Expert Systems With Applications 186 (2021) 115699

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Application of data-driven models to predictive maintenance: Bearing wear prediction at TATA steel

X. Chen, Jos Van Hillegersberg *, E. Topan, S. Smith, M. Roberts

Tata Steel Shotton, United Kingdom
University of Twente, the Netherlands

ARTICLE INFO

Keywords:
Predictive maintenance
Industry 4.0
Data-driven
Machine learning

ABSTRACT

Industries that are in transition to Industry 4.0 often face challenges in applying data-driven methods to improve performance. While ample methods are available in literature, knowledge on how to select and apply them is scarce. This study aims to address this gap reported on the design and implementation of data-driven models for predictive maintenance at TATA Steel, Shotton. The objective of the project is to predict the wearing behaviour of the components in the steel production line for maintenance activity decision support. To achieve the predictive maintenance goal, the approach applied can be summarized as follows: 1. business understanding and data collection, 2. literature review, 3. data preparation and exploration, 4. modelling and result analysis and 5. conclusion and recommendation. The data-driven methods that were analysed and compared are: Partial Least Squares Regression (PLSR), Artificial Neu- ral Network (ANN) and Random Forest(RF). After cleaning and analysing the production line data, predictive maintenance with the current available data in TATA Steel, Shotton is best feasible with PLSR. The study further concludes that, predictive maintenance is likely to be feasible in similar industries that are in transition to industry 4.0 and have growing volumes of production data with varying quality and detail. However, as illustrated in this case study, careful understanding of the industrial process, thorough modeling and cleaning of the data as well as careful method selection and tuning are required. Moreover, the resulting model needs to be packaged in a user friendly way to find its way to the job floor.

1. Introduction

Industry 4.0, also referred to as Smart Industry, includes smart manufacturing, smart factory, lights-out manufacturing and Internet of Things (IoT) (Sniderman, Mahto, & Cotteleer, 2016). The essential idea of industry 4.0 is to apply automation, connectivity and big data exchange in manufacturing processes. In smart industry applications, automation is not limited to production but also includes automating decision making. One of the application areas is predictive maintenance. While various studies have discussed the potential of using big data, such as sensor data, on improving predictive maintenance, actual industry applications and reported experiences are scarce. In this study we aim at filling this gap, reporting on the design and implementation of a data-driven model and a software application of the model for predictive maintenance of rotating metal bush at Tata Steel, Shotton, UK.

The prediction of rotating metal-to-metal contact wear is one of the critical areas 15 in predictive maintenance, as rotating mechanical components such as bearing and bush are widely implemented in machines and failures of these components cause down time of machinery and the entire production line. Existing prognostic methods can be classified into three categories: “Model-based prognostics”, “data-driven prognostics” and “Reliability-based prognostics” (Tobon-Mejia, Medjaher, Zerhouni, & Tripot, 2012). “Model-based prognostics” requires deep knowledge of system functions. Mathematical models are built to represent the system behaviour including component degradation process. However, systems are often com- plex in reality, thus mathematical modelling is computationally expensive and various assumptions need to be made while building models. Reliability based 25 prognostics can also be referred to as “experience-based prognostics” which uses historical data during a significant period of time and discover the statistical distribution of each parameter. Poisson, exponential, weibull and log-normal distribution have been proposed in the literature for failure time distribution. This approach is easy to implement when historical data from a significant period of time is available. However, the

* Corresponding author.
E-mail addresses: xinyi.chen@tatasteeleurope.com (X. Chen), j.vanhillegersberg@utwente.nl (J. Van Hillegersberg), e.topan@utwente.nl (E. Topan), steve.c.smith@tatasteeleurope.com (S. Smith), Matthew.Roberts@tatasteeleurope.com (M. Roberts).

https://doi.org/10.1016/j.eswa.2021.115699
Received 1 September 2020; Received in revised form 29 July 2021; Accepted 30 July 2021
Available online 14 August 2021

prediction result is generally less precise than results from model-based and data-driven methods. "Data-driven prognostics" aim at getting information from the raw data that is mainly from sensors. It typically uses artificial intelligence or statistical models to learn the wearing behaviour and to predict the condition in the future. A system based on a data-driven model operates automatically without considering the explanatory power to the real system or parameters. Although data-driven methods are not computational expensive, they generally still provide good prediction results for systems where it is easy to monitor data representing wearing behaviour or system failure. However, critical data such as those related to failure are often missing in industry since failure has been prevented in every way possible, due to the huge cost of down time (Zschech, Heinrich, Bink, & Neufeld, 2019). Think for example of preventive periodical replacement components, a often expensive way to prevent failure. This project conducts a data-driven bush wear prediction based on data from the steel production line in Tata steel Shotton (UK). A critical part of the production line is maintained by replacing the component every four weeks, which leads to the fact that barely any failures occur. As the bush operates in a melted zinc pot and there are currently no sensors connected to it, the remaining width of the bush is measured every four weeks, thus the wearing data is only avail- able at the end of each maintenance cycle. For simplicity reasons, cycle will be used in the rest of this paper refer to maintenance cycle. Although this study is conducted at TATA Steel and addresses its bush wear case, we believe that the approach can be extended to many other similar environments, where data increasingly becomes available from production systems and maintenance operations, but may still be lacking for parts of the production or maintenance process. In such cases, data is usually varying in volume and complexity, and extensive historical data are missing, as several production components still lack near real-time sensing capabilities.

This study aims at contributing in three ways to literature and practice:

1. We demonstrate the use of data-driven models for predictive maintenance in a steel industry context and in that way contribute to bridging the gap in the current state of the art and the theory for industry in transition to industry 4.0. 2. We demonstrate how to predict the bearing wear up to an 90% accuracy in Tata Steel UK while vibration data are not available and wearing measures are limited. The Partial Least Squares Regression (PLSR) model used for bearing wear prediction can be well generalized to situations when data are at high dimension but limited sample size.

3. We propose a method for data processing in a real-world industry context when production data are large in volume and varying in detail and accuracy. We discuss generalizability of our findings to businesses in transformation to the digital era, where the infrastructure of data logging, integration and analytic are not fully developed.

The structure of this paper is as follows: Section 2 presents the approach and research design, Section 3 presents a literature review of the current state of the art. Section 4 describes the current situation at Tata Steel and the data collection process. Section 5 explains the data preparation steps and Section 6 gives the modelling process and results, which are further discussed in Section 7. The software tool development is presented in Section 8 and we conclude the paper in Section 9.

2. Research design and approach

The main research question has been set as: 'How can data-driven methods be applied to predictive maintenance in industries that aim to benefit from industry 4.0 applications?'

To answer this main research question, we follow the approach shown in (Fig. 1) The research methods deployed include on-site interviews, data collection and production line inspection to achieve Business Understanding, Literature Re- view, Data Preparation and Exploration, Modelling and Result Analysis. Each sub-question is answered in each subsequent section. Our approach follows an agile way of working, where we always go back to the previous steps when problems appear to the current step or additional information leads to reconsider earlier steps in the process.Fig. 1

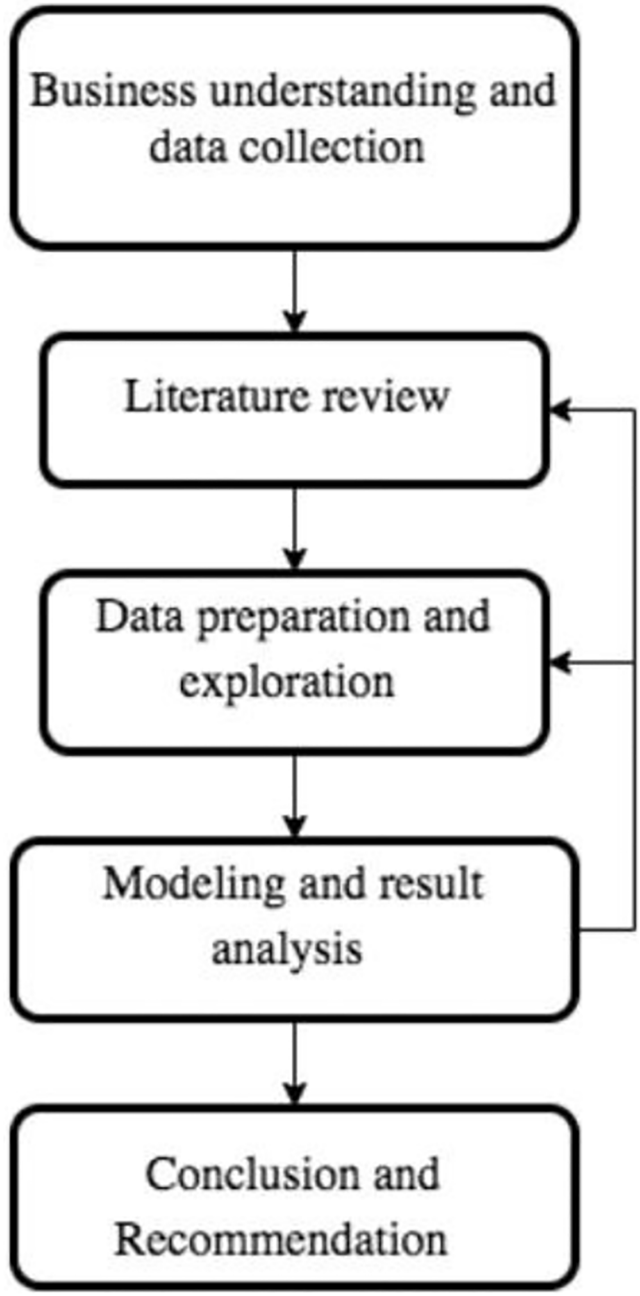

Fig. 1. Research approach.

3. Literature review

As the goal of this project is mainly to investigate the use of data-driven methods to predict wearing condition, we conducted a literature review to explore the cause of mechanical degradation and into methods to predict degradation, especially metal to metal contact wear. In (Cubillo, Perinpanayagam, & Esperon-Miguez, 2016), metal–metal contact wearing is classified into two stages: adhesive wear and fatigue wear. Adhesive wear is due to surface contact that leads to material transfer or loss from surface (Bayer, 2004). Fatigue wear, however, generates surface cracks that result in severe damage after a critical number of cycles. Adhesive wear can be modelled with Archard's law as follows:

$$\text{Volume}[\text{wear}] = K * \text{ Load* Sliding Distance}/(3\text{*Material Hardness}) \quad (1)$$

where K is the wear coefficient that can be adjusted to fit more complex wear models. Methods for predictive maintenance in terms of wearing condition can be generally classified into three categories (Tobon-Mejia et al., 2012), namely mathematical model based methods, experience based methods and data-driven meth- ods. Mathematical modelling based methods produce good results, but require deep knowledge on system functions, which is a viewed as a disadvantage as this is often hard to obtain. Experienced based methods make decisions based on empirical events and often produce poor prediction results. Data-driven methods investigate the problem mainly from a data perspective. Machine learning is often used to learn the data patterns. Predictive maintenance using data-driven methods requires failure data as target variables and featured variables to de- scribe input data (Gouriveau, Ramasso, & Zerhouni, 2013). However, failure data is often not available in industries as down times are prevented at nearly any cost because of high down time cost. Artificial neural networks, support vector machines and random forest are the most implemented data-driven methods. In addition, hidden Markov models (HMM) with strict data structures and reliable computing performance have also drawn much attention. In recent years, many researchers have applied HMM in tool condition monitoring and achieved good results (Liao, Gao, Lu, & Lv, 2016). Some recent papers are taken into consideration (Zhang, Zhang, Wang, & Habetler, 2019, Yang, Yin, Chang, Gao, & Yang, 2020, Schwendemann, 2021, Ding, Yang, Cheng, & Yang, 2021, Yang, Lei, Jia, & Xing, 2019). These work are on bearings with more state of the art methods (variants of NN such as RNN, CNN, FTNN.) However, the data used for wearing prediction are mostly experimental ones, which means machines are set up and run to failure for a number of times and data are collected from the sensors installed on the machine. In this way, critical parameters are fully obtained. Unlike these papers our contributions mainly lies in practice. We are aiming at solving existing problems. Such ideal data sets are hard to get in real industry. A limited number of studies have concentrated on predictive maintenance with missing labels. Zschech et al. (2019) investigated prognostic model development with missing labels. These models first use unsupervised learning, such as clustering techniques to create labels and then use supervised learning, namely a Recurrent neural network (RNN), to predict future labels. In Amruthnath and Gupta (2018a), Amruthnath and Gupta (2018b) unsupervised learning has been investigated to detect faults early in predictive maintenance on experimental data. T-statistics, k-means clustering, c-means clustering and hierarchical clustering are implemented and compared. The results confirm that unsupervised learning can detect faulty behaviour and the clustering results are similar. Langone et al. (2015) propose a least square support vector machine(Ls-SVM) framework for maintenance strategy optimization based on real-time condition monitoring. It uses both clustering (unsupervised learning) and a supervised leaning method namely nonlinear auto-regression (NAR) on, however, different data sets, and conclude that supervised learning can achieve better results but is computationally more expensive. The two methods combined may result in an optimised maintenance cost configuration.

In terms of data processing, statistical features are widely used in wear prediction (Zschech et al., 2019). Vibration signals are used as indicator of the condition in most of the literature. Some literature use current and or frequency (Cipollini, 2018) as it is commonly considered available even though not in the case of this paper. Acoustic emission is another commonly used variable (He, 2017) which is also not applicable in this case Data cleaning and huge data set handling, according to (Brown, 2015), should include: data analysis, definition of transformation workflow and mapping rules, schema-related data transformation, verification, transformation and back-flow of cleaned data. These steps are provided as the guide for data cleaning to detect errors and inconsistencies, transforming data into standard format with the least possible manual inspections and verifying the correctness of data transformation. Panagiotis (2018) is one of the few works using real life data. However, the data used is logged events, different from sensory and production data, and in the context of flight failure which is also different from the context industrial production settings, like the heavy steel industry.

This paper intends to address several gaps in the current state of art. We have not identified any approaches for data cleaning in the literature regarding predictive maintenance. Thus, we briefly describe our approach in handling a large data set in this paper. Moreover, we report on the feasibility of using alternative variables when either failure data or condition indicator variables are available. Finally, we focus on capturing the heavy industrial context where complex process are involved and further investigate how to develop data-driven models in a real heavy industry or manufacturing setting. This work has proven the feasibility of component condition monitoring which contributes to maintenance practice of heavy industry of over-maintaining their assets so creates cost savings. Different from most of the literature, we address the challenges faced by industries, aiming to apply industry 4.0 concepts while essential data may be missing and data may lack consistent quality.

4. Maintenance decisions at Tata steel Shotton

Till the start of this project, maintenance decisions have been made based on experience by the engineers in Tata Steel, Shotton. Specifically, the current maintenance cycle of the pot gear (Fig. 2) is approximately four weeks where no other faults occur such as strip break. At the end of every four weeks, the line is shut down and the pot gear is replaced, even if there has not been any failure. Engineers have been trying out larger bush diameters to prolong the maintenance cycle (the time interval of replacing the pot gear).

To collect data on the wear from different perspectives, we extracted data from in total four separate production systems/sources. The four data sources 185 are: The Set-up sheet, the Data warehouse and two sensor-based data sources for system tracking. The set-up sheetA.16 records data regarding component properties such as roll diameter, roll condition before the rolls are put into the zinc bath and the bush condition before and after each maintenance cycle. Bush wear data is logged after each maintenance cycle approximately every 4 weeks (data was registered starting from 15-05-2019). The data in the set up sheet are logged manually originally in paper and made digitally during this work. Human measurement errors are hard to avoid and are considered when further analysing the data. In our paper, our primary goal is not to propose a novel approach that proves to perform well in comparison to ones in the literature. Instead, our goal is to predict the wear of the bearing that is sunk in a zinc pot considering its special features and show that this provides useful information. These special features depend on the structure of the bearing which also differs from the normal bearing. No sensors are connected to the bearing which cannot be seen and heard either. The only way to maintain the bearing condition is to replace it every 4 weeks without predictions. Based on the current practice, our proposed method can provide big improvement as in to monitor the wear to avoid over-maintenance. This is applicable in every galvanizing line in steel industry where coating is involved.

The Data warehouse records mainly procurement-related dataA.17, such as the product that the customer ordered, the product delivered, the team ID and batch ID etc. Data are logged approximately every half hour. The data from different interfaces to the data warehouse are not all logged at the same time point. Some parameters such as ‘coil width’ appear repetitively, but can’t be interpreted logically. For example, the finished coil width is supposed to be narrower than the ordered and received coil width, but they are logged larger. Such phenomena, to a certain extent, reflect the poor data quality in the data warehouse. Two sensor-based data sources (IBA EMASS system) are logging process-

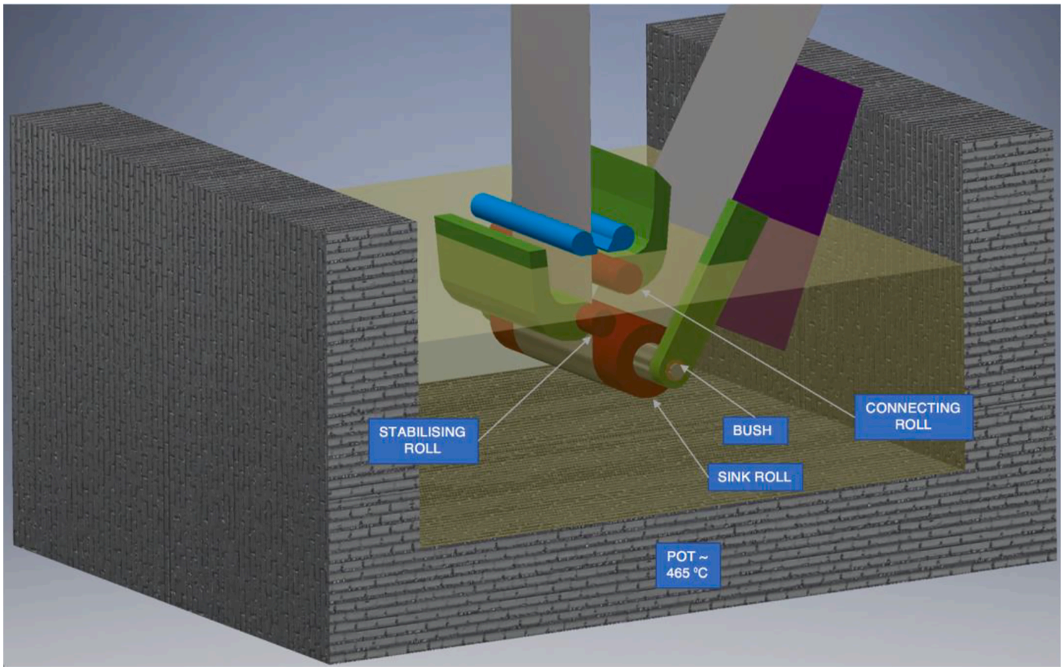

Fig. 2. Pot gear.

related parameters in a much higher frequency (more than one observations per second in EMASS and more than one observation per minute in IBA) see appendixA.19 and A.18.

We used the statistical software package "R" for data reading, cleaning and integration because R is capable of handling the large data size of the data from the sensor systems. Other tools we tried with the data sets, such as OpenRefine, could not handle the vast volumes of sensor data.

5. Data preparation and exploration

5.1. Data reading and cleaning

Four data sources are relevant to support data-driven maintenance decisions: IBA, EMASS, the Setup sheet and the data warehouse. Processed sample data can be found in Appendix A. In Appendix B lists of variables and their explanations are presented for each data source. IBA and EMASS are sensor-based data sources whereas the Set-up sheet, the data warehouse component and the product related data, are logged manually. The data size of IBA and EMASS is much larger than the other sources and the data format thus required more time when cleaning the data.

The data cleaning process is more close to data engineering, as the cleaning algorithm has to be built based on the specific raw data structure and problems will keep on occurring until the data are fully structured. It may take a long time only to read the data due to memory limit of the system and the software.

The software used here is R which can read 2 GB to 4 GB data at a time with reasonable speed. A maximum of 2 billion indices can be stored in memory. It is important to split the file into smaller pieces and clean them separately. If the sub-files are still not readable because they are over-sized, it is important to run operations before reading the file, such as skipping lines while reading to keep the data size fit for the software. The IBA data from 17/03/2019 to 13/09/2019 stores 3,119,156 observations and 687 variables in one single text file. It was not feasible to read it as a whole so the text file was split into 10 sub-files using R and read again. Based on the suggestions of domain experts, 27 variables are selected as essential variables. After reducing the variables to 27 it becomes feasible to read the whole dataset as the data size is reduced. Time to read the data set was reduced to around 15 min. After reading the data set, then it is impossible to conduct any further operations on the data due to lack of memory thus we wrote a script to reduce the size of the data set by only read certain rows in data file per day. The rows are selected such that the data used cover the whole day. EMASS data which is the largest of all stores over 5 million observations per day. At some days over 6 billion observations are recorded. The EMASS is logging over 200 text files per day and zips them automatically. On the system we used, R is able to read a 6-billion observation text file in one and half hours but no additional operations could be executed. As the data size of one day is already large and as there are thousands of observations within one minute, the data are read by skipping every 80–500 rows depending on the data size for that day.

Set-up sheets and data warehouse are logged manually, thus the logging frequency is not as high as that of the sensor-based data. 20,232 observations were read from the data warehouse using R and the parameters logged in the setup sheet were manually transferred into Excel files.(15–05-2019 to 01–12-2019) Data quality is checked by summarizing the data and by comparing the data to the norm values provided by the domain expert of Tata steel Shotton. For example when exploring tension which is a critical parameter, we have extracted its maximum, minimum, mean and median values and compare them with the norm value provided by domain expert. We found that data had hardly ever been logged within the norm, indicating that the data is generally messy with poor quality.

As this project has the time span of half year and wearing measures are recorded since the project started, the data samples that are used in modelling are gradually increased as a new sample is becoming available each time when the pot gear get replaced. Initially, we use 6 samples to train and cross-validate the model, and eventually expand to 9 samples to evaluate the model. The feature selection process is done together with modelling based on the initial 6 samples. The detailed feature selection and modelling process can be found in Section 6.

In the end, each sample contains normalized features (Table1) and the wearing measure of the corresponding maintenance cycle. A summary process of data reading and processing can be found in Fig. 3. Missing values and outliers are deleted and filled with normal value (norm of the variable) suggested by the data itself. The way we define outliers and norm value is by making a summary of each variable, list

Table 1
Selected Feature.

Features
Total Length
Scrap Length
Total Surface
Mean Tension
Minimum Tension
Maximum Tension
Median Tension
Skewness Tension
Kurtosis Tension
Standard Deviation Tension
RMS Tension
Remaining Bush Width Days
Roll Diameter

the average, maximum and minimum value and then check with domain expert to know the normal range for the specific variable. If the value of the variable falls far away from the normal range we view it as outlier. The norm value of variable is defined based on the knowledge of domain expert. Detailed code and operations can be found in Appendix C, D and E. Outliers are removed but only when we are sure that it is not the right value.

5.2. Data exploration

We implement unsupervised learning (hierarchical clustering and K-means clustering techniques) to identify abnormalities in the data. Validation for unsupervised learning has been done by looking back at the events recorded in the shift report. Nearly all dates that have been clustered out can be confirmed with either planned or unplanned issues that happened. However there are also quite some days with abnormal events that have not been clustered out. The main reason is that the data set only contains data from one section of the production line and the fact that some variables are logged wrongly decreases data quality. If data from the whole line can be collected (with good data quality), the clustering performance should improve and this provides a new direction of predictive maintenance project for the future. Moreover, advanced unsupervised learning techniques may be worth investigating to increase the clustering performance.

6. Modelling and result analysis

In this section we select models and apply to the data to predict the wear and report on the results with a comparison. We investigate the performance of partial least squares regression, artificial neural networks and random forest. The reason for choosing neural network and random forest is that they are widely used in current state of art and in some cases provide good results. Especially, the neural network has shown great performance and therefore can be considered as a benchmark method. The PLSR rarely appears in predictive maintenance context, however, PLSR in theory is suitable for small sample size with high dimension and that is the characteristic of our data set after processing and cleaning. That is why PLSR is tried out and compared with other commonly used methods. These techniques have shown promising performance in theoretical settings but far less is known on their performance in actual industry settings with big but fragmented and inconsistent data. Data are normalized before being put into models to keep different variables comparable. The target variable (maximum wear width of the bush) are not normalized. 6.1. Partial Least Squares Regression By using the features selected according to the Partial Least Squares (PLS) theory (sliding distance and force related variables), we use ”Leave One Out” cross validation to evaluate the performance of Partial Least Squared Regression (PLSR) model. Every time we run the model we will use a different maintenance cycle of the bush as test sample and the rest of the cycles as training ones. This is to make sure the model has the most variance of samples to learn from (as many training samples as possible). As PLSR is an extension of principle component regression, the data are projected into lower dimensions as “latent variables”. The number of latent variables are called “the number of components” during the experiments from now on. The number of components has to be chosen before predicting new results. Many techniques can be used to choose the most effective number of components, here we plot the number of components against Root Mean Square Error of Prediction (RMSEP) and choose the most effective number of component when the RMSEP is the lowest. The modelling process has been first conducted with the initial sample size of 6 RMSEP and using different sample as test data. The result can be found in Fig. 4. As the prediction error is very similar on the initial six samples. We plot a learning curve using 3 additional samples (Fig. 5). Feature set 3 (Table 1) is selected as it produces the least error in the learning curve. The reason we choose RMSE and R squared value as performance indicator is that to evaluate regression models we want to know 395 how much the prediction differs from the real value. We use R squared value to evaluate the variance of the data set represented by the models which will follow in model comparison. One of the advantages of using PLSR is that wearing can be found by plotting the correlation plot. As shown in one of the representative correlation plot, maximum tension, minimum

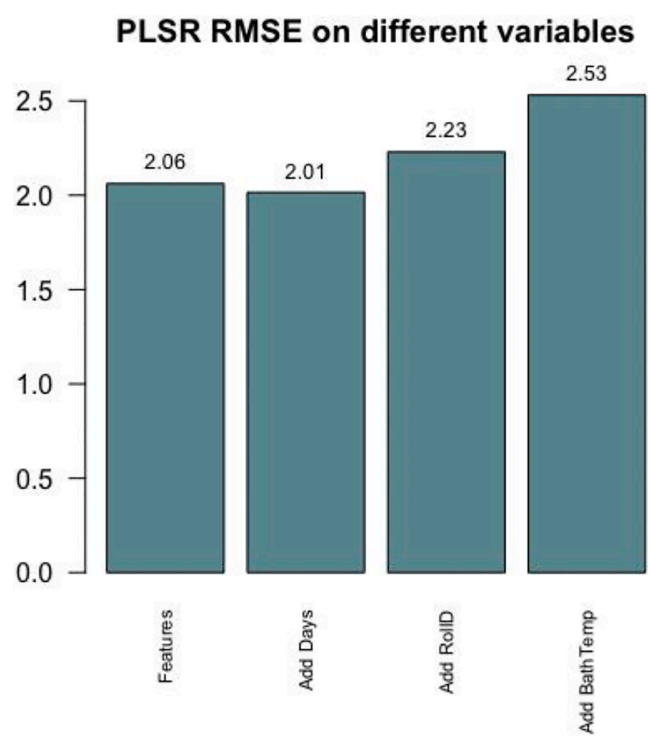

Fig. 4. Performance on initial 6 samples PLSR.

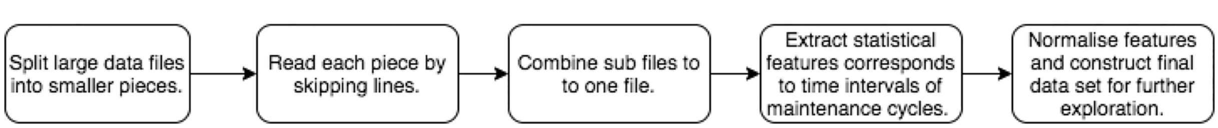

Fig. 3. Data processing.

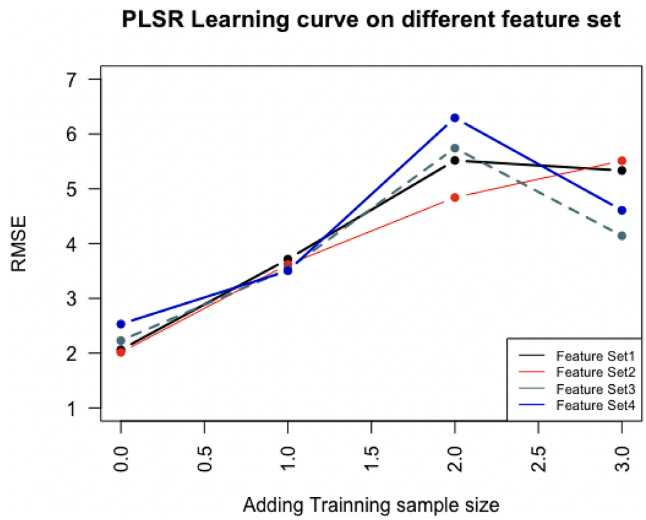

Fig. 5. Learning curve PLSR with different feature set.

tension and sink roll diameter are the most influential factors for prediction (Fig. 6).

6.1. *Artificial neural network*

In order to maintain consistency in model comparison, models are trained on the features selected. An Artifical Neural Network (ANN) has parameters to tune, namely the number of layers, number of nodes on each layer, and the initial weight of each neuron. The Default training mechanism, when using R, is that the model will stop training when the mean squared error is less than 350 0.01. Initial training weights are assigned randomly. This causes problems as the prediction result may be different every time we run the model. In order to produce consistent results, whenever a good prediction result is produced, we will record the weight matrix and set it as initial weights and then train the model using different training samples and again validate it using the test 355 sample. A neural network contains an input layer on which the number of nodes is equal to the dimension of the data set. In our case the number of input layer is 13. The output layer is set with one output node of 1. A

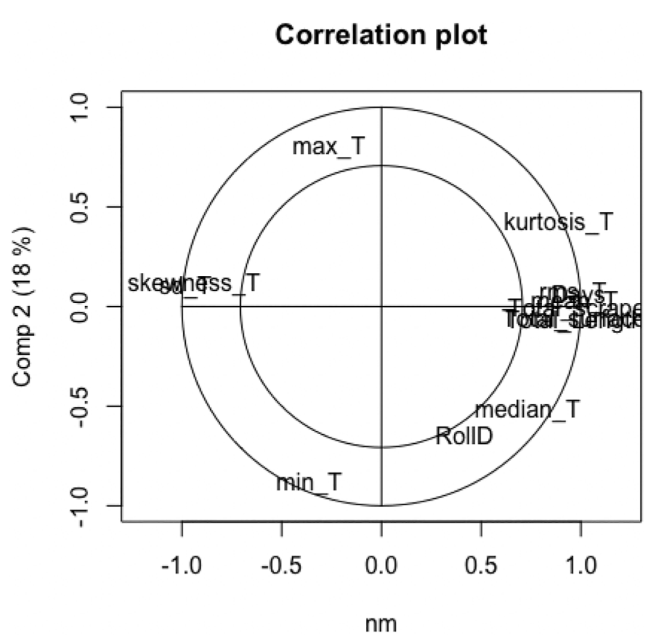

Fig. 6. Correlation plot PLSR.

decision has to be made for the number of nodes on the hidden layer. It is suggested by Heaton (2017) that the optimal size of the hidden layer is usually between the size of the 360 input and size of the output layers. It is also suggested by Heaton (2017) that the situations in which performance improves by adding a second (or third, etc.) hidden layer are very few. One hidden layer is sufficient for the large majority of problems. The number of nodes is decided by using one cycle as test set and adding up neurons from 1 to 13, and choosing the number when the prediction result is fluctuating around the label of the test sample instead of other random values. The structure of the ANN is tuned using Cycle 6 as test sample where the real label is 17 and the prediction result is 16.999712. We do not change the number of layers and number of nodes on each layer anymore for further experiments. The Neural network structure we use is 12, meaning disregarding input layer, output layer and bias nodes (Blue nodes in Fig. 7) there is 1 hidden layer with 12 nodes. After setting the NN structure, we use Leave One Out (LOO) cross validation again to select the initial weights by using one cycle as test set and training the model on the rest of the samples. The reason why we only use one sample as test set is that we want to maximize the number of training samples. A similar approach has also been used in literature (Li, Meng, Cai, Yoshino, & Mochida, 2009). The model is evaluated based on Root mean squared error (RMSE) value which is the measure of prediction deviation from the real value. A graph with a comparison of RMSE value when using different cycle as test sample can be found in Fig. 8. We can see from the graph that ANNs produces a stable result on the initial 6 samples no matter which sample was used as the test sample. All RMSE values are within 1 mm. The prediction result is always close to the real value. However, using cycle4 as test set does produce the best result.

In order to show more intuitively how close the neural network prediction result 385 is to the real value, we plot Fig. 9. A learning curve has been plotted to monitor the change of the prediction accuracy when more samples are added into training samples (Fig. 10).

6.2. *Random forest*

Based on (Biau & Scornet, 2016), Random Forest should also be able to handle small 390 sample sizes. Thus, we have tried out the random forest algorithm. The default number of trees in R is 500. Different from ANN, we do not have to tune the parameter in Random Forest as it is a random process. Samples chosen for each tree are different and the features chosen at each split are chosen randomly. We thus expect the model to produce a different result even with the same samples. This is considered one of the disadvantages of using Random Forest. The model is not able to predict special cases (Cycle 4 and Cycle 2) and the RMSE is 7.16 mm which is rather large for wear monitoring as the total diameter of the bush is only 30 mm. We also plot the learning curve to see whether the prediction gets more accurate when more samples are taken into consideration. We can see that the error slightly dropped but still does not fall into an acceptable range (Fig. 11).

6.3. *Comparison*

We compare the three models (partial least squared regression, artificial neural network and random forest) by adding one sample to the training set at a time. Starting with 5 training samples and 4 test samples and then 6 training samples, 3 test samples and so on. Then we take the RMSE (Root Mean Squared Error) and R squared value of the test samples. R squared value is the measure indicating the amount of variance the model can represent. RMSE indicates the prediction accuracy. The smaller the RMSE is the better the model performs whereas the larger the R squared value is the better the model performs. Physical or mathematical degradation models can also be explored providing sufficient knowledge on the wearing behaviour (Si et al., 2012). Our paper focuses on data-driven methods, other modelling types are not taken into consideration. The comparison table can be found in Table 3 and Table 2. By plotting the learning curve regarding each measure (Fig. 12

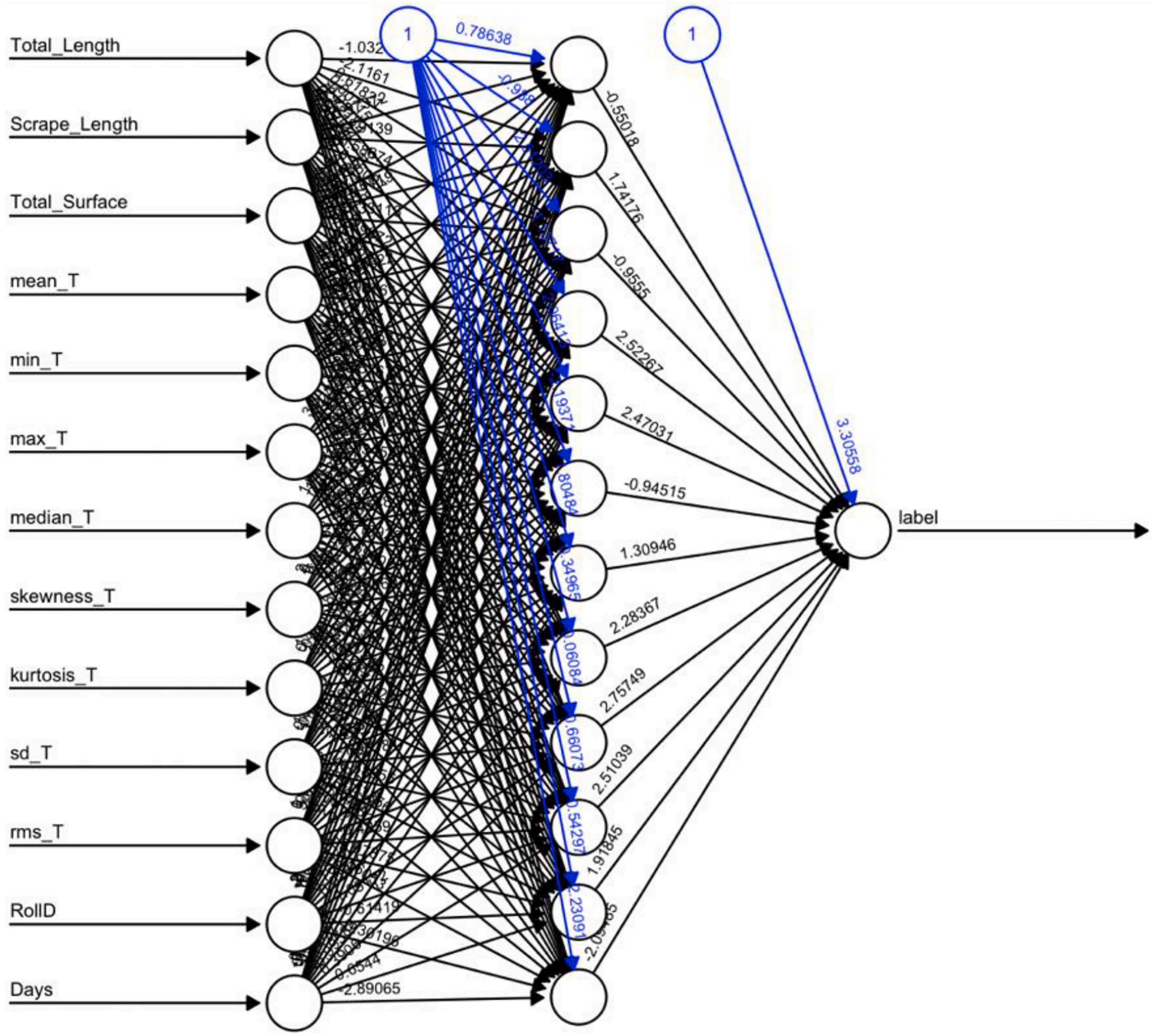

Fig. 7. Neural Network (12).

and Fig. 13) we can see that PLSR is holding the largest R squared value and the smallest RMSE in general. Thus, it is the best suitable model for current available data.

7. Discussion

The entire modelling process has been done 3 times because more data are 420 collected over time. It has been discovered that the way data is processed strongly effects the feature selection part of the modelling. The ANN model seems to suffer from over-fitting problems. ANN resulted in near perfect prediction on initial samples and drastically dropped in prediction power when new samples are taken into account. This problem may be solved by continuously adding new samples and then retraining the model, when the model has learned enough about variance from the samples, the result should be reasonable and stable. The ANN structure is also likely to require modifications when more samples are coming in. All three chosen modelling techniques have their own advantages and disadvantages. Within the current context of Tata steel and given the currently available data, PLSR is the most suitable model to implement for now. The stable pre- diction result produced by the PLSR model in turn confirms Archard's Law that the predictor variables and the target variable form a linear relation based on the currently available data samples. The strength of PLSR model is its explanatory power and its ability to allow model builders to see the contribution of each variable to the prediction result. This is significant when it comes to decision support. PLSR also produces consistent and stable results which are easier to implement and maintain as new data comes in. As the data size grows bigger, the modelling performance is subjective to change but it is likely to converge at some point.

Application of an Artificial neural network has the potential to produce prediction result with very high accuracy on training samples but the prediction performance is unstable with the currently available samples. Meanwhile, with a small sample size the ANN has to be tuned many times in order to fit the data. This needs manual tuning and requires a huge amount of iterations for updating the weights on each neuron. However, this problem should gradually disappear when more data becomes available. Another disadvantage of ANN is that it is a "black box" model thus it is hard to interpret the results as in an ANN it is very complex to obtain insight in which variables effect the wearing condition the most. The Random forest model is so far not able to either fit nor predict with current samples. The model combined too many random pro- cesses thus should require huge data sets before it can produce decent results. However, Random Forest is the least computationally expensive among all three modelling techniques as it does not need parameters tuning or component selection. Random forest seems a technique that worth re-trying when the sample size becomes sufficiently large and it can potentially be the most user friendly model for online monitoring because it is the least computational expensive.

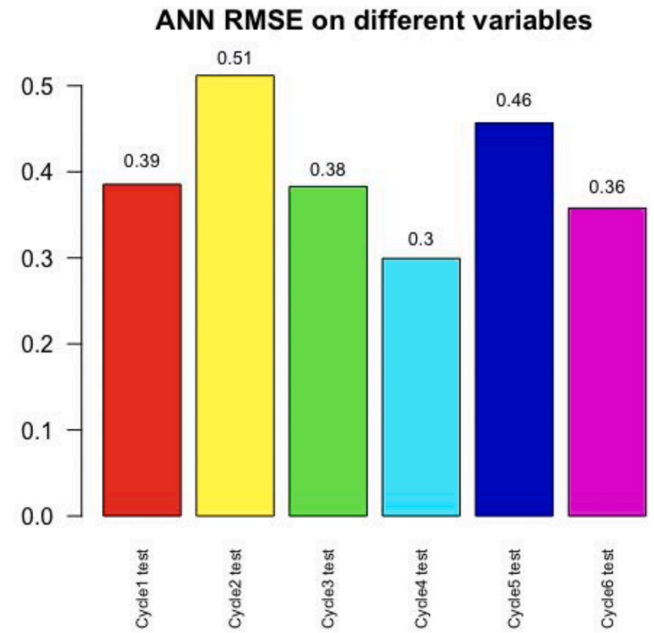

Fig. 8. ANN performance and selection.

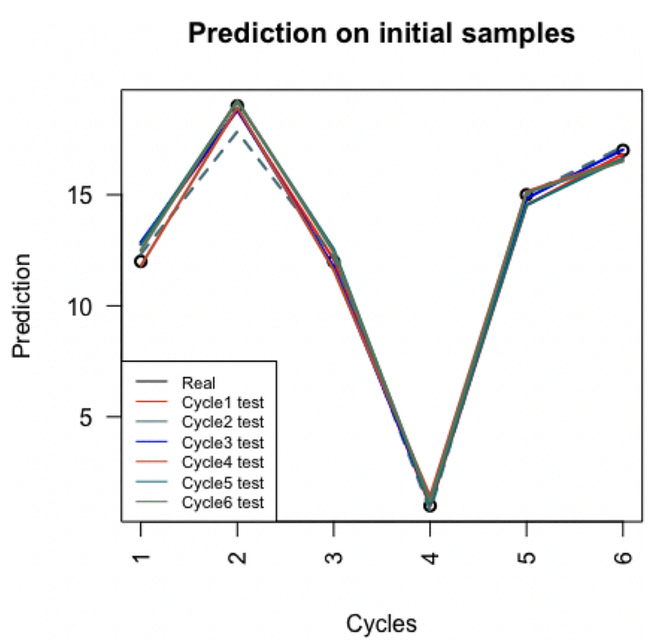

Fig. 9. RMSE comparison using different test samples to find initial weight.

A small sample size brings challenges from the very beginning of data exploration when we did correlation test between variables and bush wear. Because the sample size is too small, we cannot really draw any conclusions on relation ships between tension and wear although they seem to have a strong correlation based on the samples we have. The same holds for the linear regression model. We can not be sure that the variables and bush wear follow a linear relation based on only six cycle samples.

The linearity might be gone when more data becomes available. However, the good prediction results from the PLSR model are to some extent validating the correlation and linear relationships between tension, run length and wear measures because the underlying theory behind PLSR is linear regression.

The modelling process from feature selection to model selection

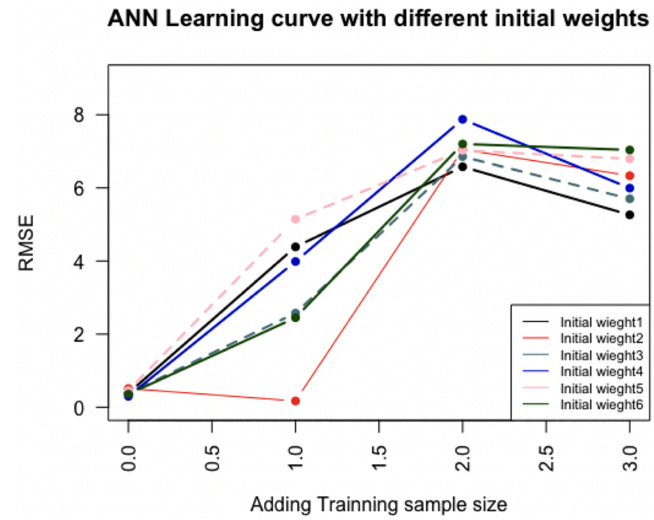

Fig. 10. Learning Curve ANN.

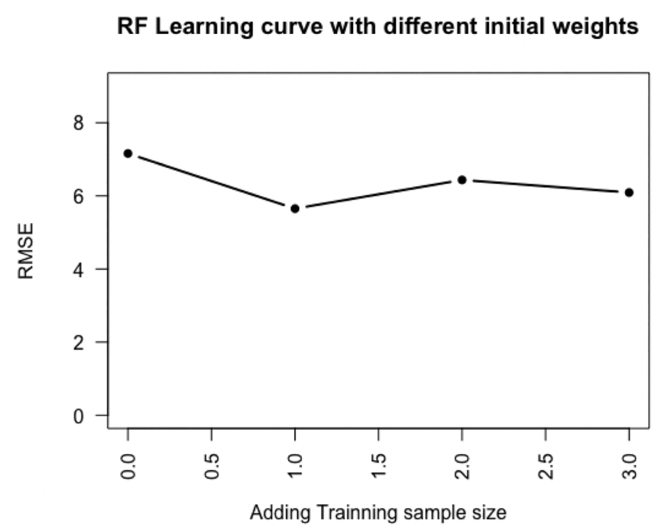

Fig. 11. Learning curve random forest.

Table 2
Modelling RMSE comparison.

Number of Training samples	PLSR	ANN	RF
5	2.23	0.39	7.16
6	3.53	4.39	5.65
7	5.74	6.58	6.43
8	4.14	5.26	6.09
Average	3.91	4.15	6.33

Table 3
R^2 comparison of models.

Number of Training samples	PLSR	ANN	RF
5	0.85	1	−0.53
6	0.61	0.4	0.01
7	0.18	−0.07	−0.02
8	0.53	0.25	−0.01
Average	0.55	0.39	−0.14

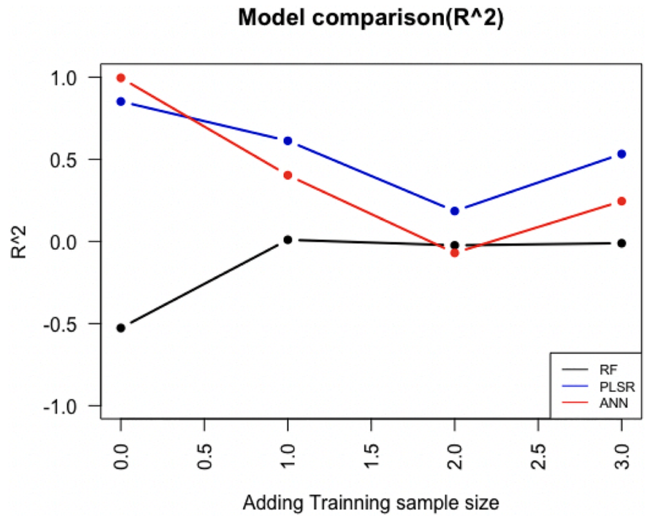

Fig. 12. Learning curve R squared value.

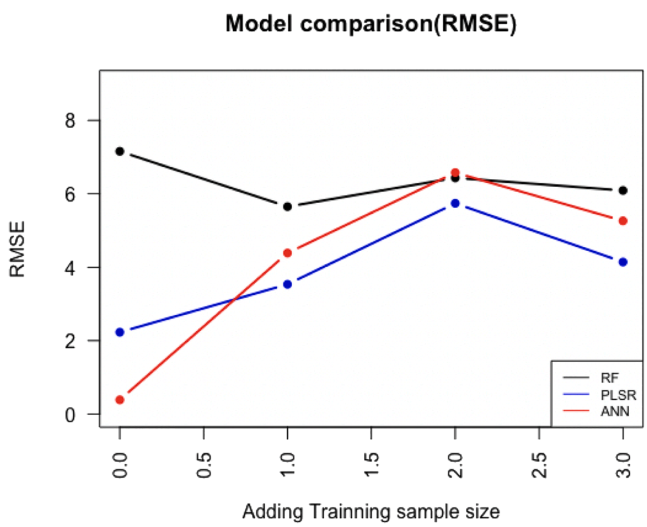

Fig. 13. Learning curve RMSE.

should be repeated over pre-set time intervals to discover more about the stability and ac- curacy in longer term. At the beginning when model performance is not stable the modelling process should be repeated more frequently, e.g. once per month. When the modelling performance converges, the maintenance interval can be longer, say every 3 months or every half year depending on the specific situation. In the worst case scenario where the modelling performance is dropping, more investigation into other techniques should be done in order to maintain the prediction accuracy.

The methods and modelling techniques can be well generalized into application areas that have similar characteristics. These are businesses that are transforming into digital era where the connectivity of the data infrastructure is limited, in which data are not integrated into one platform, instead, stored in different data sources. Data are in large volumes and in high number of features and mostly unstructured. These are areas where wear is not directly observable and failure data is limited. Possible applications having similar characteristics areas are machine tools in discrete manufacturing (Akcay, Topan, & van Houtum, 2020), wheels in rolling stocks (Li, Ren, Enblom, Stichel, & Li, 2020). These systems have data in large volumes with a high number of features (dimension) that explain wear. Yet, the wear is hid- den and precise measuring of wear is only possible during periodic inspections, and therefore, predicting wear is essential between inspections. Our data-driven prediction model can directly be applied to these sorts of systems. We believe that PLSR can be applied to the situation where data are in high dimension meaning that the data includes many features and sample sizes are small.

In our paper, we are predicting the non-observable degradation/wear of bush for monitoring instead of predicting the degradation/wear over a future time interval. Hence, our focus is monitoring rather than predicting the future. By inserting the parameters into Fig. 15, we are able to predict current remaining bush width instead of the remaining bush width in the future. Therefore, prediction horizon is not the focus of this paper. Regarding the frequency of data updates, it is fixed and outside the scope of this project. The result shows in the paper is a proof of concept that predicting/monitoring the bush condition by data-driven methods is feasible. By adding more data the monitoring performance can be increasing or decreasing. The recommended data update frequency is 4 weeks as that's a normal cycle length when the rolls are pulled out and bush condition can be inspected.

8. Implementation

To implement predictive modelling into the Tata steel operation, a web front- end has been developed for monitoring purposes. By having the web front- end, engineers will obtain insights into the current wear and predicted wearing pattern of the bush. The web interface contains two main parts; the first part is the input, including date, sink roll diameter and bush diameter (Fig. 14). As for each maintenance cycle, the sink roll diameter and bush diameter are subject to change, it is not possible to have these automated in the modelling process. Date is an essential input. When users select a date, the algorithm will automatically search for the start day of the maintenance cycle that contains the user selected date and predict the remaining bush width of each day within that cycle. The prediction result of the selected date is presented at the bottom of the graph, both the wear width and the remaining width (bush width - predicted wear width). Combining the information of the predicted wearing pattern and the predicted wear (Fig. 15), decision makers should be able to decide whether to replace the part or not and to determine when the maintenance activity should be planned. By monitoring the remaining bush width, we can avoid to over-maintain the rolls. When the predicted remaining bush width is big, then the engineers can decide not to replace it even though the rolls are in the zinc pot for 4 weeks. We can also combine replacing the rolls in combination with the product change. As when there is a product change, we need to pull the rolls out of the zinc pot and replace the pot with another chemical pot. If in that case the predicted bush width is very thin, we can replace it while changing the pot such that downtime is reduced by lowering the shutdown frequency and create savings.

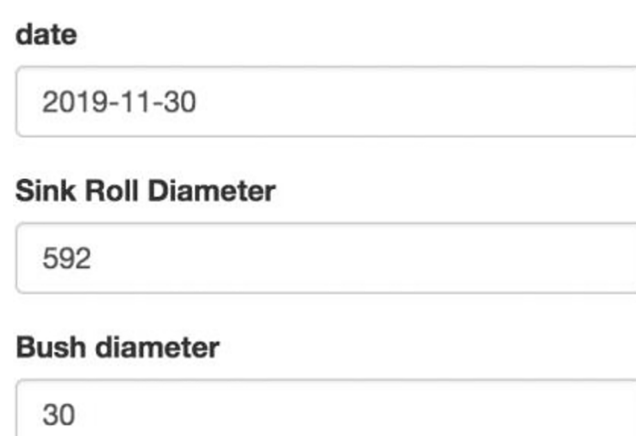

Fig. 14. Web page input.

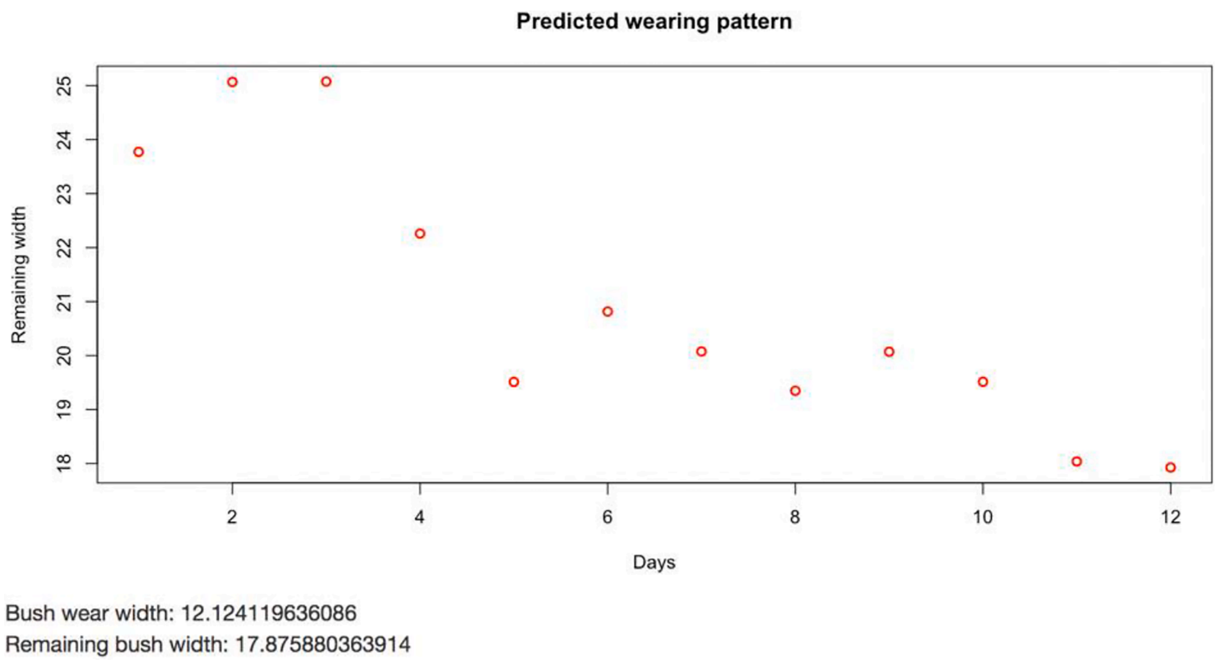

Fig. 15. Web page output.

9. Conclusions

Our paper shows that it is feasible to predict the component condition using existing data related to the component when vibration data is not available. Partial least squared regression (PLSR) predicts the bush wear with a good accuracy and produces stable results. Thus, PLSR is recommended for the currently available data in this project at Tata steel and this experience gives support for PLSR to be considered for predictive maintenance in similar real world industry settings. The value of this research project mainly lies in the ability to monitor current wear and thus provide insight on the wearing behaviour for maintenance decision support and for discovery of differences between the theoretical and practical context. The biggest limitation of this project is the sample size. By the end of this project the sample size has been increased from the initial 6 to 9 samples. According to the plotted learning curve, the prediction error first rises and then drops in the end based on 9 samples. We can not know whether the error will continue to drop or it may increase again when more samples are taken into account. Thus it is necessary for the model to be retrained and re-evaluated with new data.

Our paper contributes by predicting the real time bearing wear, which is important in steel production. We think that our findings contribute to predictive maintenance planning in steel production. Yet, there are several possible directions to extend the research in the future: (1) Predict the future bush wear. (2) Make use of unsupervised learning and correlation tests to detect abnormal events in the production line. (3) Expand the model to other sites/lines. (4) Follow a similar methodology to build models for other problem areas within the production line. (5) Investigate various data processing platforms that are emerging to reduce the time needed for data reading and processing. Such steps are likely to be similar in comparable industry settings where increasing amount of data becomes available from production lines. However, these data may have diverse volume, quality and consistency, and the feasibility of applying the method presented here for predicting maintenance would need careful examination of its suitability and reliability compared to alternative data-driven modelling methods.

CRediT authorship contribution statement

X. Chen: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Visualization, Project administration. **J. Van Hillegersberg:** Writing - review & editing, Supervision, Resources, Methodology. **E. Topan:** Writing - review & editing, Supervision, Resources. **S. Smith:** Resources, Data curation, Project administration, Funding acquisition. **M. Roberts:** Data curation, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by [Tata Steel Europe] and [University of Twente]. We thank our colleagues from Tata Steel who provided insight and expertise which greatly assisted the research. We thank [Tom Eussen] for providing the opportunity for this research. Fellow colleagues [Mark Wardle] for IT support. Peter Edwards for domain knowledge on systems and components. Glyn Derby for assistance on data exploration.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2021.115699.

References

Akcay, A., Topan, E., & van Houtum, G. J. (2020). Machine tools with hidden defects: Optimal usage for maximum lifetime value. *IISE Transactions*, 1–14.
Amruthnath, N., & Gupta, T. (2018a). A research study on unsupervised ma- chine learning algorithms for early fault detection in predictive maintenance. In *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)* (pp. 355–361). IEEE.
Amruthnath, N., & Gupta, T. (2018b). Fault class prediction in unsupervised learning using model-based clustering approach. In *2018 International Confer- ence on Information and Computer Technologies (ICICT)* (pp. 5–12). IEEE.
Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test, 25*(2), 197–227.
Brown, M. S. (2015). What IT Needs To Know About The Data Mining Process. Published by Forbes, 29.

Cipollini, F. et. al., (2018). Data-Enabled Discovery and Applications, Num:1 - Unsupervised Deep Learning for Induction Motors Bearings Monitoring, Vol:3 - 2018.

Cubillo, A., Perinpanayagam, S., & Esperon-Miguez, M. (2016). A review of 590 physics-based models in prognostics: Application to gears and bearings of rotating machinery. Advances in Mechanical Engineering, 8(8), 1687814016664660.

Ding, H., Yang, L., Cheng, Z., & Yang, Z. (2021). A remaining useful life pre- diction method for bearing based on deep neural networks. *Measurement, 172*, 108878.

Gouriveau, R., Ramasso, E., Zerhouni, N. (2013). Strategies to face imbalanced and unlabelled data in PHM applications. Chemical Engineering Transactions, 33, 115-120.

Heaton, J. The number of hidden layers, 2017. URL https://www. Heatonre-search. com/2017/06/01/hidden-layers. html.

Langone, R., Alzate, C., De Ketelaere, B., Vlasselaer, J., Meert, W., & Suykens, J. A. (2015). LS-SVM based spectral clustering and regression for predicting maintenance of industrial machines. *Engineering Applications of Artificial Intelligence, 37*, 268–278.

Li, Qiong, Meng, Qinglin, Cai, Jiejin, Yoshino, Hiroshi, & Mochida, Akashi (2009). Applying support vector machine to predict hourly cooling load in the building. *Applied Energy, 86*(10), 2249–2256.

Li, Yuyi, Ren, Zunsong, Enblom, Roger, Stichel, Sebastian, & Li, Guodong (2020). Wheel wear prediction on a high-speed train in China. *Vehicle System Dynamics, 58*(12), 1839–1858.

Liao, Z., Gao, D., Lu, Y., & Lv, Z. (2016). Multi-scale hybrid HMM for tool wear condition monitoring. The International Journal of Advanced Manufacturing 615 Technology, 84(9-12), 2437-2448.

Panagiotis, Predictive Maintenance in Aviation: Failure Prediction from Post Flight Reports 2018.

Sniderman, B., Mahto, M., & Cotteleer, M. J. (2016). Industry 4.0 and manufacturing ecosystems: Exploring the world of connected enterprises. Deloitte 620 Consulting.

Tobon-Mejia, D. A., Medjaher, K., Zerhouni, N., & Tripot, G. (2012). A data- driven failure prognostics method based on mixture of Gaussians hidden Markov 625 models. *IEEE Transactions on Reliability, 61*(2), 491–503.

Yang, Bin, Lei, Yaguo, Jia, Feng, & Xing, Saibo (2019). An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mechanical Systems and Signal Processing, 122*, 692–706.

Yang, J., Yin, S., Chang, Y., Gao, T., & Yang, C. (2020). An Efficient Method for 630 Monitoring Degradation and Predicting the Remaining Useful Life of Mechan- ical Rotating Components. *IEEE Transactions on Instrumentation and Measurement, 70*, 1–14.

Zhang, S., Zhang, S., Wang, B., & Habetler, T. G. (2019). Deep learning algorithms for bearing fault diagnostics-a review. In *2019 IEEE 12th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)* (pp. 257–263). IEEE.

Zschech, Patrick, Heinrich, Kai, Bink, Raphael, & Neufeld, Janis S. (2019). Prognostic model development with missing labels. *Business & Information Systems Engineering, 61*(3), 327–343.

Van Hillegersberg received his MSc in computer science at the Leiden University in 1991. In 1997 he received his PhD at the Erasmus Universiteit Rotterdam under Kuldeep Kumar with the thesis, entitled "Metamodelling-based integration of object-oriented systems development." In 1990 Van Hillegersberg had started his career as research assistant at the Expert Systems Group of IBM, and in 1998 he became Component Manager at AEGON. In 2000 he returned to the academic world to become in Associate Professor at the Erasmus University Rotterdam. In 2005 at the University of Twente he was appointed Professor of Business Information Systems at its School of Business, Public Administration & Technology, where he also became Head of the Department of Information Systems and Change Management. Van Hillegersberg is also working as consultant for the Wagner Group.