KMD Seminar on: Neural Networks for Event Prediction in the Domain of Machine Maintenance.

Mehedi Hasan

February 1, 2018

1 Literature Search Procedure

Interface for searching: scholar.google.com
Keywords: Event Prediction, Neural Networks, Machine Maintenance, Predictive Maintenance

Selection Procedure:

1. In first search using above mentioned list of keywords: 32,200 results
2. Narrowed down the search by applying filter-year range from 2010-2017: 17,210 results
3. Tried to find relevant papers for my problem of interest based on title, Google scholar search result summary and then abstract, also introduction sometimes. For top 20 search results: found 18 papers in machine maintenance domain but irrelevant with my topic (papers are on fault detection based on other methods and also not based on neural networks) and 2 papers are partially relevant (Neural Networks are used to solve problem but not in event prediction).
4. Further narrowed down the search by limiting it to year since 2013: 17,200 results which is almost same as previous result.
5. If I again reduce the time range from 2014 to 2017: 16,800 results.
6. By repeating the step-3 i found 4 papers: From first 20 search results [Martínez-Martínez et al., 2015] has evidence that Neural Networks is used for event data, in next 20 search results no interesting paper detected, considered further 20 results where I found [Wu et al., 2014], [Arabgol et al.,] and from next 20 [Ramos et al.,].
7. Again I performed the search for year 2016 and 2017 using same list of keywords: 8,500 results where i selected [Xiao et al., 2017], [Chen X, 2017] from top 100 results.
8. Further narrow downing search for year of 2017 gave me: 3,710 results.
9. Then i looked for relevant papers iterating top 500 results in random fashion (forward, backward and also accessing random index) based on paper title, Google search result summary, abstract and sometimes introduction part: [Guo et al., 2017], [Yang et al., 2017], [Nikula et al.,], [Zhang et al.,], [ElSaid et al., 2017].
10. In this point i refined the search again by rearranging and and reducing one keyword exactly like "Neural Networks Event Prediction Machine Maintenance" and applying filter since 2017 which gave me: 5,900 results. By analyzing through the top 200 results i found [Zhao et al., 2017], [Tao et al., 2017].

For your information, search result was sorted by relevancy which is a default. I have got different number of results using same set of keywords and same sequence as previous search in different time but difference was not more than 200 results. Anyway i was able to reproduce the papers i selected even though i was getting more results than previous.

2 Clustering of Selected Literatures

Selected literatures are grouped based on actual problem each paper solved and similarities among the papers.

Cluster 1 –> Small Fault Event Detection for Machine Maintenance:
[ElSaid et al., 2017] predicted excess vibration values for aircraft engine from flight data. According to author, excess vibration is not an normal event. Although it is a small fault for an aircraft engine, this might lead to future unpleasant situations and engine maintenance activities. Author said: "These neural networks provide a promising means for the future development of warning systems so that suitable actions can be taken before the occurrence of excess vibration to avoid unfavorable situations during flight". Based on author's opinion, i would say, "suitable actions" could be in-flight maintenance or maintenance actions after emergency landing. [Arabgol et al.,] predicted energy output to detect abnormality in wind turbine with high accuracy. After predicting energy output, Exponentially Weighted Moving Average (EWMA) chart was applied to control the residuals between predicted output and expected output under certain conditions (e.g. in windy weather condition). Based on the difference between predicted and expected output value, the EWMA control chart might detect that there is something abnormal. The small discrepancy between predicted and regular output was considered as small fault which might lead to future catastrophic failure events. This is how failure could be detected in advance and wind turbine damages could be avoided. [Nikula et al.,] predicted substantial mechanical stress inflicted on the machines during the processing of modern high-strength steels. Predicting mechanical stress of a machine is important for detecting if the capability of a machine is exceeded. The incapable machines results damage and flaw in final product. So, in other words, author predicted damage and flaw events in the manufactured product (e.g. steel). And author said: "The continuous development of steel products generates new challenges for the maintenance of manufacturing machines in steel mills". So, this study was targeted for machine maintenance domain.

Cluster 2 –> Remaining Useful Life Prediction for Machine Maintenance:
According to [Si et al., 2011]: "The remaining useful life (RUL) of an asset or system is defined as the length from the current time to the end of the useful life." And [Sankararaman,] said: "The remaining useful life of an engineering component/system is defined as the first future time-instant in which a set of safety threshold conditions are violated." Safety conditions of a machine also known as Health-Indicator(HI). Analyzing HI for for RUL prediction is also known as prognostics [Tao et al., 2017]. According to [Guo et al., 2017] and others, RUL prediction problem is actually failure time prediction problem. As definition suggests, the failure time is a time-instant in future when a machine face failure event. Generally, at failure time, HI exceeds certain threshold. So, actually, [Guo et al., 2017], [Wu et al., 2014], and [Tao et al., 2017] predicted failure time of machines or components of the machine.

Cluster 3 –> Anomaly Detection for Machine Maintenance:
Anomaly detection is the identification of events or observations which do not conform to an expected pattern or other items in a dataset [Deepthi and Rao, 2014]. [Ramos et al.,] predicted disc replacement event (because of malfunctions) by applying Artificial Neural Network (ANN) over data obtained from a monitoring system that continuously keeps track of the relevant equipment parameters of manufacturing machines. [Zhao et al., 2017] predicted the actual tool wear event using ANN based on raw sensory data collected from manufacturing machines. Tool wear is the gradual failure of cutting tools due to regular operation i.e. flank wear (the portion of the tool in contact with the finished part erodes) [Kadirgama et al., 2011]. And [Zhang et al.,] proposed a ANN based (rare) event (e.g. hard disk failure prediction, heating system fault prediction etc) prediction method based on hard drive datasets collected from data center environment. Consequently, this group literature solved anomaly detection problem for machine maintenance in general. However, authors in cluster 1 predicted small faults which are common event and not actual faults but could be reason of future catastrophic failure events if they remain undetected.

Small Fault Event Detection for Machine Maintenance:

[Arabgol et al.,] According to author: "Generally, there are two types of problem that happens on wind turbines: 1) sudden and unexpected failures which are usually hard to predict, such as a turbine tower breakage after an environmental catastrophe; and 2) small problems that are difficult to detect, e.g. a small crack on the blade, and a light wear on gearbox that could get larger and ultimately cause severe damages". Small and slowly developing fault events may lead to future catastrophic events in wind turbine if those are remain undetected and not fixed on time. Author addressed this problem and proposed an intelligent Multi-layer Perceptron(MLP) Feed-Forward Neural Network based fault detection system which predicts the output of wind turbine based on environmental parameters. This is a two step approach where first step is applying ANN to predict energy output from a wind turbine and, second step is comparing the predicted energy output with expected value continuously over time (If the residual between predicted and actual output is more than some threshold, the wind turbine is in fault). This is how small faults could be detected, preventing the wind turbine from potential damage events which causes huge downtimes and costs.

As i mentioned earlier, author used Multi-layer Perceptron(MLP) Feed-Forward Neural Network (FFNN) to solve this problem. "Feed-forward networks are a class of ANN in which the information will be communicated only downstream in the network". It has a input layer, one or two hidden layer and a output layer was trained on a dataset consists of three features (Wind Speed, Number of rotation per unit of time and, Voltage). The network was trained based on a back-propagation algorithm. It yields the predicted amount of energy. The energy output of a wind turbine depends on mechanical impacts of the machine and also on environmental and electrical conditions. Equation (1) shows the energy output variation.

$$"\epsilon_{Output} = \epsilon_{Environmental/Electrical} + \epsilon_{Mechanical}" \tag{1}$$

After getting predicted the energy output, the value was compared with actual output obtained from the wind turbine Exponentially-Weighted Moving Average (EWMA) chart. The difference between real and predicted energy output, which was monitored to detect small changes in residuals, are the indicators of small faults.

The entire dataset was divided into training (for training the ANN), validation (validating the developed network), and test datasets (evaluating based on prediction accuracy). The criteria for performance evaluation were Root Mean Square Error(RMSE) which gives the average prediction error and the coefficient of determination(R^2) which gives the goodness of fit. Author said based on the results obtained, "the performance of the ANN in predicting energy output using the three mentioned input variables is significant". Author also mentioned, the predictions generated by the ANN were reasonably accurate. Consequently, the proposed method can predict the failure in advance and can be considered as baseline for intelligent early failure event detection.

[Nikula et al.,] The research problem discussed in this literature is the prediction of the relative stress that is inflicted during the leveling of steel strips to prevent from damage events. Regression-Based Artificial Neural Network(ANN) technique was used to predict relative stress level. According to author, "the relative stress defines the relative level of the mechanical stress that each leveling event inflicts on the machine". For this purpose, the Generalized Regression Neural Network(GRNN) was used and compared with Multiple Linear Regression(MLR) and Partial Least Squares(PLS) based model in prediction of the stress level.The generalized regression neural network is a memory-based network was used here to solve this problem. It has a one-pass learning algorithm with parallel structure. It approximates any arbitrary function between input and output vectors and draws the function estimate directly from the training data. The proposed, GRNN consists of four layers, which include the input layer, pattern layer, summation layer, and output layer.

The steel strip features (Yield strength, length, weight, width and thickness, respectively) were used as explanatory variable. Also, the relative stress features were extracted as response variable from acceleration signal. The response variable is the focus of a question in a study and explanatory variable is one that explains changes in that variable. The criteria for performance evaluation were Root Mean Square Error(RMSE) and the coefficient of determination(R^2). The results showed that "GRNN had better performance compared with linear regression models in model training". This means that the neural network learned the training data effectively. However, the performance in model testing was almost the same with linear models, but the prediction accuracy for the test data was similar to the other models. The GRNN had the best performance based on the RMSE, but the R-value of testing was almost the same in comparison with other models.

[ElSaid et al., 2017] This literature provided an Artificial Neural Network(ANN) based generalizable and robust method to predict aircraft engine vibration events from different flight parameters. This method used Long Short Term Memory(LSTM) Recurrent Neural Networks(RNN) variations of ANN for prediction task. The LSTM network was trained on a large dataset of flight data records collected from an airline containing flights that suffered from excessive vibration. The entire flight dataset was divided into two distinct sets : the training set (contains data for 28 flights) and the testing set (57 flights). Total 15 features were chosen based on the degree of their contribution to the vibration based on aerodynamics/turbo-machinery background for training. Three different architectures (Architecture I, Architecture II, Architecture III are capable of predicting vibration values in 5, 10 and 20 seconds in advance) of LSTM RNNs based method were trained in Forward Propagation. Mean Squared Error(MSE) was used as an error measure to train the three architectures because it provides a smoother optimization surface for back-propagation of error. The error is calculated at the output and distributed back through the network layers. The performance of these three different architectures were compared based on error in prediction. Mean Absolute Error(MAE) was used as a final measure of accuracy for the three architectures. Author said: "As the parameters were normalized between 0 and 1, the MAE is also the percentage error". Architecture I provides the best results regarding the overall accuracy of the vibration prediction where Architecture II yields the least accurate prediction. Higher peaks of the vibration signal predicted by Architecture I was more accurate than the lower peaks prediction as if the neural network is tending to learn more about the max critical vibration value, anyway this is favored by this project because this project was aimed at predicting excess vibration event, where Architecture II is good at predicting lower peaks but weak at higher peaks of the vibration signal. Although Architecture III was the most computationally expensive and deeper network, the results were not as accurate as expected compared to most simple Architecture I.

According to author, one serious problem of traditional deep RNN is that it suffers from vanishing/exploding gradients when trained with back propagation. It means, when RNN is trained by back propagation through time, and therefore unfolded into feed forward net with multiple layers, gradient is passed back through many time steps, it tends to grow or vanish, same way as it happens in deep feed-forward nets. LSTM RNNs provide a solution to that by providing a "memory" of the contribution of previous time series data which can further improve predictions of future vibration values.

Comparisons: To prevent the machines or parts of the machine from its deteriorating working condition, small faults i.e. excess vibration [ElSaid et al., 2017], excessive stress [Nikula et al.,], small crack on the blade, and a light wear on gearbox [Arabgol et al.,] etc. must be detected at least in some certain time advance. This is how future catastrophic damages could be avoided which causes not only huge downtime but also huge amounts of recovery costs. This grouping of literatures dealt with such similar problems (small fault detection) and proposed mainly Artificial Neural Network(ANN) based techniques to address small fault event detection problem to prevent catastrophic event. And also based on cluster 1 introduction section, the literatures included in this group are similar. But they are different according to their target machine and target variable i.e. [Arabgol et al.,] detected small and unnoticeable faults by predicting energy output using an artificial neural network (ANN) for wind turbine setup where [Nikula et al.,] predicted relative stress level inflicted on a roller leveler using measurements from an steel mill industry case. [ElSaid et al., 2017] predicted vibration values of aircraft engine in certain time (5, 10, and 20 seconds) advance from flight data records.

These literatures could be compared based on the criteria: Predictive-quality, Simplicity and Stability of the prediction algorithm. Predictive-quality is the most important criteria which defines how accurately the proposed method can predict [Arabgol et al.,]. In other words, how close is the predicted value and actual value. Simplicity defines how simple the model is. It is also known as Network-complexity. Typical criteria are: a small number of input features, small number of neurons and layer [Romero Ugalde et al., 2014]. [Xu et al., 2012] said: "In a broad sense, stability means that an algorithm is well posed so that, given two very similar data sets, an algorithm's output varies little."

From predictive-quality point of view: [ElSaid et al., 2017] predicted excess vibration value with given the accuracy of the predictions rather far in the future – 3.3 percent error for 5 second predictions, 5.51 percent error for 10 second predictions, and 10.19 percent error for 20 second predictions. [Arabgol et al.,] predicted energy output with MSE of 51[W]. And according to author: "Coefficient of determination value (R) was above 99 percent for all processes which means that the input parameters were good estimators for the output with high accuracy". [Nikula et al.,] predicted stress level with given prediction error approximately .095. And Coefficient of determination value (R) was above 72 percent. Based on these results obtained, i would rank [Arabgol et al.,] as best paper, where [Nikula et al.,] comes next and finally [ElSaid et al., 2017] in prediction of small fault events.

From model simplicity point of view: [ElSaid et al., 2017] proposed three different architecture of LSTM RNNs (Architecture I, Architecture II, Architecture III). Among them Architecture I is the simplest one and produce best results. Architecture I has three layer with a input layer (15 input nodes, it takes 15 features as input), a hidden layer and a output layer. [Nikula et al.,] proposed a deep GRNN method which takes 90 features with three explanatory variables as input and the network has 4 layer. [Arabgol et al.,] 's MLP FFNN has total three layer which takes 3 features, is the simplest architecture. So, [Arabgol et al.,] is the best paper from model simplicity point of view, next [ElSaid et al., 2017] and at last comes [Nikula et al.,].

From stability point of view: [ElSaid et al., 2017] obtained 3.3 percent error for 5 second predictions in testing and .03 percent in training. [Nikula et al.,] obtained error in training 6.96 percent 8.31 percent in testing where [Arabgol et al.,] 's Coefficient of determination value (R) was always above 99 percent but MSE for training and testing was almost same. I would rank [Arabgol et al.,] as 1st , [Nikula et al.,] as 2nd and [ElSaid et al., 2017] as 3rd paper.

(2) Remaining Useful Life Prediction for Machine Maintenance:

[Guo et al., 2017] The bearing is a most frequent and important rotating component of industrial machines. According to author, "Prediction accuracy of bearing remaining useful life (RUL) mainly depends on the performance of bearing health indicators which are usually fused from some statistical features extracted from vibration signals". Available literatures regarding health indicators computation has the following two limitations: (1) different features/parameters contribute in different degree on machine health indicators. (2) difficult to fix a failure threshold value for health indicators. A Recurrent Neural Network based HI is presented to overcome these two shortcomings. Most informative and sensitive features were selected based on their degree of contribution to the health indicators and correlations among them (8 features). The RNN-HI was evaluated by two bearing data sets collected from experiments and wind turbine industry. The proposed approach was compared with a self-organizing map based Health Indicator (SOM-HI). The comparison criteria was percent of error in prediction. According to author, for experimental dataset, "the proposed method has the lower percent error than the SOM based method". And "the industrial data verification showed that the proposed method can predict the wind turbine generator bearing RUL effectively". RNN-HI performed better than the SOM-HI.

[Tao et al., 2017] Machine health prognostics is aimed at predicting the time at which a machine or components of a machine will no longer perform its intended function meaning machine will fail. Predicting failure time is actual challenge here. One of the most important factors of failure time prediction is previous failure history of the machine. However, the prognostics become more challenging if one has to consider truncated failure history of the machine component for accurate prediction of machine health. Truncated history means if there is no failure history available. Author addressed this problem by offering a Feed Forward Neural Network(FFNN) based method to predict machine component health of the individual component. To train the FFNN, the Neighborhood Geodesic Distance(NGD) data which is a geometrical metrics dataset with truncated failure history was used as inputs to the Neural Network(NN). For evaluation, proposed method was tested on degradation data generated by bearing test rig that can produce run-to-failure data. It was evaluated based on accuracy in predicting failure time. The results demonstrate that the proposed method can predict failure time even with only truncated histories. According to author, "the failure threshold predetermined in this study was 0.5 that means once the survival probability drops below 0.5, the failure time interval is reached".

[Wu et al., 2014] To avoid Sudden damage caused by temporary failure and ensure excellent operation of the equipment, author presented an effective Remaining Useful Life(RUL) prediction model which combined the back propagation neural network (BPNN) with multi-agent cooperation grouping algorithm. The three features (Bearing Temperature, Cylinder Pressure, frame displacement) were input to the BPNN. Author said that "The values of weights and thresholds of BPNN were obtained through optimization results of the multi-agent cooperation grouping algorithm". BPNN was repeatedly trained based on above mentioned initialization parameters to predict machine health indicator(HI). For validation, a case study on continuous casting equipment dataset was conducted and compared with performance of traditional BPNN techniques based on convergence error, average time of each iteration. According to author, the proposed model can predict HI to compute RUL with increased prediction accuracy, compared with classical BPNN prediction method. Author said, "The suggested method is superior to classic BPNN in convergence speed."

Comparisons: To compute RUL of a machine or components of a machine, modeling the relation between HI and Failure Time is most important. This group literatures discussed about this. All the authors in this cluster used ANN based approach to solve this problem as like cluster 1. To compare the literatures inside this cluster, criteria could be: Predictive quality, Reliability, and Robustness. [Javed, 2014] mentioned, "Ability of a prognostics approach to be insensitive to inherent variations of input data". It means that, whatever the subset from the entire training dataset is used, the performances of a robust prognostics model should not impair. [Javed, 2014] also said, "Even if the prognostics approach appear to be robust to tolerate uncertainty, it should also be reliable enough to be used for the context that is different from the one considered during the modeling phase". It means, the prognostics should cope with the variations of the context, such as, multiple operating conditions, geometric scale or materials differences of components, etc [Javed, 2014].

From predictive quality point of view: [Guo et al., 2017] predicted RUL with mean error 23.24 percent in prediction. So accuracy is 76.76 percent where [Tao et al., 2017] predicted RUL with mean accuracy of 98.32 percent. And [Wu et al., 2014] achieved 89.08 percent accuracy. Consequently, ranking would be first [Tao et al., 2017], then [Wu et al., 2014] and finally [Guo et al., 2017].

From robustness point of view: In [Tao et al., 2017] and [Guo et al., 2017], the uncertainty of from training data were handled explicitly. That means they are robust. [Wu et al., 2014] said, "The values of weights and thresholds of BPNN were obtained through optimization results of the multi-agent cooperation grouping algorithm." So, this approach is much more robust. Thus, according to robustness, i vote [Wu et al., 2014] as best paper and, [Guo et al., 2017] and [Tao et al., 2017] together comes next.

From reliability point of view: [Guo et al., 2017] said: "In particular, whatever the operating condition of a bearing is, the RNN-HI is almost equal to one at the failure condition." So, this approach can cope with multiple operating condition. [Tao et al., 2017] utilized multiple geometric metrics but it is not described explicitly in the paper whether the proposed approach is reliable. According to [Wu et al., 2014], proposed method indicated as reliable but it was not clear. According to reliability, i would rank [Guo et al., 2017] as best paper next [Wu et al., 2014] and at last [Tao et al., 2017].

(3) Anomaly Detection for Machine Maintenance:

[Ramos et al.,] "The prediction of failures and maintenance actions of industrial machines is a problem with interesting characteristics". The author interested in forecasting certain rare events (i.e. disc replacement), which are dependent on the recent values of a set of time series values. The time series describe the recent values of a set of sensors that monitor several aspects of the industrial machines. To maintain well working condition, the machine's sensors are expected to have a certain typical behavior. Variations from this typical behavior are good indicators of a failure prediction or some maintenance action. So, author's approach to the first step for addressing a predictive study was the construction of a good quality data set which provided the models with examples of the rare events to forecast. Author's focus was on the maintenance of the defibrator discs of the refiner machine, in particular on their replacement event. To be able to predict such event, failure and maintenance activity histories must be analyzed. The data set was consisting of a time tagged sequence of observations of the machine state. The second step was to predict the future values of the sensors of the machine which has been done by applying Artificial Neural Network(ANN) technique. Autoregressive Feed-Forward Neural Networks(FFNN), a variation of ANNs, was trained and tested on refiner machine data. The available data was divided into three parts. The first training part was used for training the neural network based model, while the second part was for model selection by comparing neural network based model with Auto Regressive Moving Average(ARIMA) model. The last test sample was then used for true prediction evaluation. To evaluate and compare the forecasting performance of the ANN model with ARIMA model, they used three most popular overall error measures in this study: the root mean squared error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE). The result says that neural networks are able to model and forecast better than ARIMA in the in-sample judged by the three performance measures. However, the out-of-sample forecasting performance of ARIMA models evaluated via RMSE, MAE and MAPE was better than the NN models.

[Zhao et al., 2017] Tool wear is the gradual failure of machine or parts due to regular operations. The proposed method is able to predict the actual tool wear event based on raw sensory considering the time series, noise, varying length and irregular sampling behind sensory data. However, feeding raw sensory data directly into regression/classification model is difficult and useless. So, at first, Convolutional Neural Network(CNN) was used to extract local features that are robust and informative from the sequential time series input. Then, Convolutional bi-directional Long Short Term Memory(CBLSTM) was used to encode temporal information where Long Short-Term Memory networks (LSTMs) are able to capture long-term dependencies and model time series data, and the bi-directional structure enables the capture of past and future contexts to predict the target value. To validate the proposed model, a real-life tool wear test was introduced and performance of proposed method was compared with several baseline tool wear event predictors i.e. Linear Regression(LR) on extracted features of raw signal, Support Vector Regression(SVR) on extracted features of the raw signal, Multi-layer Perceptron(MLP) Neural Network on extracted features of the raw signal, Basic RNN, Deep RNN etc. To compare performance, two well-known measures are used - Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The result shows that proposed CBLSTM is able to predict the actual tool wear based on raw sensory data and is able to perform better than several state-of-the-art baseline methods mentioned earlier.

[Zhang et al.,] In this paper, author considered, the problem of event prediction with multivariate time series data consisting both continuous and categorical variables, and reconstructed the (rare) event prediction task as a classification problem. First, time-series data was converted into symbolic representations which is a simple way of reducing the dimensionality by turning time series into sequences of symbols. Then, symbolic representations were fed into an a Long Short Term Memory Neural Network (LSTM) Recurrent Neural Network (RNN) which was trained to learn from discriminative features (based on importance). The proposed approach showed effective performance on real-world industrial datasets. The hard drive dataset released from Backblaze data center was randomly split into a training set (containing 486 positives) and a test set (containing 100 positives) using hard disk serial number and with the temporal information. The results showed that the proposed method achieves cutting edge performance on two problems of hard disk failure prediction and heating system fault prediction. The proposed method was compared against Logistic Regression Classification (LR) on normalized raw data (without symbolization). The performance metrics for evaluation were balanced accuracy, arithmetic mean of the true positive and true negative rates, the area under curve (AUC) of ROC. Author mentioned based on the results, "it is seen that the proposed LSTM RNN based method with symbolization achieves the best performance".

Comparisons: As i mentioned in definition of this cluster, this group of literature predicted certain rare events. The approaches are also based on ANN like other clusters except variations in their architecture. The literatures inside cluster could be compared based on: Predictive quality, simplicity and ability to capture temporal dependency.

From predictive quality point of view: RMSE and MAE of [Zhao et al., 2017] are 10.8 and 7.5 receptively. For [Ramos et al.,], RMSE 1.796 and MAE 1.561. From predictive quality point of view i would recommend [Ramos et al.,] than [Zhang et al.,].

From simplicity point of view: [Ramos et al.,] used FFNN which is the most simplest one, [Zhao et al., 2017] used CBLSTM RNN which is least simplest one and [Zhang et al.,] used LSTM RNN. So, in my opinion, [Ramos et al.,] is the best according to model simplicity , next [Zhang et al.,] and at last [Zhao et al., 2017].

From ability of encoding temporal information point of view: LSTM is capable of encoding long and short term dependencies. [Zhao et al., 2017] and [Zhang et al.,] used CBLSTM and LSTM variations of RNN. Both of them are capable of encoding temporal dependency but CBLSTM is comparatively more as it is bidirectional. In other hand, FFNN is least capable. So, i would rank [Zhao et al., 2017] as best, then [Zhang et al.,] and finally [Ramos et al.,]

References

[Arabgol et al.,] Arabgol, S., Ko, H. S., and Esmaeili, S. Artificial Neural Network and EWMA-based Fault Prediction in Wind Turbines. *2015 Industrial and Systems Engineering Research Conference.*

[Chen X, 2017] Chen X, Xu D, X. L. (2017). Joint optimization of replacement and spare ordering for critical rotary component based on condition signal to date. *Eksploatacja i Niezawodnosc – Maintenance and Reliability*, 19(1, 2017):76–85.

[Deepthi and Rao, 2014] Deepthi, A. S. and Rao, K. V. (2014). Anomaly Detection Using Principal Component Analysis. *InternatIonal Journal of Computer SCIenCe and teChnology*, 5(4).

[ElSaid et al., 2017] ElSaid, A., Wild, B., Higgins, J., and Desell, T. (2017). Using LSTM recurrent neural networks to predict excess vibration events in aircraft engines. In *Proceedings of the 2016 IEEE 12th International Conference on e-Science, e-Science 2016.*

[Guo et al., 2017] Guo, L., Li, N., Jia, F., Lei, Y., and Lin, J. (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing.*

[Javed, 2014] Javed, K. (2014). *A robust and reliable data-driven prognostics approach based on Extreme Learning Machine and Fuzzy Clustering.* PhD thesis, Université de Franche-Comté.

[Kadirgama et al., 2011] Kadirgama, K., Abou-El-Hossein, K. A., Noor, M. M., Sharma, K. V., and Mohammad, B. (2011). Tool life and wear mechanism when machining Hastelloy C-22HS. *Wear.*

[Martínez-Martínez et al., 2015] Martínez-Martínez, V., Gomez-Gil, F. J., Gomez-Gil, J., and Ruiz-Gonzalez, R. (2015). An Artificial Neural Network based expert system fitted with Genetic Algorithms for detecting the status of several rotary components in agro-industrial machines using a single vibration signal. *Expert Systems with Applications.*

[Nikula et al.,] Nikula, R.-P., Karioja, K., Leiviskä, K., and Juuso, E. Prediction of mechanical stress in roller leveler based on vibration measurements and steel strip properties. *Journal of Intelligent Manufacturing.*

[Ramos et al.,] Ramos, P., Oliveira, J. M., and Silva, P. Predictive maintenance of production equipment based on neural network autoregression and ARIMA.

[Romero Ugalde et al., 2014] Romero Ugalde, H. M., Carmona, J. C., Reyes-Reyes, J., Alvarado, V. M., and Corbier, C. (2014). Balanced simplicity–accuracy neural network model families for system identification. *Neural Computing and Applications.*

[Sankararaman,] Sankararaman, S. Remaining Useful Life Prediction through Failure Probability Computation for Condition-based Prognostics.

[Si et al., 2011] Si, X. S., Wang, W., Hu, C. H., and Zhou, D. H. (2011). Remaining useful life estimation - A review on the statistical data driven approaches. *European Journal of Operational Research.*

[Tao et al., 2017] Tao, L., Yang, C., Cheng, Y., Lu, C., and Ragulskis, M. (2017). Machine component health prognostics with only truncated histories using geometrical metric approach. *Mechanical Systems and Signal Processing.*

[Wu et al., 2014] Wu, W., Zhang, F., Liu, M., and Shen, W. (2014). A multi-agent based failure prediction method using neural network algorithm. In *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2268–2272.

[Xiao et al., 2017] Xiao, S., Yan, J., Chu, S. M., Yang, X., and Zha, H. (2017). Modeling The Intensity Function Of Point Process Via Recurrent Neural Networks. *AAAI17.*

[Xu et al., 2012] Xu, H., Caramanis, C., and Mannor, S. (2012). Sparse algorithms are not stable: A no-free-lunch theorem. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

[Yang et al., 2017] Yang, H.-C., Li, Y.-Y., Hung, M.-H., and Cheng, F.-T. (2017). A cyber-physical scheme for predicting tool wear based on a hybrid dynamic neural network. *Journal of the Chinese Institute of Engineers*, 40(7):614–625.

[Zhang et al.,] Zhang, S., Bahrampour, S., Ramakrishnan, N., Schott, L., and Shah, M. DEEP LEARNING ON SYMBOLIC REPRESENTATIONS FOR LARGE-SCALE HETEROGENEOUS TIME-SERIES EVENT PREDICTION. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, 2017, pp. 5970-5974*.

[Zhao et al., 2017] Zhao, R., Yan, R., Wang, J., and Mao, K. (2017). Learning to monitor machine health with convolutional Bi-directional LSTM networks. *Sensors (Basel). 2017 Jan 30;17(2).*