

Article



Comparison of Different Features and Neural Networks for Predicting Industrial Paper Press Condition

João Antunes Rodrigues ^{1,2,*}, José Torres Farinha ^{3,4}, Mateus Mendes ^{3,5,*}, Ricardo J. G. Mateus ² and António J. Marques Cardoso ¹

- ¹ CISE—Electromechatronic Systems Research Centre, University of Beira Interior, Calçada Fonte do Lameiro, 6200-358 Covilhã, Portugal
- ² EIGeS—Research Centre in Industrial Engineering, Management and Sustainability, Universidade Lusófona, Campo Grande 376, 1749-024 Lisboa, Portugal
- ³ Polytechnic of Coimbra—ISEC, Quinta da Nora, 3030-199 Coimbra, Portugal
- ⁴ Department of Mechanical Engineering, Centre for Mechanical Engineering, Materials and Processes, University of Coimbra, 3030-290 Coimbra, Portugal
- ⁵ Department of Electrical and Computer Engineering, Institute of Systems and Robotics, University of Coimbra, 3030-194 Coimbra, Portugal
- * Correspondence: p5942@ulusofona.pt (J.A.R.); mmendes@isec.pt (M.M.)

Abstract: Forecasting has extreme importance in industry due to the numerous competitive advantages that it provides, allowing to foresee what might happen and adjust management decisions accordingly. Industries increasingly use sensors, which allow for large-scale data collection. Big datasets enable training, testing and application of complex predictive algorithms based on machine learning models. The present paper focuses on predicting values from sensors installed on a pulp paper press, using data collected over three years. The variables analyzed are electric current, pressure, temperature, torque, oil level and velocity. The results of XGBoost and artificial neural networks, with different feature vectors, are compared. They show that it is possible to predict sensor data in the long term and thus predict the asset's behaviour several days in advance.

Keywords: maintenance; neural networks; XGBoost; forecast; sensor prediction

1. Introduction

Advanced sensing technology, combined with high performance computing, help industries run with increasing reliability and competitiveness.

Industries are striving to constantly improve industrial processes and equipment. Maintenance plays a fundamental role in this field, being very important to prevent disruptions in production chains.

1.1. The Importance of Maintenance

Maintenance is a combination of technical and administrative activities required to maintain equipment, facilities, and other physical assets. The goal is to maintain those assets in the desired operational condition, or restore them so that they can fulfil their function with quality [1–3]. The main objectives of a good maintenance policy are: safety, quality, cost reduction, and availability [4]. The optimization of those four objectives at the same time is challenging, since they often conflict with each other. In those cases, it is the maintenance management's responsibility to find the best compromise solution based on the company's strategic objectives.

Predictive maintenance is one of the fastest growing types of maintenance in the industry nowadays [5]. It aims to predict the occurrence of failures before they happen, using data from sensors and state-of-the-art augmented intelligence algorithms. The algorithms are trained based on historical data, the operating condition of the assets is monitored, and the trends are predicted in near real time.



Citation: Rodrigues, J.A.; Farinha, J.T.; Mendes, M.; Mateus, R.J.G.; Cardoso, A.J.M. Comparison of Different Features and Neural Networks for Predicting Industrial Paper Press Condition. *Energies* 2022, 15, 6308. https://doi.org/ 10.3390/en15176308

Academic Editor: Valentina Colla

Received: 26 July 2022 Accepted: 25 August 2022 Published: 29 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Industrial systems currently use tens, hundreds, or thousands of sensors to collect data to be used primarily to monitor processes and equipment condition [6,7].

Due to developments in data processing, along with storage algorithms and hardware, it is currently possible to store and process large quantities of data to predict the future behaviour of equipment, thus making it possible to forecast failures in advance [8].

The asset's behaviour, after being observed and analyzed, can be predicted with stateof-the-art algorithms. Such techniques have a positive impact on production reliability, security, availability and quality [9]. It should also be noted that predictive maintenance promotes environmental sustainability, as it contributes to reduce industrial downtimes, unnecessary maintenance interventions, production surpluses, and non-conforming products [10].

1.2. Industry 4.0 in Maintenance

Industry 4.0 is a consequence of scientific and technological advances, including predictive maintenance.

The amount of data extracted from industrial processes has exponentially increased due to the rise of non-invasive sensing technologies and decreasing hardware costs. However, it is essential to calibrate the sensors correctly, so that the acquired data are reliable [7,11]. Poor or incorrect data do not add value and can lead to prediction errors [12,13].

Analysis of reliable data with predictive computational techniques can avoid unnecessary equipment changes, save costs and improve safety, availability, and efficiency of processes [14].

1.3. Predictive Maintenance from an Economic Point of View

Maintenance was often seen as a source of unnecessary cost by industry, so it was often overlooked by companies. Nowadays, the role of maintenance is better understood. It is considered a key factor for the success of companies, helping them to reduce production costs and, consequently, increase profits [15].

Although applying predictive maintenance policies may involve significant costs, those costs are often less than the benefits generated from a well-planned system [16].

Most devices involve an expensive hardware network, formed by many sensors for data collection and storage. In addition to hardware, predictive maintenance requires additional costs for training staff, as well as analysing data and developing and training prediction and classification methods.

By enabling more efficient, sustainable, and higher quality production, the application of predictive maintenance also affects the company's image in the market and contributes to increase its value.

Predictive maintenance can be applied to almost all industrial equipment. However, due to its high implementation costs, technical and economic analyses must be performed before proceeding to modelling and deployment, namely determining the criticality of the equipment in case of failure or anomaly, and the potential economic losses for the company.

According to François Monchy, the more expensive the unavailability of an equipment, the more important its maintenance [17]. In other words, direct and indirect costs of equipment unavailability along with the value generated by the equipment are the most important factors to consider when choosing a maintenance policy.

The greatest advantage of predictive maintenance is that it can assess the current condition of any machine and predict when it needs maintenance before a fault happens. With a properly implemented and updated maintenance policy, it is possible to schedule equipment maintenance for times that will have the least impact in production schedule and deadlines, minimizing disruptions in production lines and improving the quality of the items produced by the factory, contributing to the profitability and sustainability of any company's business.

1.4. Artificial Neural Networks

Artificial neural networks are machine learning models with interconnected nodes distributed over several layers. The networks can be trained to recognize hidden patterns, to classify input samples into a few classes and to perform predictions. This type of model was inspired by the human brain [18,19].

The neuron is the atomic unit of a neural network. When an input vector is given, the neuron provides an output which is a function of the weighted average of the input vector's coordinates. The neurons' outputs can then be fed as inputs to other neurons in the subsequent layers.

Optimization of neural networks is a challenging problem, and it has been the topic of many works [20,21].

Feed-forward (FF) neural networks are a type of neural network in which the data flow in a single direction, from input to output, without any feedback. On the contrary, outputs in recurrent neural networks (RNN) can be fed back into the network, allowing the network to remember past events and operate in non-episodic environments.

Multi-layer perceptron (MLP) is a type of FF neural network. It comprises three types of layers: one input layer, several hidden layers, and one output layer. The main applications of MLP networks are pattern classification, recognition, and prediction [22].

As the computing power and big data increase, deep learning models are becoming more popular in several fields of science. Deep models are characterized by containing several layers, while shallow models rarely have more than three layers. For instance, deep networks are the preferred architecture in object detection or classification problems. Shallow neural networks are more adequate for prediction problems. Despite many clear distinctions between deep and shallow neural networks, some techniques developed for deep learning can help improving shallow models, and vice versa [23].

The importance of the present work is reinforced by several authors that have emphasized the necessity to change the focus from short-term (15 days) maintenance policies to long-term ones (90 days). The importance of these contributions corresponds to the increase of equipment's availability, which permits increased productivity and, at last, the success of the company [24–26].

1.5. XGboost and Random Forest

XGBoost is a scalable and highly accurate implementation of gradient boosting that pushes the limits of computing power for boosted tree algorithms, being built largely for energizing machine learning model performance and computational speed [27].

Random forest is also a popular and effective ensemble machine learning algorithm. It is widely used for classification and regression predictive modeling problems [28].

1.6. Objectives

The present research aims to propose a model to forecast sensor values of an industrial pulp paper press for 15 days, 30 days, and 90 days.

The goal was to compare the performance of multiple prediction models, including neural networks and other machine learning methods, optimizing different features and architectures.

The team defined that forecasts of most variables must have MAPE errors of less than 10%.

1.7. Contributions

Predicting in advance the values of the sensors allows us to anticipate the future state of the monitored equipment and to predict its expected operating conditions. The main contributions are:

• The approach proposed for the predictions in the present research compares and determines the best features, time windows and architectures for feed-forward shallow networks and XGBoost.

Results are compared to LSTM and GRU.

To the best of the authors' knowledge, these are novel contributions for the area of equipment maintenance, allowing us to maximize the useful life of equipment while still minimizing risk of failure.

Similar works on industrial sensor prediction use other machine learning models, including deep networks, which require larger computational networks, which demonstrates the contribution of this study using shallow networks.

1.8. Paper Structure

The structure of the paper is as follows. Section 2 addresses the work related to this field of research. Section 3 presents the data and explains how they were treated and filtered. Section 4 shows the architecture of the underlying neural network. Section 5 presents the metrics for evaluating the neural model. Section 6 displays the tests and results of this study. Sections 7 and 8 present time series for overlapping and non-overlapping sliding windows, respectively. Section 9 shows a comparison of different feature vectors and forecast models. Finally, Section 10 presents the conclusions.

2. Related Work

2.1. Neural Networks for Prediction and Classification

This section reviews relevant works using neural networks for prediction and classification, namely in the field of predictive maintenance.

Rodrigues et al. used a neural network to predict and classify the degradation state of diesel engine oils from laboratory analysis data on 21 oil parameters, achieving an accuracy over 90% [29].

Effective maintenance is essential to keep assets at maximum availability and accident free. For these reasons, Bukhsh et al. developed a model to predict the need for railway maintenance [30].

Elhag and Wang presented an application of artificial neural networks to assess bridge risk by computing their risk scores and categories [31].

Balluff and his team developed a model to predict wind speed and pressure through recurrent neural networks [32].

Deepika and Prakash predicted the power consumption of a virtual machine with the help of backwards predictive analytics using a multi-layer perceptron, achieving a 91% accuracy [33].

Hongxiang et al. developed an algorithm using artificial neural networks (ANNs) to analyze spectroscopy data from lubricant oils. Results proved that ANNs can be used to classify distinct types of lubricants and to distinguish routine conditions of a diesel engine from operating conditions [34].

An algorithm based on a multi-layer feed-forward neural network model was developed to control a steel pickling process in several simulation cases [35].

Okoh et al. presented an approach to determine when a system needs to undergo maintenance, repair, and overhaul, before a failure occurs. One of the main innovations of this project is that forecasts were made in the long-term [36].

One of the main challenges of maintenance is to increase the availability of equipment and, hence, it is important to prognose failures before they happen. Makridis et al. presented a machine learning approach for detecting anomalies from data collected through sensors installed on vessels, predicting the condition of specific parts of the vessels' main engine [37].

In 2021, Zhagparov et al. proposed a solution to automate the prediction of grain yield based on machine learning using the XGBRegressor algorithm on the territory of the Republic of Kazakhstan. Comparisons were made with linear regression and decision tree regressor algorithms [38].

Dong et al., in 2020, developed a prediction model based on the XGBoost algorithm that considers all potential influential factors simultaneously; the objective of this model was to predict the electrical resistivity based on an experimental database [27].

In summary, according to the authors referred, among others, neural networks have high prediction accuracy and can improve support in decision making [39,40].

2.2. Condition Monitoring in Paper Press

Condition monitoring plays a central role in the maintenance of paper machines; the main objective is to maximize the availability and reduce the costs of these manufacturing units and to prevent unexpected damage or mechanical breakdowns.

The results of the tests by Suomela et al. in 2002 make it clear that thermal imaging combined with adaptive drive has great potential for monitoring paper machine components [41].

The work by Bissessur et al. features the ability to detect faults and provide early warning of impending problems based on collected vibration data and pre-processing spectra. These data processed by a neural network provide an instant decision about the state of the felt that is monitored. This method can be extended to diagnose faults in a wide range of mechanical and rotating equipment in industries [42].

Mateus et al. developed predictive models based on long-term deep memory neural networks applied to a dataset of sensor readings. The results show that it is possible to predict future behaviour up to a month in advance with reasonable confidence (errors in general inferior to 10%) using long short-term memory and gated recurrent unit deep neural networks [43,44].

3. Dataset and Pre-Processing

For the present analysis, a paper pulp company provided a three-year data set containing the time series of six variables: electric current (Sensor 1), oil level (Sensor 2), pressure (Sensor 3), rotation velocity (Sensor 4), temperature (Sensor 5), and torque (Sensor 6). All data were collected from sensors with a sampling frequency of one minute.

The dataset contains several repeated values as well as discrepant samples (outliers) that may be due to reading errors or production line stops. Upper outliers might have resulted from errors in sensor reading or recording, while lower outliers are most probably a result of those causes along with programmed or non-programmed downtimes.

In a predictive algorithm, the quality of the underlying data is of extreme importance. Poor quality data implies inaccurate results. For that reason, the dataset was previously processed to increase confidence in the results and facilitate convergence during the learning process.

The units of the several variables are as follows: electric current is measured in amperes (A); oil level is measured in percent of full tank (%Tank); pressure is measured in pascals (Pa); rotation velocity is measured in rotations per minute multiplied by 1000 (RPM \times 1000); temperature is measured in degrees Celsius (°C); torque is measured in Newton-meter (N \times m).

Figure 1 presents the time series collected by each sensor on the six variables.

Figure 1 shows that there are many outliers in the dataset (e.g., null, zeroes and repeated values); repeated values arise by sensory errors or even at the time change. The outliers are replaced by the average value of the variable in the sliding window before the outlier. This method has been described in more detail by Mateus et al. [45].

Therefore, the dataset was filtered using a Python algorithm developed by the authors as follows:

- Repeated values as well as lower and upper discrepant values were removed and replaced by the corresponding variable average value;
- Values beyond three standard deviations from the first and third quartiles on each variable were also replaced by the mean value of the variable in question.

Figure 2 shows the six time series of the variable values collected by the sensors after being filtered by the previously described pre-processing method. As the chart shows, there are no more sudden variations, probably representing outliers, which could impair the machine learning process. Previous studies show that pre-processing discrepant data improves the learning process [43].



Figure 1. Plot of sensor data collected from 2018 to 2020 with a sampling period of 1 min. Raw data as provided by the company.



Figure 2. Data filtered from outliers with a sampling period of 1 min.

4. Artificial Neuronal Network

A sliding window encompasses a continuous subset of a time series dataset, which slides over the latter with a certain step. The window size determines the number of data point samples from the whole dataset to be included in this subset.

The window started with the first w data points (samples) of the time series and slid to the end of the series, in steps of one for an overlapping window, or steps of w samples for a non-overlapping window.

For *n* variables (n = 6 in this case), data from each variable *i* in each sliding window with size *w* were grouped into 15 equal-width bins *j* and the corresponding absolute frequency values (vector $S_{i,j}$), along with the respective average (A_i), median (M_i), standard deviation (SD_{*i*}), variance (V_i) statistics and, finally, 30 ratios ($R_{i1,i2}$), between each pair of

variables *I*, where $i_1 \neq i_2$ make up the input vector *I* that feeds the neural network, as represented in Equation (1).

$$I = (S_{1,1}, S_{1,2}, \dots, S_{n,15}, A_1, \dots, A_n, SD_1, \dots, SD_n, M_1, \dots, M_n, V_1, \dots, V_n, R_{1,2}, \dots, R_{n,n-1})$$
(1)

For each window w: $S_{i,j}$ represents the value of variable i in bin j; A_i is the average value of variable i; M_i is the median value of variable i; V_i is its variance; and, finally, $R_{i1,i2}$ represents the ratios between the variables collected by the sensors.

Data inputs were further standardized using the Standard Scaler library from Sklearn before being fed into the ANN model. Standardization is a technique applied in the preparation of data, with the objective of placing them in a range of common values.

Note that each variable *i* was predicted not only from its respective past data but also from the other five variables.

Time series data were separated into two groups: the first 80% from 1 January 2018 to 27 May 2020 were used for training the model; and the remaining 20% for carrying out the tests.

Tests were carried out with the application of various sizes w of sliding windows. Time windows w of 12, 24, 48, and 72 h were tested (720, 1440, 2880, and 4320 data point samples, respectively).

Neural Network Architecture

The architecture type chosen for the neural network is the multi-layer perceptron, one of the most popular feed-forward architectures, implemented using the Python Sklearn library named MLPRegressor.

The MLPRegressor uses multiple hyper parameters to optimize the generalization of the network model for prediction. Several architecture combinations were tested to find the best possible network configuration.

Adam solver was chosen as the algorithm for optimizing ANN weights, since it is a graph-based optimization algorithm recommended for large datasets, using a logistic sigmoid as the activation function, as represented in Equation (2), for x being the independent variable.

$$f(x) = \frac{1}{(1 + exp(-x))}$$
(2)

Creation vector and tests to find the best value for each alternative ANN configuration took about two days to perform, due to the complexity and size of the dataset, in a shared GPU server AMD EPYC 7552, with 16 Core + Nvidia Tesla T4/V100S.

The authors tested alternative networks with one, two, three, and four hidden layers. Using one layer only yielded quite bad results, while using four layers was quite time consuming. Results from using two and three layers were quite similar, so the authors chose two layers only as the training time was faster without loss of accuracy. Alternative ANN configurations further varied the number of neurons in each layer.

Hence, a network with two hidden layers (150 and 75 neurons, respectively) was chosen, as it showed results very similar to the three hidden layers' architecture but was much faster. Figure 3 depicts the chosen ANN architecture.



Figure 3. Architecture of the artificial neural network.

5. Model Evaluation

To assess the accuracy of the forecast model developed, three popular metrics were used: mean squared error (MSE) presented in Equation (3), and the mean absolute percent error (MAPE) presented in Equation (4) [46].

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2$$
(3)

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|Y_t - \hat{Y}_t|}{|Y_t|}$$
(4)

where Y_t represents the actual value, \hat{Y}_t the predicted value, t is discrete instant time that varies between 1 and n, and n is the total number of data point samples.

6. Tests and Results for Overlapping Sliding Windows

The developed algorithm was tested for forecasts of 15, 30, and 90 days in advance. The training took up to 1000 learning epochs in each of the tests, with overlapping sliding window sizes w of 720, 1440, 2880, and 4320 samples.

Best results were achieved for windows sizes with either 12 or 24 h (720 or 1440 samples). Hence, detailed results will be presented only for these two window sizes.

Tables 1–3 show the results achieved for the six variables (sensors) in terms of MAPE, MSE, and number of iterations (ITER), which the training requires to be completed.

Evaluation results show that it is possible to predict variable (sensor) values with 3 months, 1 month, and 15 days in advance with a reasonable degree of accuracy. Most variables show MAPE errors below 10%.

In general, a window size of 720 samples (12 h) over 1440 samples (24 h), not only has a shorter learning time, but it also yields better accuracy results in terms of MAPE and MSE. Hence, a window size of 720 samples was selected as a good sampling size.

Current Oil Level Velocity Pressure Temperature Torque 1440 Samples 720 Samples 90 Days 3.441 6.231 16.286 4.977 2.696 4.053 30 Days 2.295 4.642 14.643 4.0062.332 4.112 15 Days 2.205 4.306 12.453 3.717 1.734 3.678 Oil Level Current Pressure Temperature Torque Velocity 90 Days 3.623 6.696 21.426 4.8782.612 4.451 30 Days 2.310 5.124 14.034 4.686 2.049 4.234 15 Days 2.541 4.319 4.363 13.633 1.864 3.896

Table 1. Comparative MAPE (%) results.

Table 2. Comparative MSE results.

les		Current	Oil Level	Pressure	Temperature	Torque	Velocity
du	90 Days	1.374	5.804	3.954	2.336	0.638	1.485
) Sa	30 Days	1.094	5.358	4.153	2.201	0.735	1.379
72(15 Days	1.080	5.013	4.056	2.060	0.670	1.454
les		Current	Oil Level	Pressure	Temperature	Torque	Velocity
dun	90 Days	1.413	6.476	4.518	2.271	0.621	1.584
0 Sı	30 Days	1.010	5.385	4.123	2.336	0.632	1.668
144	15 Days	1.173	4.739	3.808	2.374	0.643	1.517

Table 3. Comparative number of iterations.

les		Current	Oil Level	Pressure	Temperature	Torque	Velocity
iduu	90 Days	122	161	363	201	58	164
) Sa	30 Days	120	173	353	230	56	159
720	15 Days	104	132	425	188	63	190
1440 Samples		Current	Oil Level	Pressure	Temperature	Torque	Velocity
	90 Days	136	164	415	217	64	179
	30 Days	126	144	435	249	64	213
	15 Days	105	115	432	208	71	168

7. Results with Overlapping Sliding Windows

Almost all variables show large fluctuations, including striking peaks (see Figures 1 and 2). Hence, to stabilize the output and to visualize better actual and predicted time series values on each variable, they were smoothed using a rolling average filter of 1 day.

Figures 4 and 5 present two examples of actual time series in blue and 90-day forecasts in orange, after smoothing has been applied. Pressure is the most difficult variable and torque is the easiest variable to predict, as shown in Table 1, so they were chosen as examples. According to Table 1, pressure was the variable that had the highest MAPE error and torque was the one that overall had the lowest MAPE error.



Figure 4. Real and 90-day forecast values (after smoothing) for pressure.



Figure 5. Real and 90-day forecast values (after smoothing) for torque.

8. Results with Non-Overlapping Sliding Windows

Using overlapping windows showed good prediction accuracies for all variables. However, their training times are quite large, taking on average more than two days for each variable (using a Cirrus workstation). Hence, non-overlapping windows were assessed to reduce learning time.

Using non-overlapping windows, the input vector in the neural network contains fewer data points, thus making its processing much faster. On average, this method allowed us to reduce the learning time to only seven minutes (using a MacBook Pro M1 from 2020 with 8 GB of RAM with MacOS Monterey).

Using non-overlapping windows yields worse long-term (90 days) forecasts than the previous overlapping window method. However, the short-term (15 days) results are good (see Figure 6). It should be noted though that the neural network is the same, regardless of whether it is for short/medium or long-term predictions. It is only the data included in the input vector that change.



Figure 6. Real and 15-days forecast values (after smoothing) using non-overlapping windows for current.

9. Discussion

9.1. Comparation with TEPEN Vector

The present research corresponds to an optimization of the features of the neural network vector already developed by the authors. The new vector contains new ratios among variables [47]. This is the difference between the two vectors, as is presented in Table 4. This comparison is made for 90-day forecast results.

ples		Current	Oil Level	Pressure	Temperature	Torque	Velocity
720 Samı	Old vector	1.564	5.943	4.741	3.553	0.544	1.842
	New vector	1.374	5.804	3.954	2.336	0.638	1.485
ples		Current	Oil Level	Pressure	Temperature	Torque	Velocity
1440 Sam	Old vector	1.672	7.231	3.670	7.571	0.802	2.843
	New vector	1.413	6.476	4.518	2.271	0.621	1.584

Table 4. Comparative MSE results between old vector and new vector.

Analyzing Table 4, the prediction results of the new vector are generally much better than the old vector, except for the torque parameter that maintains identical values.

9.2. Comparison between LSTM, GRU and Feed-Forward Network

Long short-term memory network is an advanced RNN, a sequential network, that allows information to persist. It can handle the vanishing gradient problem faced by RNN. Long short-term memory network (LSTM) extracts patterns from sequential data and stores these patterns in internal state variables. Each cell can retain important information for a longer period when it is used. Such information properties allow the LSTM to perform well in predicting dynamic sequences [46–48].

The gated recurrent unit (GRU) was designed by Cho et al. [49]. The closed recurrent unit is a special type of optimized recurrent neural network based on LSTM [50,51]. The difference is that the GRU combines the input port and the forgetting port in the LSTM into a single update port [52,53].

Table 5 shows a comparison of prediction models using LSTM, GRU [43,44]. A Short-Term Electric Load Forecast Method Based on Improved Sequence-to-Sequence GRU with Adaptive Temporal Dependence. and the feed-forward model presented in this paper. The comparison is made by analysing the MAPE errors of forecast 30 days in advance for each variable. The 30-day forecast was selected because it was the time gap defined as the objective for the project.

Table 5. Comparison of MAPE results of the prediction models using LSTM, GRU and the feedforward model presented in this paper.

	Current	Oil Level	Pressure	Temperature	Torque	Velocity
GRU-ReLU	2.52	2.94	9.91	2.84	3.03	15.05
GRU-Sigmoid	2.22	2.72	<u>9.29</u>	2.74	2.88	12.42
LSTM-ReLU	2.42	2.92	10.36	2.30	3.72	17.19
MLP (720 Samples)	2.30	4.64	14.64	4.01	2.33	4.11
MLP (1440 Samples)	2.54	4.32	13.63	4.36	<u>1.86</u>	<u>3.90</u>

The first row and second row of the column present the results of the GRU prediction models using the ReLU and Sigmoid activation functions. The third row of the column presents the results of a traditional LSTM model using the ReLU activation function. Finally, the last two lines present the results of the MLP neural network developed, presented, and explained in the previous chapters. Table 5 presents the MAPE results, which were used to evaluate the performance of the algorithm.

Analyzing Table 5, it is concluded that the current parameter has very similar prediction results in all models. The GRU and LSTM models have similar prediction results; however, the GRU-SIGMOID has a slightly lower MAPE error. In the case of the pressure parameter, the GRU models have the best prediction results, with the GRU-SIGMOID the one that achieves the best results. In the temperature parameter, although the results do not show a significant difference, it is the LSTM-ReLU model that presents the smallest prediction error. Regarding torque, the MLP models present the best prediction results, with the best model the MLP-1440 SAMPLES. The biggest difference in results occurs in the velocity parameter, where the MLP models present much better prediction results than the other models, obtaining much lower MAPE errors.

MLP networks are simpler than GRU models and, in turn, the GRU network is simpler than LSTM. Observing the results, it is concluded that both prediction models can predict the future values of an industrial paper press, 30 days in advance, with MAPE, in general less than 10%. The difference in results in the velocity variable is noted, where only feed-forward networks, despite their simplicity, achieved a MAPE error of less than 5%. In short, there is no better overall model, because each variable has the forecast model that best suits its data, as shown in Table 5. Therefore, for optimal prediction, it is important to plan and optimize the machine learning models, with the best model that which achieves the minimum difference for each variable.

9.3. Comparison between XGBoost and Feed-Forward Network

In this section, we compare the ML feed-forward network with another alternative model, namely XGBoost.

Random forest models have, for many problems, a low error. However, for the present problem, they do not have the ability to follow the trend of the parameters to be predicted. They were tested, but the results were not acceptable, and they are not included in Table 6. The first row displays the results of the XGBoost forecast model.

The XGBoost model, in addition to having a very fast training, presents good results in the prediction of electric current, pressure and temperature, with the only disadvantage being the difficulty of identifying peaks of values. It is noted that only MLP models can follow oil level parameter trends. Figure 7 shows an example of a prediction using XGBoost. The XGBoost algorithm needs four minutes to create the vector and train the model. It should be noted that the input vector is with non-overlapping sliding windows and the machine used is a MacBook Pro M1 of 2020 with 8GB of RAM with MacOS Monterey.

Table 6. Comparison between XGBoost and feed-forward network using MSE.

	Current	Oil Level	Pressure	Temperature	Torque	Velocity
XGBOOST	1.075	4.953	0.618	1.304	1.576	3.356
MLP (720 Samples)	1.374	5.804	3.954	2.336	0.638	1.485
MLP (1440 Samples)	1.413	6.476	4.518	2.271	0.621	1.584



Figure 7. Real and 30-day forecast values (after smoothing) for temperature using XGBoost.

9.4. Discussion

XGBoost is a scalable and highly accurate implementation of gradient boosting that pushes the limits of computing power for boosted tree algorithms, being built largely for energizing machine learning model performance and computational speed.

One of advantages is that the proposed method can perform short-, medium-, and long-term prediction on other equipment, providing there are the necessary data and processing power available. The features used and developed to feed the machine learning models should be available in a wide range of industrial equipment. Nonetheless, for each specific situation the data pre-processing or the neuronal network architecture may need to be modified and there are no *a priori* guarantees of similar results.

This study is not intended to predict sensor failures. This study focuses on predicting the future behaviour of the machine in the short, medium, and long term. Sensor failures will be detected as a malfunction, which needs further analysis and diagnosis. That has also been clarified in the paper.

The results of those predictions can then be processed, for instance, using a classifier neural network, to classify the asset's condition into one of the following states: failure, alert, or good functioning.

10. Conclusions

Forecasting is very important to make better decisions in maintenance and other areas. Predicting the probable future behaviour of an asset brings numerous benefits. For instance, based on accurate predictions, and knowing the respective nominal operating values recommended by the asset's manufacturer, it is possible to identify anomalies in advance for the equipment in the short, medium, and long term.

The proposed algorithm makes it possible to know the behaviour of a pulp paper press in the long term, supported by the time series acquired from sensors installed on it. This way it is possible to optimize long-term programmed stops and to avoid production downtimes. This prediction model will be enhanced by the addition of a classification network that will classify the machine into one of three states: failure, alert, or good functioning.

This paper presents a valuable comparison between the input vector of the neural network using overlapping and non-overlapping sliding windows, presenting the results of the tests performed, with unequivocal conclusions about the advantages and limitations of each technique used.

The number of data points present in the neural network input vector, as well as the prediction gap, have a direct impact on the prediction accuracy. On the one hand, a larger sliding window increases the prediction errors, but a smaller window has difficulty in predicting peaks. On the other hand, the larger the prediction gap, the more difficult the prediction becomes.

The results achieved for the short term, midterm and long term were comparable to or better than the state of the art. Long-term forecasts using overlapping windows showed very good accuracies, because the predictions of most parameters present MAPE errors below 10%, that is the objective of the research presented in this paper, as shown in Section 9.2. However, they take a large processing time. Short-term forecasts using non-overlapping windows can significantly reduce this shortcoming.

The XGBoost Model presents fast and good results in the prediction of electric current, pressure, and temperature. XGBoost had errors of MSE less than two.

Future work includes applying this method to other variables and comparing it against alternative machine learning models for prediction. Additionally, other machine learning methods, such as unsupervised clustering, will be studied to classify the future condition state of the asset based on the forecasts resulting from the presented ANN.

The number of input features can also be optimized using techniques such as principal component analysis (PCA), or probabilistic principal component analysis (PPCA). Other approaches, namely hidden Markov models will also be explored.

Author Contributions: Conceptualization, J.A.R.; Formal analysis, J.A.R.; Investigation, J.A.R.; Methodology, M.M.; Project administration, J.T.F. and M.M.; Software, J.A.R.; Supervision, J.T.F. and M.M.; Validation, M.M. and R.J.G.M.; Writing–original draft, J.A.R.; Writing–review & editing, J.T.F., M.M., R.J.G.M. and A.J.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 ProjectPOCI-01-0145-FEDER-029494, and by National Funds through the FCT—Portuguese Foundationfor Science and Technology, under Projects PTDC/EEI-EEE/29494/2017, UIDB/04131/2020, and UIDP/04131/2020. 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 n° 022153.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
FF	Feed Forward
ITER	Iterations
MAPE	Mean Absolute Percentage Error
MLP	Multi-Layer Perceptron
MSE	Mean Square Error
PCA	Principal Component Analysis

- PPCA Probabilistic Principal Component Analysis
- RF Random Forest.
- RNN Recurrent Neural Network
- RPM Rotations Per Minute

References

- 1. Kumar, U.; Galar, D.; Parida, A.; Stenström, C.; Berges, L. Maintenance Performance Metrics: A State-of-the-art Review. J. Qual. Maint. Eng. 2013, 19, 233–277. [CrossRef]
- Standards, E. BS EN 13306:2017 Maintenance. Maintenance Terminology. Available online: http://hadidavari.com/wp-content/ uploads/2018/12/BS-EN-13306-2017.pdf (accessed on 9 November 2021).
- 3. Rao, B.K.N. Handbook of Condition Monitoring; Elsevier: Amsterdam, The Netherlands, 1996. ISBN 978-1-85617-234-9.
- Carnero, M.C. Selection of Diagnostic Techniques and Instrumentation in a Predictive Maintenance Program. A Case Study. Decis. Support Syst. 2005, 38, 539–555. [CrossRef]
- Selcuk, S. Predictive Maintenance, Its Implementation and Latest Trends. Proc. Inst. Mech. Eng. Part B: J. Eng. Manuf. 2017, 231, 1670–1679. [CrossRef]
- Patwardhan, A.; Verma, A.K.; Kumar, U. A Survey on Predictive Maintenance Through Big Data. In Proceedings of the Current Trends in Reliability, Availability, Maintainability and Safety; Kumar, U., Ahmadi, A., Verma, A.K., Varde, P., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 437–445.
- 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]
- 8. 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]
- Rodrigues, J.; Torres Farinha, J.; Marques Cardoso, A. Predictive Maintenance Tools—A Global Survey. Wseas Trans. Syst. Control 2021, 16, 96–109. [CrossRef]
- 10. Rodrigues, J.; Farinha, J.; Mendes, M.; Mateus, R.; Cardoso, A.J.M.; Rodrigues, J. Short and Long Forecast to Implement Predictive Maintenance in a Pulp Industry. *Eksploatacja i Niezawodnosc—Maint. Reliab.* **2021**, *24*, 33–41. [CrossRef]
- 11. Carvalho, T.P.; Soares, F.A.A.M.N.; Vita, R.; da Francisco, R.P.; 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]
- Galar, D.; Stenström, C.; Parida, A.; Kumar, R.; Berges, L. Human Factor in Maintenance Performance Measurement. In Proceedings of the 2011 IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, 6–9 December 2011; pp. 1569–1576.
- 13. Sahal, R.; Ali, M.I.; Breslin, J. 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]
- 14. Hashemian, H.M. State-of-the-Art Predictive Maintenance Techniques. *IEEE Trans. Instrum. Meas.* 2011, 60, 226–236. [CrossRef]
- 15. Shrivastav, O.P. Industrial Maintenance: A Discipline in Its Own Right. *World Trans. Eng. Technol. Educ.* 2005, 4, 4.
- Poór, P.; Basl, J.; Zenisek, D. Predictive Maintenance 4.0 as next Evolution Step in Industrial Maintenance Development. In Proceedings of the 2019 International Research Conference on Smart Computing and Systems Engineering (SCSE), Kelaniya, Sri Lanka, 28 March 2019; pp. 245–253.
- 17. Monchy, F.; Mirochnikoff, Y. La Fonction Maintenance: Formation à La Gestion de La Maintenance Industrielle. Engineering 1987.
- Wang, S.-C. Artificial Neural Network. In *Interdisciplinary Computing in Java Programming*; Wang, S.-C., Ed.; The Springer International Series in Engineering and Computer Science; Springer: Boston, MA, USA, 2003; pp. 81–100. ISBN 978-1-4615-0377-4.
- 19. Nigrin, A. Neural Networks for Pattern Recognition; MIT Press: Cambridge, MA, USA, 1993. ISBN 978-0-262-14054-6.
- Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; Vinyals, O. Understanding Deep Learning (Still) Requires Rethinking Generalization. Commun. ACM 2021, 64, 107–115. [CrossRef]
- Oymak, S.; Soltanolkotabi, M. Toward Moderate Overparameterization: Global Convergence Guarantees for Training Shallow Neural Networks. *IEEE J. Sel. Areas Inf. Theory* 2020, 1, 84–105. [CrossRef]
- Multilayer Perceptron—An Overview. ScienceDirect Topics. Available online: https://www.sciencedirect.com/topics/computerscience/multilayer-perceptron (accessed on 9 November 2021).
- 23. Impact of Deep Learning-Based Dropout on Shallow Neural Networks Applied to Stream Temperature Modelling. Elsevier Enhanced Reader. Available online: https://reader.elsevier.com/reader/sd/pii/S0012825219305549?token=CF6227805A1730A8 59BAEE2F77EBDCB50B400FCE09411CAF51B569500652A4EE3FA108283C5770F90F0470C4BF3EE5EE&originRegion=eu-west-1&originCreation=20211109021345 (accessed on 9 November 2021).
- Jonsson, P. Company-Wide Integration of Strategic Maintenance: An Empirical Analysis. Int. J. Prod. Econ. 1999, 60–61, 155–164. [CrossRef]
- Carnero, M. An Evaluation System of the Setting up of Predictive Maintenance Programmes. *Reliab. Eng. Syst. Saf.* 2006, 91, 945–963. [CrossRef]
- 26. Yamashina, H. Japanese Manufacturing Strategy Competing with the Tigers. Bus. Strategy Rev. 1996, 7, 23–36. [CrossRef]
- 27. Dong, W.; Huang, Y.; Lehane, B.; Ma, G. XGBoost Algorithm-Based Prediction of Concrete Electrical Resistivity for Structural Health Monitoring. *Autom. Constr.* **2020**, *114*, 103155. [CrossRef]

- Qi, Y. Random Forest for Bioinformatics. In *Ensemble Machine Learning: Methods and Applications*; Zhang, C., Ma, Y., Eds.; Springer: Boston, MA, USA, 2012; pp. 307–323. ISBN 978-1-4419-9326-7.
- Rodrigues, J.; Cost, I.; Farinha, J.T.; Mendes, M.; Margalho, L. Predicting Motor Oil Condition Using Artificial Neural Networks and Principal Component Analysis. *EiN* 2020, 22, 440–448. [CrossRef]
- Allah Bukhsh, Z.; Saeed, A.; Stipanovic, I.; Doree, A.G. Predictive Maintenance Using Tree-Based Classification Techniques: A Case of Railway Switches. *Transp. Res. Part C Emerg. Technol.* 2019, 101, 35–54. [CrossRef]
- Elhag, T.M.S.; Wang, Y.-M. Risk Assessment for Bridge Maintenance Projects: Neural Networks versus Regression Techniques. J. Comput. Civ. Eng. 2007, 21, 402–409. [CrossRef]
- Balluff, S.; Bendfeld, J.; Krauter, S. Short Term Wind and Energy Prediction for Offshore Wind Farms Using Neural Networks. In Proceedings of the 2015 International Conference on Renewable Energy Research and Applications (ICRERA), Palermo, Italy, 22–25 November 2015; pp. 379–382.
- Deepika, T.; Prakash, P. Power Consumption Prediction in Cloud Data Center Using Machine Learning. Int. J. Electr. Comput. Eng. 2020, 10, 1524–1532. [CrossRef]
- Hongxiang, T.; Yuntao, L.; Xiangjun, W. Application of Neural Network to Diesel Engine SOA. In Proceedings of the 2011 Third International Conference on Measuring Technology and Mechatronics Automation, Shanghai, China, 28 January 2011; Volume 1, pp. 555–558.
- Kittisupakorn, P.; Thitiyasook, P.; Hussain, M.A.; Daosud, W. Neural Network Based Model Predictive Control for a Steel Pickling Process. J. Process Control 2009, 19, 579–590. [CrossRef]
- Okoh, C.; Roy, R.; Mehnen, J. Predictive Maintenance Modelling for Through-Life Engineering Services. *Procedia CIRP* 2017, 59, 196–201. [CrossRef]
- Makridis, G.; Kyriazis, D.; Plitsos, S. Predictive Maintenance Leveraging Machine Learning for Time-Series Forecasting in the Maritime Industry. In Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 20–23 September 2020; pp. 1–8.
- Zhagparov; Buribayev, Z.; Joldasbayev, S.; Yerkosova, A.; Zhassuzak, M. Building a System for Predicting the Yield of Grain Crops Based on Machine Learning Using the XGBRegressor Algorithm. In Proceedings of the 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), Nur-Sultan, Kazakhstan, 28–30 April 2021; pp. 1–5.
- 39. 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]
- 40. Fonseca, D.J.; Navaresse, D.O.; Moynihan, G.P. Simulation Metamodeling through Artificial Neural Networks. *Eng. Appl. Artif. Intell.* **2003**, *16*, 177–183. [CrossRef]
- Suomela, J. Condition Monitoring of Paper Machine with Thermal Imaging; Maldague, X.P., Rozlosnik, A.E., Eds.; SPIE: Orlando, FL, USA, 2002; pp. 143–150.
- Bissessur, Y.; Martin, E.B.; Morris, A.J. Machine Condition Monitoring for Consistent Paper Production. Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng. 1999, 213, 141–151. [CrossRef]
- 43. 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. [CrossRef]
- Mateus, B.; Mendes, M.; Farinha, J.; Assis, R.; Cardoso, A.J.M. Comparing LSTM and GRU Models to Predict the Condition of a Pulp Paper Press. *Energies* 2021, 14, 6958. [CrossRef]
- 45. Mateus, B.; Farinha, J.T.; Mendes, M.; Martins, A.B.; Cardoso, A.M. Data Analysis for Predictive Maintenance Using Time Series and Deep Learning Models—A Case Study in a Pulp Paper Industry. *TEPEN* **2021**, *IncoME-VI* 2021. in press.
- 46. Chicco, D.; Warrens, M.J.; Jurman, G. The Coefficient of Determination R-Squared Is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [CrossRef] [PubMed]
- Rodrigues, J.A.; Farinha, J.T.; Cardoso, A.J.M.; Mendes, M.; Mateus, R.J.G.M. Prediction of Sensor Values in Paper Pulp Industry Using Neural Networks. TEPEN 2021, IncoME-VI 2021. in press.
- 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]
- Simulating Time-Series Data for Improved Deep Neural Network Performance. IEEE Jorunal Magazine. IEEE Xplore. Available online: https://ieeexplore.ieee.org/document/8835043 (accessed on 21 July 2022).
- 50. Frontiers. Continuous Timescale Long-Short Term Memory Neural Network for Human Intent Understanding. Available online: https://www.frontiersin.org/articles/10.3389/fnbot.2017.00042/full (accessed on 21 July 2022).
- 51. Santra, A.S.; Lin, J.-L. Integrating Long Short-Term Memory and Genetic Algorithm for Short-Term Load Forecasting. *Energies* **2019**, *12*, 2040. [CrossRef]
- 52. Li, D.; Sun, G.; Miao, S.; Gu, Y.; Zhang, Y.; He, S. A Short-Term Electric Load Forecast Method Based on Improved Sequence-to-Sequence GRU with Adaptive Temporal Dependence. *Int. J. Electr. Power Energy Syst.* 2022, 137, 107627. [CrossRef]
- Liu, X.; Lin, Z.; Feng, Z. Short-Term Offshore Wind Speed Forecast by Seasonal ARIMA—A Comparison against GRU and LSTM. Energy 2021, 227, 120492. [CrossRef]