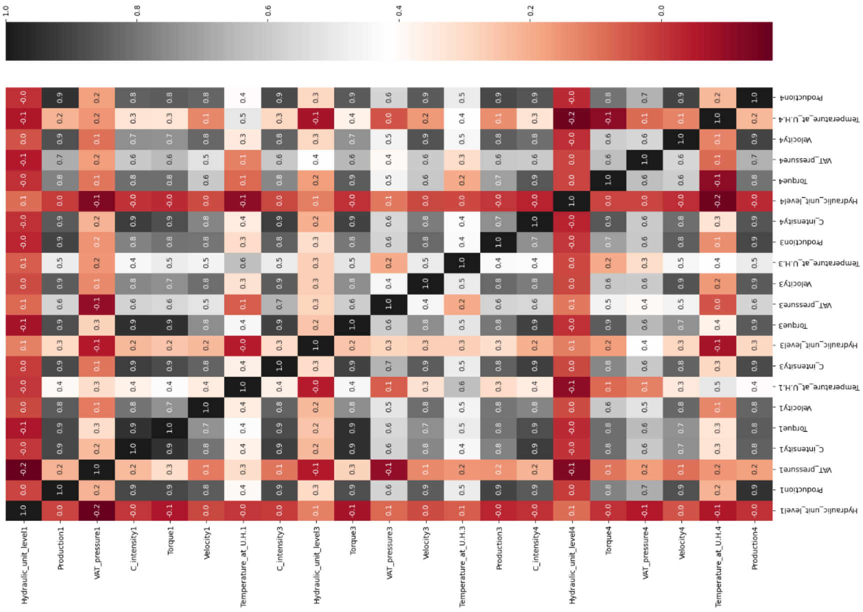


Figura 7- Correlação entre todas as variáveis das prensas de polpa de papel.

Além da análise das correlações entre variáveis, foram também analisadas as auto-correlações, de forma a compreender-se até quantas amostras se pode esperar fazer uma boa previsão. Verifica-se que individualmente cada variável apresenta autocorrelações que decaem relativamente depressa. De qualquer forma, os modelos de previsão multivariados foram testados para previsão a 30 dias, esperando-se dessa forma obter o máximo ganho da informação presente no conjunto de todas as variáveis. Para a realização dos testes foram apenas consideradas as 2 prensas de papel.

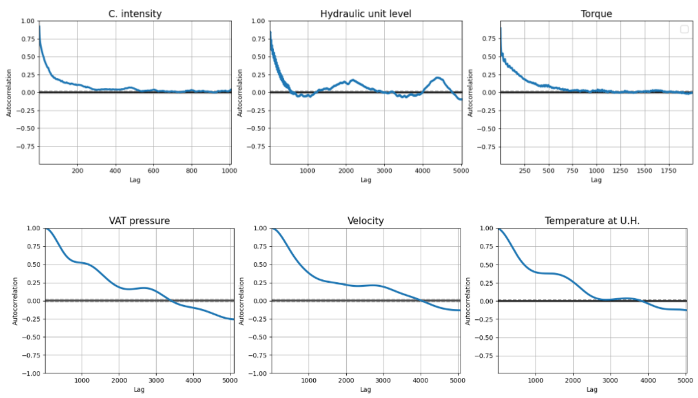


Figura 8- Autocorrelações variáveis da prensa 4.

A Figura 8 mostra as autocorrelações das variáveis da prensa 4 e o nível de óleo da unidade hidráulica é o que tem um decaimento mais rápido da autocorrelação. As outras variáveis mostram uma boa melhoria, indicando maiores probabilidades de pequenos erros de previsão.

Conforme apresentado atrás, redes neuronais recorrentes apresentam uma boa capacidade de previsão de séries temporais, embora ainda não havia previsões feitas para dados deste tipo de prensa. Nisto foram explorados os parâmetros e hiperparâmetros do modelo em questão, o que resultou em dois trabalhos científicos (Mateus, Mendes, Farinha, & Cardoso, 2021) (Mateus, Mendes, Farinha, Assis, et al., 2021).

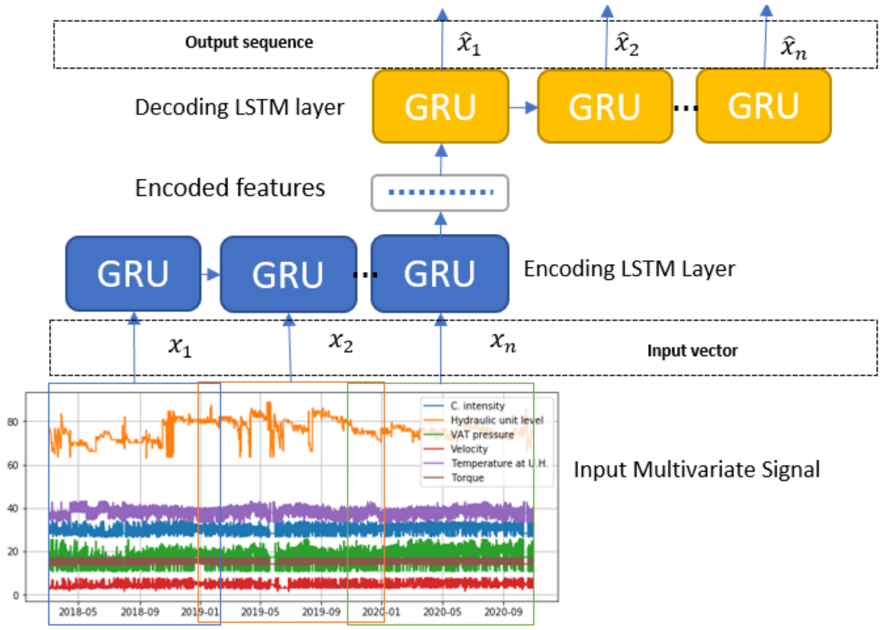


Figura 9 – Arquitetura do modelo de previsão 30 dias para frente.

As experiências foram realizadas utilizando um computador com um processador i5 de terceira geração, com 8 GB de RAM. A Figura 9 descreve a arquitetura de um dos modelos de rede utilizados. Os modelos foram implementados em linguagem de programação *Python* utilizando a biblioteca TensorFlow e Keras. A mesma está composta por unidades GRU.

Os resultados mostram que o modelo proposto é capaz de aprender e prever o comportamento das seis variáveis estudadas: Intensidade de corrente elétrica; Nível de óleo da unidade hidráulica; pressão do VAT; Velocidade do motor; Temperatura na unidade hidráulica e binário, como mostra a Figura 10.

Figura 10 - Gráfico das previsões da prensa 2 com diferentes combinações de funções de ativação.

Usando uma abordagem comparativa das redes neuronais GRU, é possível verificar que a GRU oferece bons resultados. Os erros de previsão são menores do que aqueles apresentados pela rede neural LSTM no estudo de Mateus et al. (2021) e o GRU é mais imune a problemas de explosão ou desaparecimento do gradiente durante a aprendizagem, como referido atrás. Portanto ele aprende numa gama mais ampla de configurações.

A rede com unidades GRU suporta taxas de reamostragem mais altas, portanto, pode trabalhar com janelas de dados menores. O otimizador Adam é eficaz para minimizar o problema de *exploding gradient*, e foi também utilizado nos modelos GRU. Um modelo GRU otimizado oferece melhores resultados com uma janela deslizante de amostragem de 12 dias de dados com um período de amostragem de 1 hora. Isto para um modelo GRU com 50 unidades na camada oculta. As melhores funções de ativação dependem do modelo, porém, o *relu-tanh* talvez seja um dos melhores modelos, em média.

Foi feita também uma análise do impacto de remover dados discrepantes do conjunto de dados. A limpeza de dados discrepantes elimina muitos valores extremos, designadamente valores próximos de zero que são lidos durante as paragens da máquina, ou registados por falhas de comunicação. Esses dados foram eliminados usando o intervalo interquartil e método LOESS (que é regressão local ou regressão polinomial local).

No entanto, mesmo depois da eliminação de dados discrepantes, a amplitude e a frequência das variações ainda tornam as leituras muito instáveis. Os dados têm muito ruído e tornam a aprendizagem e a leitura dos resultados difíceis. Utilizando método de eliminação dos dados discrepantes e o método de filtragem e alisamento LOESS (ou LOWESS) foi possível aumentar a precisão do modelo e produzir gráficos mais estáveis e, portanto, fáceis de ler. A Figura 11 e a Figura 12 mostra uma das representações da previsão da variável da Intensidade da Corrente num período de 30 dias para frente. Com esta precisão é possível prever com antecedência variáveis que poderão originar uma paragem.

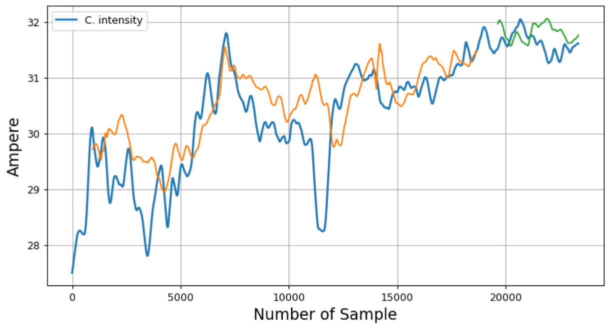


Figura 11 - Gráfico das previsões da prensa 2 da Intensidade da corrente com previsão a 30 dias. A azul o sinal original alisado com LOESS, a laranja a previsão no conjunto de treino e a verde a previsão no conjunto de teste.

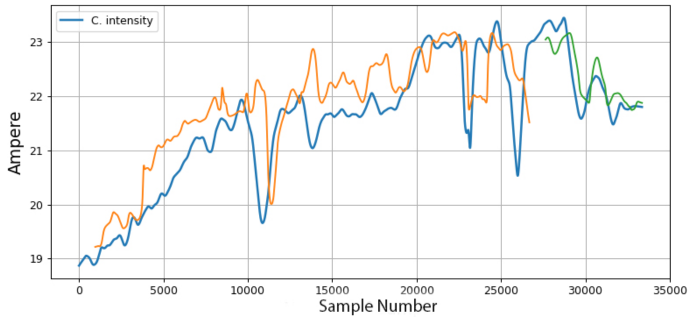


Figura 12 - Gráfico das previsões da prensa 4 da intensidade da corrente com previsão a 30 dias. A azul o sinal original alisado com LOESS, a laranja a previsão no conjunto de treino e a verde a previsão no conjunto de teste.

Aplicando os métodos de processamento, os resultados com GRU e LOESS apresentaram uma superioridade com relação aos resultados publicados no trabalho (Mateus, Mendes, Farinha, Assis, et al., 2021). O MAPE (Erro Percentual Absoluto Médio) para a Intensidade Atual para prensa de papel 2 diminuiu de 2,30% para 0,62%. Para o nível de óleo hidráulico o MAPE diminuiu de 2,8% para 1,85%. Para o Torque, a MAPE diminuiu de 2,85% para 2,24%. Para a pressão do VAT, a MAPE passou de 9,87% para 3,91%. Para a velocidade, a MAPE baixou de 11,8% para 10,27%. Finalmente, para a Temperatura, a MAPE baixou de 2,66% para 0,96%.

Com as previsões de antecipação de 30 dias das variáveis da prensa, aplicamos o mesmo modelo com objetivo de realizar também uma previsão da nossa produção total. Na Tabela 1 é possível verificar que os erros apresentados são significativamente baixos. O coeficiente de determinação é alto, o que nos leva a ter um modelo preparado para assim ser aplicado na previsão da produção. Na Figura 13 está presente o resultado da previsão da produção total das prensas de pasta de papel, o mesmo encontra-se com uma taxa de amostragem de uma amostra por mês.

Tabela 1 - Resumo dos melhores erros de predição da variável da produção total.

Error	Produção Total
MAPE	12.23
R^2	0.85
MAE	0.08
RMSE	0.08

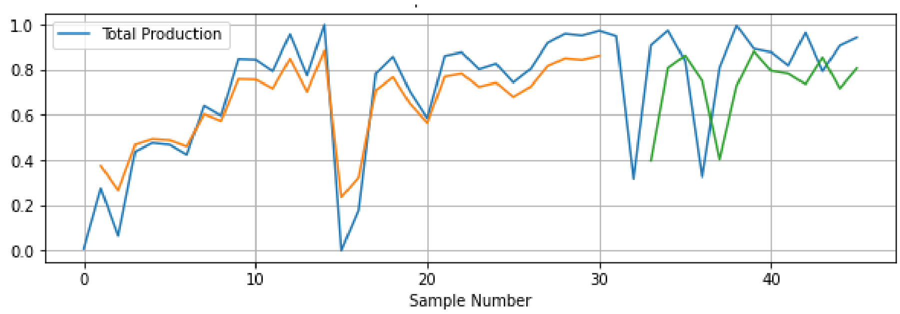


Figura 13 - Gráfico das previsões da produção total a 30 dias.

Os resultados mostram que é possível prever com bom grau de certeza o comportamento futuro das prensas de pasta de papel industrial e a sua capacidade de produção com até 30 dias de antecedência. Isso pode ser uma boa oportunidade para otimizar as decisões de manutenção, reduzir o tempo de paragem e os custos.

6. Aplicabilidade prática e perspetivas de futuro

Pelos registos dos técnicos de manutenção houve uma avaria no mês de setembro. Aa mesma foi prevista pelo nosso modelo por via do comportamento da variável da corrente mostrada na Figura 14, em que o momento está marcado com uma cruz. Não foi possível obter informação sobre outros momentos de avaria até à data de conclusão do presente trabalho.

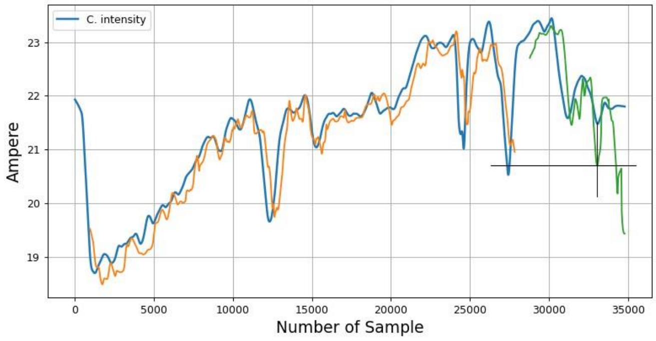


Figura 14 – Previsão da prensa 4 da intensidade da corrente á 30 dias para frente.

Fazendo uma previsão a 30 dias a Figura 15, apresenta a previsão de uma forma interativa do mês de dezembro do ano de 2021, o que torna mais fácil a visualização da previsão e por sua vez nas tomadas de decisão com relação a produção do mês em questão.

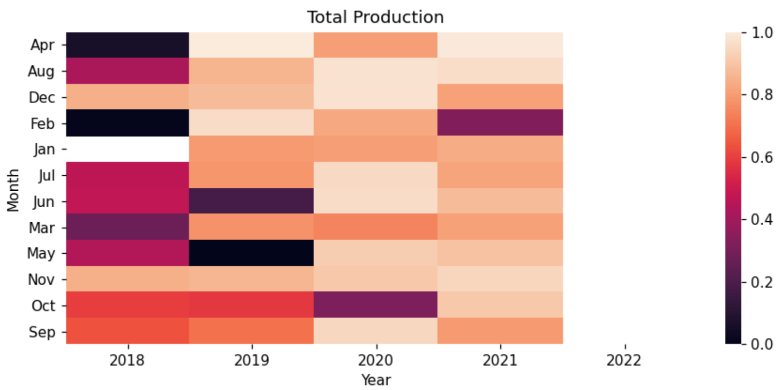


Figura 15 – Placar da previsão da produção total com 30 dias para frente.

Tendo a disponibilidade da produção das três prensas de pasta de papel, no trabalho futuro outras variáveis podem ser incluídas nos algoritmos de previsão, nomeadamente as cotações bolsistas que se sabe afetaram a procura e, portanto, as necessidades de produção. Segundo a Figura 16, variáveis de cotação da bolsa, obtidas no site da EURONEXT e FRED , que apresentam uma grande relevância a nível de correlação com a nossa produção total (Soma das produções), estas variáveis da bolsa são: 1) ALTRI

SGPS, 2) Consumer_Prices e 3) Pulp_Paper. Com estas conclusões deu-se início ao estudo de um modelo de previsão para assim ser feita a previsão das variáveis mais relevantes.

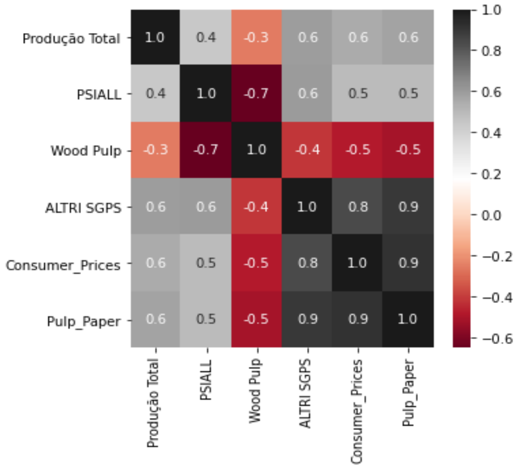


Figura 16- Correlação entre a produção total e os valores da bolsa.

7. Conclusão

A manutenção preditiva assume uma importância cada vez maior para as empresas, e o crescente poder computacional para adquirir e processar dados tornam possível a sua aplicação prática. No âmbito do presente trabalho foram desenvolvidos algoritmos pioneiros para previsão do estado de prensas de pasta de papel. O método e algoritmos são replicáveis entre prensas, tendo sido testados com duas prensas diferentes. Cada prensa tem um estado de funcionamento diferente, sendo necessário treino para ajuste dos modelos.

A ferramenta desenvolvida é um auxílio na tomada de decisão, com base nos comportamentos das variáveis, sendo que com uma previsão de 30 dias para frente a empresa pode ajustar as intervenções nas máquinas. Esta torna possível a redução das falhas das prensas de papel, por conseguinte evita possíveis paragens não programadas e os custos que as mesmas provocam.

No âmbito deste projeto também se demonstrou uma boa capacidade de previsão da produção total da prensa, com modelo GRU. Isto trará grandes benefícios na empresa, para otimização e eventual redução nos ‘stocks’ sem que haja falhas nem atrasos nas entregas.

O projeto apresenta uma desvantagem que é, para elevadas quantidades de dados é necessário uma elevada capacidade computacional, sendo que o tempo de treino do modelo pode levar horas ou até mesmo dias, no caso de várias prensas serem consideradas. De qualquer forma, atualmente o processamento em GPU é relativamente barato e com baixo consumo energético.

Referências bibliográficas

Aydin, O., & Guldamlasioglu, S. (2017). Using LSTM networks to predict engine condition on large scale data processing framework. *2017 4th International Conference on Electrical and Electronics Engineering, ICEEE 2017*, 281–285. https://doi.org/10.1109/ICEEE2.2017.7935834

Bangalore, P., & Tjernberg, L. B. (2015). An artificial neural network approach for early fault detection of gearbox bearings. *IEEE Transactions on Smart Grid*, *6*(2), 980–987. https://doi.org/10.1109/TSG.2014.2386305

Baptista, M., Sankararaman, S., de Medeiros, I. P., Nascimento, C., Prendinger, H., & Henriques, E. M. P. (2018). Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. *Computers & Industrial Engineering*, *115*, 41–53. https://doi.org/10.1016/J.CIE.2017.10.033

Bruneo, D., & de Vita, F. (2019). On the use of LSTM networks for predictive maintenance in smart industries. *Proceedings - 2019 IEEE International Conference on Smart Computing, SMARTCOMP 2019*, 241–248. https://doi.org/10.1109/SMARTCOMP.2019.00059

Carvalho, D. M., & Nascimento, M. C. V. (2022). Hybrid matheuristics to solve the integrated lot sizing and scheduling problem on parallel machines with sequence-dependent and non-triangular setup. *European Journal of Operational Research*, *296*(1), 158–173. https://doi.org/10.1016/j.ejor.2021.03.050

Choi, Y., Kwun, H., Kim, D., Lee, E., & Bae, H. (2020). Method of Predictive Maintenance for Induction Furnace Based on Neural Network. *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 609–612. https://doi.org/10.1109 / BigComp48618.2020.00021

Chui, K. T., Gupta, B. B., & Vasant, P. (2021). A Genetic Algorithm Optimized RNN-LSTM Model for Remaining Useful Life Prediction of Turbofan Engine. *Electronics 2021, Vol. 10, Page 285*, *10*(3), 285. https://doi.org/10.3390/ELECTRONICS10030285

Du, S., Li, T., Yang, Y., & Horng, S. J. (2020). Multivariate time series forecasting via attention-based encoder–decoder framework. *Neurocomputing*, *388*, 269–279. https://doi.org/10.1016/J.NEUCOM.2019.12.118

Freitas, F. D., de Souza, A. F., & de Almeida, A. R. (2009). Prediction-based portfolio optimization model using neural networks. *Neurocomputing*, *72*(10), 2155–2170. https://doi.org/https://doi.org/10.1016/j.neucom.2008.08.019

Gejo-García, J., Reschke, J., Gallego-García, S., & García-García, M. (2022). Development of a System Dynamics Simulation for Assessing Manufacturing Systems Based on the Digital Twin Concept. *Applied Sciences*, *12*(4). https://doi.org/10.3390/app12042095

Ghaboussi, J., & Joghataie, A. (1995). Active Control of Structures Using Neural Networks. *Journal of Engineering Mechanics*, *121*(4), 555–567. https://doi.org/10.1061/(ASCE)0733-9399(1995)121:4(555)

Gorski, E. G., Loures, E. de F. R., Santos, E. A. P., Kondo, R. E., & Martins, G. R. D. N. (2022). Towards a smart workflow in CMMS/EAM systems: An approach based on ML and MCDM. *Journal of Industrial Information Integration*, *26,* 100278. https://doi.org/10.1016/J.JII.2021.100278

Jabeur, S. ben, Gharib, C., Mefteh-Wali, S., & Arfi, W. ben. (2021). CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technological Forecasting and Social Change*, *166*, 120658. https://doi.org/10.1016/J.TECHFORE.2021.120658

Koprinkova-Hristova, P. D., Hadjiski, M. B., Doukovska, L. A., & Beloreshki, S. v. (2011). Recurrent Neural Networks for Predictive Maintenance of Mill Fan Systems. *International Journal of Electronics and Telecommunications*, *57*(Vol. 57, No. 3), 401–406. https://doi.org/10.2478/v10177-011-0055-2

Li, Y., & Lu, Y. (2019). LSTM-BA: DDoS Detection approach combining LSTM and bayes. *Proceedings - 2019 7th International Conference on Advanced Cloud and Big Data, CBD 2019*, 180–185. https://doi.org/10.1109/CBD.2019.00041

Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, *108*, 33–47. https://doi.org/10.1016/J.YMSSP.2018.02.016

Lynn, H. M., Pan, S. B., & Kim, P. (2019). A Deep Bidirectional GRU Network Model for Biometric Electrocardiogram Classification Based on Recurrent Neural Networks. *IEEE Access*, *7*, 145395–145405. https://doi.org/10.1109/ACCESS.2019.2939947

Mateus, B. C., Mendes, M., Farinha, J. T., & Cardoso, A. M. (2021a). Anticipating Future Behavior of an Industrial Press Using LSTM Networks. *Applied Sciences*, *11*(13), 6101. https://doi.org/10.3390/app11136101

Mateus, B. C., Mendes, M., Farinha, J. T., & Cardoso, A. M. (2021b). Anticipating Future Behavior of an Industrial Press Using LSTM Networks. *Applied Sciences 2021, Vol. 11, Page 6101*, *11*(13), 6101. https://doi.org/10.3390/APP11136101

Paolanti, M., Romeo, L., Felicetti, A., Mancini, A., Frontoni, E., & Loncarski, J. (2018). Machine Learning approach for Predictive Maintenance in Industry 4.0. *2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications, MESA 2018,* 1–6. https://doi.org/10.1109/MESA.2018.8449150

Reynolds, J., Ahmad, M. W., Rezgui, Y., & Hippolyte, J.-L. (2019). Operational supply and demand optimisation of a multi-vector district energy system using artificial neural networks and a genetic algorithm. *Applied Energy*, *235*, 699–713. https://doi.org/https://doi.org/10.1016/j.apenergy.2018.11.001

Rivas, A., Fraile, J. M., Chamoso, P., González-Briones, A., Sittón, I., & Corchado, J. M. (2019). A Predictive Maintenance Model Using Recurrent Neural Networks. *Advances in Intelligent Systems and Computing*, *950*, 261–270. https://doi.org/10.1007/978-3-030-20055-8_25

Sana, S. S. (2012). Preventive maintenance and optimal buffer inventory for products sold with warranty in an imperfect production system. *Http://Dx.Doi.Org/10.1080/00207543.2011.623838*, *50*(23), 6763–6774. https://doi.org/10.1080/00207543.2011.623838

Shin, J. H., Jun, H. B., & Kim, J. G. (2018). Dynamic control of intelligent parking guidance using neural network predictive control. *Computers & Industrial Engineering*, *120*, 15–30. https://doi.org/10.1016/J.CIE.2018.04.023

Soares, S. G. (2015). *Ensemble Learning Methodologies for Soft Sensor Development in Industrial Processes | Estudo Geral*. Repositório Científico Da UC. https://estudogeral.sib.uc.pt/handle/10316/28313

Stewart, G. E., Gorinevsky, D. M., & Dumont, G. A. (2003). Feedback Controller Design for a Spatially Distributed System: The Paper Machine Problem. *IEEE Transactions on Control Systems Technology*, *11*(5), 612–628. https://doi.org/10.1109/TCST.2003.816420

Su, C. T., Yang, T., & Ke, C. M. (2002). A neural-network approach for semiconductor wafer post-sawing inspection. *IEEE Transactions on Semiconductor Manufacturing*, *15*(2), 260–266. https://doi.org/10.1109/66.999602

Sugiyarto, A. W., & Abadi, A. M. (2019). Prediction of Indonesian palm oil production using long short-term memory recurrent neural network (LSTM-RNN). *Proceedings - 2019 1st International Conference on Artificial Intelligence and Data Sciences, AiDAS 2019*, 53–57. https://doi.org/10.1109/AIDAS47888.2019.8970735

Tsibulnikova, M. R., Pham, V. A., Yu Aikina -, T., Xue-feng, L., Xiao-ben, L., Jan-ding, H., -, al, Abbasi, T., Hann Lim, K., & San Yam, K. (2019). Predictive Maintenance of Oil and Gas Equipment using Recurrent Neural Network. *IOP Conference Series: Materials Science and Engineering*, *495*(1), 012067. https://doi.org/10.1088/1757-899X/495/1/012067

Wang, Q., Bu, S., & He, Z. (2020). Achieving Predictive and Proactive Maintenance for High-Speed Railway Power Equipment with LSTM-RNN. *IEEE Transactions on Industrial Informatics*, *16*(10), 6509–6517. https://doi.org/10.1109/TII.2020.2966033

Wu, Q., Ding, K., & Huang, B. (2018). Approach for fault prognosis using recurrent neural network. *Journal of Intelligent Manufacturing 2018 31:7*, *31*(7), 1621–1633. https://doi.org/10.1007/S10845-018-1428-5

Wuest, T., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine learning in manufacturing: advantages, challenges, and applications.

Http://Mc.Manuscriptcentral.Com/Tpmr, *4*(1), 23–45. https://doi.org/10.1080/21693277.2016.1192517

Yam, R. C. M., Tse, P. W., Li, L., & Tu, P. (2001). Intelligent Predictive Decision Support System for Condition-Based Maintenance. *The International Journal of Advanced Manufacturing Technology 2001 17:5*, *17*(5), 383–391. https://doi.org/10.1007/S001700170173

Zhang, J. T., & Xiao, S. (2012). A note on the modified two-way MANOVA tests. *Statistics & Probability Letters*, *82*(3), 519–527. https://doi.org/10.1016/J.SPL.2011.12.005

Züfle, M., Moog, F., Lesch, V., Krupitzer, C., & Kounev, S. (2021). A machine learning-based workflow for automatic detection of anomalies in machine tools. *ISA Transactions*. https://doi.org/10.1016/J.ISATRA.2021.07.010

---

i