

# Research on Tool Wear Prediction Based on LSTM and ARIMA

Zhenkun Zhang

College of Mechanical & Engineering  
Guangxi University  
Nanning, Guangxi, China  
86-15271857990  
923182971@qq.com

Gang Zhou

College of Mechanical & Engineering  
Guangxi University  
Nanning, Guangxi, China  
86-15578870801  
1374736576@qq.com

Juan Lu

Department of Mechanical & Engineering  
Qinzhou University  
Qinzhou, Guangxi, China  
86-15907773431  
lujuan3623366@163.com

Xiaoping Liao

College of Mechanical & Engineering  
Guangxi University  
Nanning, Guangxi, China  
86-13977109215  
xpfeng@gxu.edu.cn

## ABSTRACT

Precise tool wear prediction is the key to improving the productivity of the entire workpiece. Reliable tool wear prediction technology can reduce the machine downtime caused by the tool change process, and can also make the entire machining process more efficient. This paper briefly reviews the research of tool wear prediction technology, and then uses the ARIMA (p, d, q) model in the time series and the LSTM model to predict the wear of the tool respectively. The results show that the LSTM model has better prediction performance.

## CCS Concepts

•Mathematics of computing → Time series analysis • Applied computing → Forecasting • Information systems → Temporal data.

## Keywords

Tool wear prediction; ARIMA; LSTM

## 1. INTRODUCTION

With the continuous development of automation and intelligence in the machining process, the research of intelligent monitoring technology for mechanical faults is crucial. Modern sustainable manufacturing can bring about better product quality, higher flexibility and productivity [1-3]. In general, these benefits depend on the trouble-free operation of various machine components [4]. In the industry, 20% of downtime is attributed to tool failure [5]. The tool is one of the most important processing elements in machining. Tool wear not only directly affects the dimensional accuracy and surface quality of the workpiece, but also indirectly affects the processing efficiency and production cost. By

predicting the tool wear condition and implementing an efficient tool change strategy, it is possible to ensure that the entire manufacturing process is in a normal wear state. This plays an important role in improving machine productivity, maintaining the quality and integrity of machined parts, minimizing material waste, and reducing the cost of sustainable manufacturing.

How to predict tool wear through more effective methods has become the focus of researchers [6-7], and they have done lots of research on tool wear prediction and have made great progress. Time series analysis is very important in a series of research topics and many engineering problems, such as analyzing time series data to estimate meaningful statistical and pattern characteristics of continuous data. Specifically, future values are predicted based on previously observed measurements. The causality between tool wear timings is not completely clear and nonlinear, it is difficult to analyze by establishing a mathematical model, but the time series method can solve this problem better [8]. Many works use statistical methods to study time series data. Kalman filter [9] and autoregressive (AR) models are important statistical methods, in which the ARMA [10] and ARIMA [11] (autoregressive integral moving average) models are AR model and moving average (MA) model's further promotion. Liang et al. [12] used autoregressive time series to simulate the acoustic emission generated during the cutting process. Yao et al. [13] used a stochastic technique based on the Multiple Autoregressive Moving Average Vector Model (ARMAV) to estimate the various wears generated on different surfaces of the tool. Kumar et al. [14] studied the AR model parameters and AR residual signals and found that it showed an effective and consistent tool wear trend.

In 2006, Hinton et al. [15] proposed an unsupervised layer-by-layer greedy training algorithm based on Deep Belief Networks (DBN) to open the wave of deep learning[16](DL) applications in academia and industry. In many deep learning models, the recurrent neural network (RNN) introduces the concept of timing into network structure design, making it more adaptable in the analysis of time series data. Among the many RNN variants, the long short-term memory (LSTM) model [17] compensates for the disappearance of gradients, gradient explosions, and lack of long-term memory capacity of RNNs, making recurrent neural networks possible to effectively use long-range timing

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

BDET 2018, August 25–27, 2018, Chengdu, China

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-6582-6/18/08...\$15.00

DOI: <https://doi.org/10.1145/3297730.3297732>

information. The LSTM model is applied successfully in the research of time series data in different fields, including language modeling, speech recognition, machine translation [18], multimedia-related audio and video data analysis, picture title modeling [19-20], prediction of traffic flow related to road transportation [21], and prediction of secondary sequence of protein related to medicine [22]. The LSTM model can also be used for the analysis and processing of big data [23]. Chen et al. [24] used the LSTM model to fit the training of 900,000 sequences in the Chinese stock market and uses other 311,361 sequences for testing, which increased the accuracy of stock returns prediction from 14.3% to 27.2%. Wang et al. [25] proposed a hybrid space-time prediction deep learning model based on the LSTM model to analyze a large data set of China Mobile and improved the prediction accuracy significantly. However, the application of the LSTM model is very limited in the field of reliability, especially for the study of fault time series prediction, and no related research has been found yet.

In this paper, the tool wear is predicted by the time series and LSTM models. The results show that the LSTM model has a better prediction effect. The rest of this paper is organized as follows. Section II introduces the methodology of the time series model and the LSTM model. ARIMA (p, d, q) model and the LSTM model are applied to predict the tool wear time series in Section III. The last part is summarized.

## 2. METHODOLOGY

### 2.1 Methodology of Time Series Model

Let  $\{Y_t\}$  be a zero-mean stationary time series, and have the following relational expression between any time  $t$  and the previous  $q$  time series values:

$$Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (1)$$

Where  $\varphi_i$  is the autoregressive coefficient;  $p$  is the order of autoregression;  $\theta_j$  is the moving average coefficient;  $q$  is the order of the moving average;  $\{\varepsilon_t\}$  is the zero mean white noise process, and there is  $E(\varepsilon_s X_t) = 0$ ,  $s > t$ , then  $Y_t$  is an autoregressive moving average sequence of order  $p$  and  $q$ , abbreviated as ARMA( $p$ ,  $q$ ). When  $q=0$ , the sequence is an AR ( $p$ ) sequence; when  $p=0$ , the sequence is a MA ( $q$ ) sequence.

The ARMA ( $p$ ,  $q$ ) model is for stationary random sequences. The characteristic of a stationary random sequence is that its statistical properties do not change over time. However, many phenomena in reality will show some kind of trend line or periodicity over time. If the non-stationary time series  $Y_t$  becomes stable after  $d$  differences, the time series model is an ARIMA model, usually denoted as ARIMA ( $p$ ,  $d$ ,  $q$ ), where  $p$ ,  $d$ , and  $q$  are the number of autoregressive items, difference times, moving average terms respectively [26]. In practical applications, these three parameters can be determined by observing the autocorrelation function (ACF) and the partial autocorrelation function (PACF), or by calculating the minimum AIC (Akaike information criterion) or BIC (Bayesian information criterion) value.

### 2.2 Methodology of LSTM

For a given sequence  $\vec{x} = (x_1, x_2, \dots, x_n)$ , applying a standard RNN model (as shown in Fig.1) [27], we can calculate a hidden layer sequence  $\vec{h} = (h_1, h_2, \dots, h_n)$  and an output sequence  $\vec{y} = (y_1, y_2, \dots, y_n)$  by iterating equations (2)-(3), where  $W$  denotes a weight coefficient matrix (such as  $W_{xh}$  denotes the weight coefficient matrix of the input layer to the hidden layer), and  $b$  denotes an offset vector (For example,  $b_h$  denotes the hidden

layer's offset vector) and  $f$  denotes the activation function (e.g. tanh function).

$$h_t = f(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (2)$$

$$y_t = W_{hy}h_t + b_y \quad (3)$$

Although RNN can effectively deal with non-linear time series, there are still the following two problems [21]: (1) RNN cannot handle long-delayed time series due to the disappearance of gradients and gradient explosion. (2) Training the RNN model requires a predetermined delay window length, but it is difficult to obtain the optimal parameter automatically in practical.

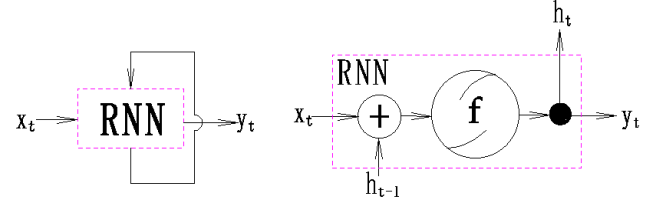


Figure 1. RNN network (left) and cell structure in hidden layer (right)

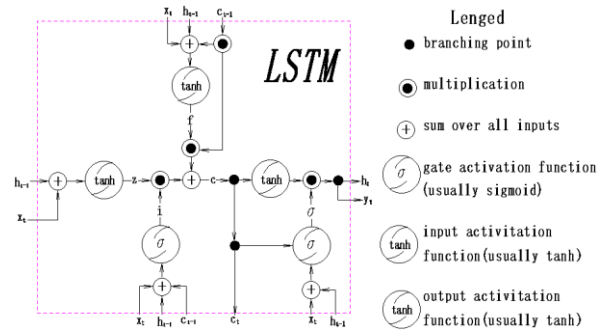


Figure 2. LSTM cell structure in hidden layer.

The LSTM model includes forward calculation methods, time-based back propagation through time (BPTT) algorithms, Adam parameter optimization algorithms, and related RNN and GRU models. The single node structure of the LSTM is shown in Fig.1 below. The ingenious aspect of LSTM lies in adding the input threshold, the forget threshold, and the output threshold, which leads to the changing weight of the self-looping. In this way, the integral scales at different times can be dynamically changed when the model parameters are fixed, thus avoiding gradient disappearance or gradient expansion. The most widely used LSTM model cell structure is shown in Fig.2 [27], and its forward calculation method can be expressed as

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (4)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (5)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (6)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (7)$$

$$h_t = o_t \tanh(c_t) \quad (8)$$

Among them,  $i_t$ ,  $f_t$ ,  $c_t$ ,  $o_t$ ,  $h_t$  represent the input gate, forget gate, cell state, and output gate, respectively.  $W$  and  $b$  are the corresponding weight coefficient matrix and offset term,  $\sigma$  and  $\tanh$  are sigmoid and hyperbolic tangent activation function, respectively. The LSTM model training process adopts a BPTT algorithm similar to the classical back propagation (BP) principle [28]. It can be roughly divided into four steps: (1) According to the forward calculation method (Equation (4)-(8)) calculate the

output value of LSTM cells. (2) Calculate the error term for each LSTM cell in reverse, including two counter-propagating directions by time and network hierarchy. (3) Calculate the gradient of each weight based on the corresponding errors. (4) Apply gradient-based optimization algorithm to update the weights.

### 3. EVALUATION AND RESULT

#### 3.1 Evaluation and Results of Time Series

The figure below shows the actual measured flank wear and the number of passes. The tool wear data adopted in this paper comes from the paper of Li [29]. The first 275 values in the tool wear time series are used to predict the next 40 values in the tool wear time series (i.e. tool wear from 276 to 315). In order to avoid the parameters in the ARMA model being too small, the amount of wear is converted to microns in the predicted calculation. According to the amount of wear of the first 275 passes shown in Fig.3, it can be clearly seen that the amount of wear on the tool increases rapidly, and the different growth rate indicates that the series has both an upward trend and a variance. At this time, the tool wear time series needs to be smoothed. Differential conversion is performed on the sequence data, and the autocorrelation function of the sequence after one-stage differential conversion is shown in the Fig.4 below. The autocorrelation function plot and partial autocorrelation function plot of the first-order differential conversion sequence are shown in Fig.5. It can be seen from the figure that the sequence after the first-order differential has no trend, and its autocorrelation function and partial correlation function can be considered as trailing, so the first-order differential sequence is a stationary sequence.

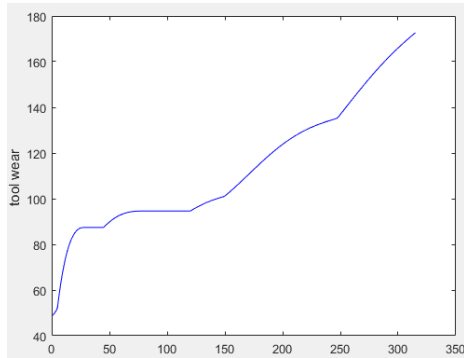


Figure 3. Tool flank wear.

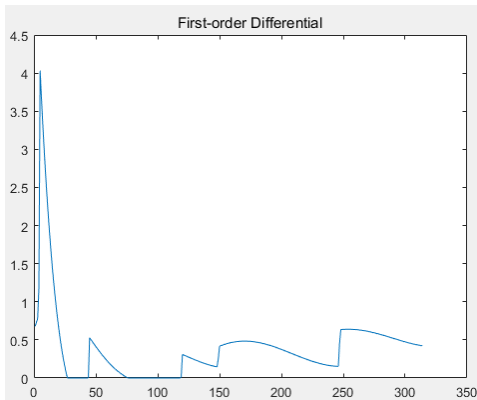


Figure 4. First-order differential sequence.

Calculate the AIC parameters and BIC parameters of each ARIMA(p,1,q) model. As shown in Table 1 and Table 2, ARIMA(2,1,0) is selected as the optimal model for the original sequence prediction according to the minimum information criterion. (Some of the grids in Table I and Table II are blank because the related values cannot be calculated but do not affect the result.)

Mean absolute percentage error(MAPE) is usually used as an index of accuracy measurement, mainly reflecting the degree of conformity of the model. Its mathematical expression is

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\varphi(t) - y(t)}{y(t)} \right| \times 100\%, t = 1, 2, \dots, n. \quad (9)$$

This is an overall evaluation method, and the model is evaluated from the point of "average". The ARIMA(2,1,0) model established in this paper was evaluated using the MAPE criterion and the forecast result is shown in Fig.6, MAPE=0.2043. The autocorrelation test and partial correlation test of the selected optimal model ARIMA (2, 1, 0) are shown in Fig.7, which shows that the model is appropriate.

Table 1. AIC value of each order

	MA(0)	MA(1)	MA(2)	MA(3)	MA(4)	MA(5)
AR(0)	465.89	150.49				
AR(1)	-199.36	-203.52	-201.76	-200.03	-198.86	-197.33
AR(2)	-203.81	-202.78	-202.24	-200.68		
AR(3)	-201.88	-202.09	-200.40			
AR(4)	-200.19	-200.87				
AR(5)	-200.24	-200.49	-198.59			

Table 2. BIC value of each order

	MA(0)	MA(1)	MA(2)	MA(3)	MA(4)	MA(5)
AR(0)	473.39	161.73				
AR(1)	-188.11	-188.52	-183.01	-177.54	-172.62	-167.33
AR(2)	-188.81	-184.04	-179.74	-174.44		
AR(3)	-183.14	-179.59	-174.16			
AR(4)	-177.69	-174.62				
AR(5)	-173.99	-170.50	-164.85			

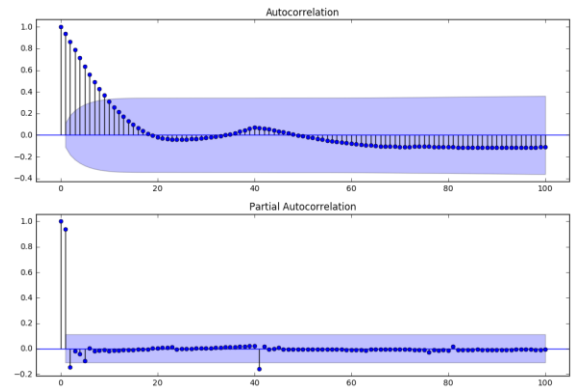


Figure 5. First-order Differential autocorrelation function diagram and its partial correlation function diagram.

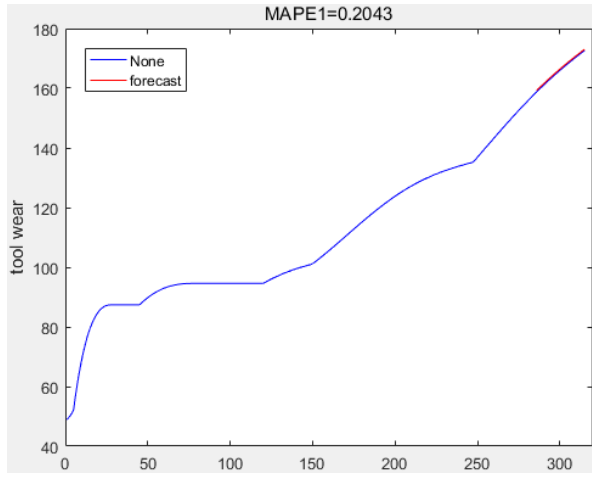


Figure 6. Prediction of ARIMA.

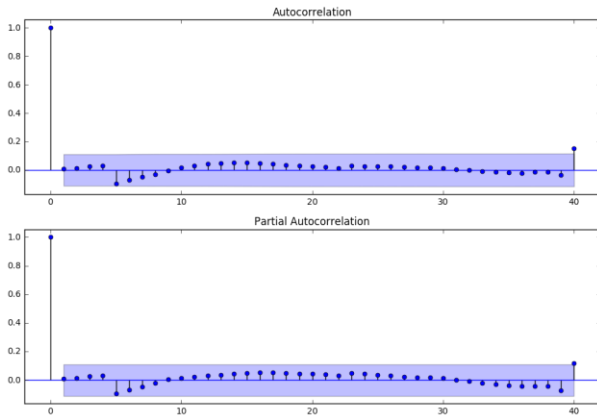


Figure 7. Test of predictive model.

### 3.2 Evaluation and results of LSTM

Based on the method described above, an LSTM prediction model is also established for tool wear time series. The LSTM prediction model established in this paper includes five functional modules: input layer, hidden layer, output layer, network training, and network prediction. The input layer is responsible for the initial processing of the original fault time series to meet the network input requirements. The hidden layer uses the LSTM cells shown in Fig.2 to build a single-layer recurrent neural network. The mean square error (MSE) is used as the loss function, and the output layer provides the prediction result. The network training uses RMSProp optimization method, network prediction adopts iterative method to point-by-point prediction. We apply the trained LSTM network for prediction which can improve final performance [30]. The forecasting process uses an iterative method. The model was also evaluated using the MAPE criterion and the forecast result is shown in Fig.9, MAPE=0.0414.

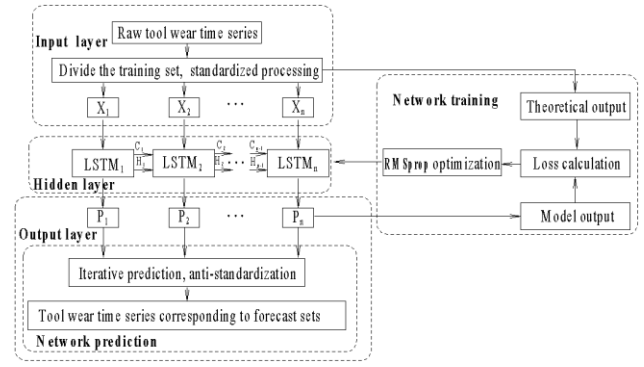


Figure 8. Prediction process of tool wear time series based on LSTM

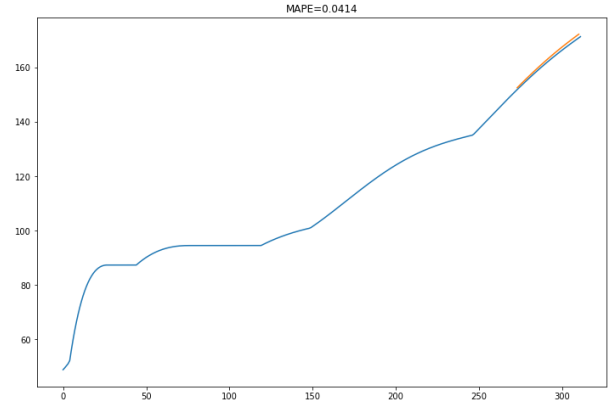


Figure 9. Prediction of LSTM model.

## 4. CONCLUSION

The accurate prediction of tool wear in real time has been something many researchers have been striving for. This paper briefly summarized the study of tool wear prediction for LSTM model and time series model, and described the basic principles of these two models, and then used the time series model and the LSTM model to predict tool wear. Through the ARIMA (2, 1, 0) model, the predicted MAPE was 0.2043, and the predicted MAPE was 0.0414 by the LSTM model. Although the time series model is widely used in tool wear prediction, it can be found that its accuracy is not high and the operation is more complex than the LSTM model, because it needs to determine the parameters before it determines the final model and then predicts it. The LSTM model is more convenient, and it can get the predicted result by calling the function block directly. In conclusion, LSTM performs better in tool well prediction.

## 5. ACKNOWLEDGMENTS

The work was supported by the National Natural Science Foundation of China (NSFC) (51665005), Innovation Project of Guangxi Graduate Education (YCBZ2017015), and the Project of Guangxi Colleges and Universities Key Laboratory Breeding Base of Coastal Mechanical Equipment Design, Manufacturing and Control (GXLH2016ZD-06).

## 6. REFERENCES

- [1] Wiendahl, H. P., ElMaraghy, H. A., Nyhuis, P., Zih, M. F., Wiendahl, H. H., Duffie, N., and Brieke, M. 2007. Changeable manufacturing-classification, design and operation. *CIRP annals*, 56(2), 783-809. DOI=<https://doi.org/10.1016/j.cirp.2007.10.003>

- [2] Bi, Z. M., Lang, S. Y., Shen, W., and Wang, L. 2008. Reconfigurable manufacturing systems: the state of the art. *International Journal of Production Research*, 46(4), 967-992. DOI= <https://doi.org/10.1080/00207540600905646>
- [3] Mähring, H. C., Litwinski, K. M., and Gümmer, O. 2010. Process monitoring with sensory machine tool components. *CIRP annals*, 59(1), 383-386. DOI= <https://doi.org/10.1016/j.cirp.2010.03.087>
- [4] Zhou, Z. D., Chen, Y. P., Fuh, J. Y. H., and Nee, A. Y. C. 2000. Integrated condition monitoring and fault diagnosis for modern manufacturing systems. *CIRP Annals-Manufacturing Technology*, 49(1), 387-390. DOI= [https://doi.org/10.1016/S0007-8506\(07\)62971-0](https://doi.org/10.1016/S0007-8506(07)62971-0)
- [5] Kurada, S., and Bradley, C. 1997. A review of machine vision sensors for tool condition monitoring. *Computers in industry*, 34(1), 55-72. DOI= [https://doi.org/10.1016/S0166-3615\(96\)00075-9](https://doi.org/10.1016/S0166-3615(96)00075-9)
- [6] Waydande, P., Ambhore, N., and Chinchani, S. 2016. A review on tool wear monitoring system. *Journal of Mechanical Engineering and Automation*, 6(5A), 49-53.
- [7] Sick, B. 2002. On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research. *Mechanical Systems and Signal Processing*, 16(4), 487-546. DOI= <https://doi.org/10.1006/mssp.2001.1460>
- [8] Yang Shuzi, Wu Ya, Xuan Jianping et al. Tool wear time series in engineering application, 2nd ed., vol. 1, Wuhan, China: Huazhong University of Science and Technology Press, June 2007, 1-7.
- [9] Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1), 35-45.
- [10] Durbin, J. 1959. Efficient estimation of parameters in moving-average models. *Biometrika*, 46(3/4), 306-316.
- [11] Box, G. E., and Pierce, D. A. 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332), 1509-1526.
- [12] Liang, S. Y., and Dornfeld, D. A. 1989. Tool wear detection using time series analysis of acoustic emission. *Journal of Engineering for Industry*, 111(3), 199-205.
- [13] Yao, Y., & Fang, X. D. 1992. Modelling of multivariate time series for tool wear estimation in finish-turning. *International Journal of Machine Tools and Manufacture*, 32(4), 495-508.
- [14] Kumar, S. A., Ravindra, H. V., and Srinivasa, Y. G. 1997. In-process tool wear monitoring through time series modelling and pattern recognition. *International Journal of Production Research*, 35(3), 739-751. DOI= <https://doi.org/10.1080/002075497195687>
- [15] Hinton, G. E., and Salakhutdinov, R. R. 2006. Reducing the dimensionality of data with neural networks. *science*, 313(5786), 504-507.
- [16] LeCun, Y., Bengio, Y., and Hinton, G. 2015. Deep learning. *nature*, 521(7553), 436.
- [17] Graves, A. Supervised sequence labelling with recurrent neural networks. 2012. ISBN 9783642212703. URL <http://books.google.com/books>.
- [18] Srivastava, N., Mansimov, E., and Salakhutdinov, R. 2015. Unsupervised learning of video representations using lstms. In *International conference on machine learning*, (Jun.2015), 843-852.
- [19] Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., and Darrell, T. 2015. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2625-2634.
- [20] Vinyals, O., Toshev, A., Bengio, S., and Erhan, D. 2015. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3156-3164.
- [21] Ma, X., Tao, Z., Wang, Y., Yu, H., and Wang, Y. 2015. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, 54, 187-197. DOI= <https://doi.org/10.1016/j.trc.2015.03.014>
- [22] Hanson, J., Yang, Y., Paliwal, K., and Zhou, Y. 2016. Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks. *Bioinformatics*, 33(5), 685-692. DOI= <https://doi.org/10.1093/bioinformatics/btw678>
- [23] Scardapane, S., Wang, D., and Panella, M. 2016. A decentralized training algorithm for echo state networks in distributed big data applications. *Neural Networks*, 78, 65-74. DOI= <https://doi.org/10.1016/j.neunet.2015.07.006>
- [24] Chen, K., Zhou, Y., and Dai, F. 2015. A LSTM-based method for stock returns prediction: A case study of China stock market. In *Big Data (Big Data), 2015 IEEE International Conference on*, (Oct. 2015), 2823-2824.
- [25] Wang, J., Tang, J., Xu, Z., Wang, Y., Xue, G., Zhang, X., and Yang, D. 2017. Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach. In *INFOCOM 2017-IEEE Conference on Computer Communications, IEEE*, (May. 2017), 1-9.
- [26] Rangwala, S., and Dornfeld, D. 1990. Sensor integration using neural networks for intelligent tool condition monitoring. *Journal of Engineering for Industry*, 112(3), 219-228.
- [27] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. 2017. LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28(10), 2222-2232.
- [28] Graves, A., and Schmidhuber, J. 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5-6), 602-610.
- [29] Li, X., Lim, B. S., Zhou, J. H., Huang, S., Phua, S. J., Shaw, K. C., and Er, M. J. 2009. Fuzzy neural network modelling for tool wear estimation in dry milling operation. In *Annual conference of the prognostics and health management society*, (Sep. 2009), 1-11.
- [30] Graves, A., Beringer, N., and Schmidhuber, J. 2005. Rapid retraining on speech data with lstm recurrent networks. *Technical Report IDSIA-09-05, IDSIA*.