Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 118 (2023) 431–436

www.elsevier.com/locate/procedia

16th CIRP Conference on Intelligent Computation in Manufacturing Engineering, CIRP ICME '22, Italy

Multi-stream big data mining for industry 4.0 in machining: novel application of a Gated Recurrent Unit Network

Federica Garghetti[a,*], Marco Grasso[a], Massimo Pacella[b], Giuseppe Fogliazza[c], Bianca Maria Colosimo[a]

[a]*Department of Mechanical Engineering, Politecnico di Milano, Via La Masa 1, 20156, Milan, Italy*
[b]*Department of Engineering for Innovation, Università del Salento, Lecce, Italy*
[c]*MCM S.p.a, Viale Federico e Guido Celaschi 19, 29020, 29020 Borgo di Sotto (PC), Italy*

* Corresponding author. Tel.: 02.2399.8575; *E-mail address:* federica.garghetti@polimi.it

Abstract

In Industry 4.0, the availability of signals from multiple sensors stimulates the investigation of novel quality monitoring and prediction methods. This paper tackles the in-line machining process monitoring by exploiting big data in the shape of multi-stream complex signals, eventually containing degradation and tool wear signatures. The proposed novel solution is fed by real-time multichannel data to identify anomalous states in machining applications. We investigate the effectiveness of a category of ANNs specifically conceived to predict process patterns based on time series of sensor signals, i.e., the Gated-Recurrent-Unit-Network. A real case study shows the efficiency of the proposed solution in predicting wild, complex and drifting patterns, typical of real productions, highlighting its provided benefits for in-line big data mining in industrial applications.

Keywords: Industry 4.0; Multichannel signals; Recurrent Neural Network; Gated-Recurrent-Unit

1. Introduction

In Industry 4.0, the increasing availability of complex signal patterns coming from multiple sensors during every production stage open various challenges, stimulating the investigation and the development of novel methodologies. Since data-rich environments represent a suitable playground for machine learning and artificial neural networks (ANNs), these techniques have been more and more integrated into a wide variety of business workflows and industrial applications, from enhanced product design to in-line monitoring and statistical quality control to process optimization.

In the field of machining process monitoring, decades of research and hundreds of works have been devoted to machine learning and ANN solutions for in-line detection of process anomalies, tool wear analysis and prediction, machine degradation states, etc (the reader is referred to Serin et al., 2020, Mohanraj et al., 2020, Wong et al., 2020, Zhou et al., 2018, Lee et al., 2019 and Li et al., 2022 for recent reviews on this research area). Nevertheless, industrial off-the-shelf toolkits for machining process monitoring still lack any advanced multi-stream / multi-channel data analysis capabilities, as they commonly rely on simple and conservative alarm rules. Indeed, there is still a wide gap between the several methods investigated in the literature and their actual industrial implementation. One reason is that ANN-based techniques are commonly applied to highly repeating operations, where big training datasets consisting of several realizations of a simple process run are needed. However, such condition represents a limited portion of actual production scenarios, where continuous tuning and recalibration of process settings is performed, cutting conditions are highly time-varying, signals exhibit dynamically changing patterns, etc. Most ANNs are not able to adapt to drifting and non-stationary processes, and they need quite extensive training phases.

10.1016/j.procir.2023.06.074

RNNs have the potential to overcome both these two limitations, as they can be initialized even in the presence of a limited training data availability, and they are specifically designed to adapt to time-varying process patterns. They are composed of a sequence of recursively connected hidden layers, allowing the storage of historical information inside the network's internal state. RNNs are flexible models, capturing the time-varying dynamics of the system. Thanks to their internal memory, they are suitable to deal with historical dependency and complex patterns in signals analysis, which make them particularly suitable to deal with drifting and non-stationary processes typically observed in discrete manufacturing (Tuttle et al., 2021).

Previous studies investigated the use of RNNs in the field of process monitoring of manufacturing operations. More specifically, recent examples of seminal studies were devoted to assessing the i) three-directional force based on in-line electric drive signal measurement (Denkena et al., 2020, Li et al., 2021, Zhao et al., 2019) and ii) spindle vibrations relying on spindle speed and load signals (Pian et al., 2021).

Another field where RNNs have attracted researchers' interest is tool condition monitoring: various works focused on estimating the remaining useful life (RUL) of the cutting tool and on the in-line prediction of the tool wear state. In this context, long-term predictions are still problematic issues (Wang et al., 2019). Some authors investigate a tool wear estimation based on single-channel RNNs, acquiring vibration signals (Zhang et al., 2021), current signals (Marani et al., 2021), and decomposed cutting force (Wu et al., 2019). Other works investigated the possibility to train multichannel RNNs (i.e., combining several signals like vibrations, forces, currents, acoustic emissions, and eventually the working conditions, depth of cut, feed rate etc. as inputs) to predict tool wear and estimate its remaining useful life (Kerboua et al., 2018, Zhou et al., 2019, Zhao et al., 2016, Wang and Huang, 2009, Lee et al., 2020, Wang et al., 2019, Proteau et al., 2019).

Despite the intrinsic potential of ANNs in dealing with high dimensional data, the literature devoted to multi-channel RNNs is still limited, as they are more commonly adopted for in-line prediction of single signals or in a multi-input / single-output form, where one output quantity of interest is estimated starting from multiple input process parameters and in-line measurable variables (Nasir and Sassani, 2021). Li et al., 2022 also pointed out the scarcity of multichannel solutions in tool breakage detection, where some works applied parametric modeling methods (Balsamo et al., 2016, Heinemann et al., 2012, Kang et al., 2019) or shallow networks (Haili et al., 2003, Lo, 2002), but the use of multichannel RNNs has been poorly explored to this aim.

This paper presents a technique belonging to the family of RNNs known as Gated Recurrent Unit (GRU) network (Cho et al., 2014) to model multi-channel, non-stationary and complex, signal patterns, typical of a series production of complex shapes where the cutting tool engagement varies along time, with a part-to-part repeating signature affected by non-stationary conditions due to the tool wear state. The current work aims to achieve a robust prediction of the next-run process pattern, i.e., the process pattern during the production of the next part, with a limited training phase. An accurate prediction of the process signature is of key importance to detect sudden anomalies quickly and effectively, like tool chipping and breakage events, even in the presence of highly time-varying cutting conditions that commonly mask such events, making them quite difficult to detect.

One typical issue in the machine learning framework consists of how to select and tune the several hyperparameters of the network. In most cases, such selection is based on the data analyst's experience and commonly shared guidelines. In this study, a Bayesian optimization (BO) algorithm was combined with the GRU methodology for the fast and robust identification of the optimal values of the GRU hyperparameters (Yang and Shami, 2020). Since the future evaluation points in the GRU adaptive scheme are based on the previous results, BO achieves more efficient performances than other commonly used algorithms (e.g., grid search). Furthermore, BO is recommended for tuning a small set of hyperparameters, as in the current study, differently from the more complicated metaheuristic algorithms (i.e., particle swarm optimization) suitable for large configuration spaces (Yang and Shami, 2020).

Section 2 describes the proposed methodology and section 3 presents the real industrial case study. Section 4 presents the performance of the network and section 5 concludes the work, opening future challenges and applications.

2. Proposed Methodology

The following bullet points summarize the relevant steps of the proposed solutions. The steps are described in detail in the following sub-sections.

- *Training*. Network architecture & coefficients estimation.
- *Testing*. One cycle-ahead prediction of the signals.

Gated Recurrent Unit network

Starting from the structure of the Long-Short-Term-Memory NN (LSTM), Cho et al., 2014 coined a new formulation of Recurrent NN (RNN), the Gated Recurrent Unit network (GRU). It is based on a memory cell with a reset and an update gate. The reset gate has the role to decide how much past information is relevant, while the update gate provides the output. The GRU memory cell is the fundamental component for storing and transferring information along the network (Tuttle et al., 2021). Figure 1 below represents an example of a GRU memory cell, where α stands for the input activation function and σ stands for the sigmoid gate activation function.

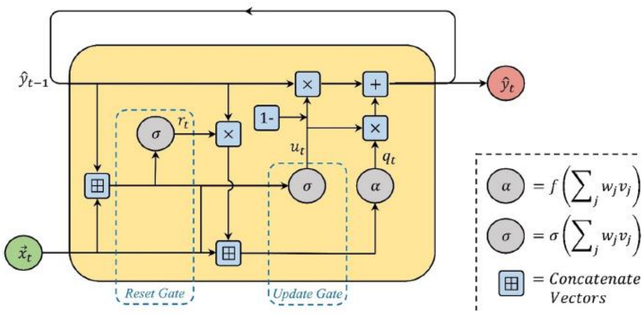

Fig. 1. Gated Recurrent Unit memory cell (Tuttle et al., 2021).

Tuttle et al., 2021 summarized the equations used for GRU cells as follows.

$$u_t = \sigma\left(w_u\left[x_t, \hat{y}_{t-1}\right] + b_u\right) \quad (1)$$

$$r_t = \sigma\left(w_r\left[x_t, \hat{y}_{t-1}\right] + b_r\right) \quad (2)$$

$$q_t = \alpha\left(w_t\left[x_t, r_t, \hat{y}_{t-1}\right] + b_y\right) \quad (3)$$

$$\hat{y}_t = (1 - u_t)\hat{y}_{t-1} + u_t q_t \quad (4)$$

Where x is the input time series, y is the cell output time series, w are the weights, b are the biases, u is associated to the update gates and r to the reset gates. A deeper discussion about the characteristics of the GRU networks can be found in Chung et al., 2014.

Training: During the training GRU receives M input signals, which spread over P channels. This phase is devoted to identifying the optimal GRU network. It is composed of two parallel steps. i) Bayesian optimization - for defining the optimal architecture of the network and estimating the hyperparameters. ii) Leave one out cross-validation (LOOCV) - for assessing the robustness of the network concerning the input dataset and estimate the coefficients. The coefficients characterize the network internal weights.

Bayesian optimization. Subsequent Eq. (5) reports the well-known Bayes' rule conditional probability (Shin et al., 2020):

$$p(v|D) = \frac{p(D|v)p(v)}{p(D)} \quad (5)$$

where $p(v|D)$ is the posterior distribution, $p(D|v)$ is the likelihood, v is an unobserved quantity and $p(v)$ is the prior distribution. The posterior probability requires knowledge about the prior distribution. Since the selection of the values of the next iteration depends on the results of the previous ones, the Bayesian approach demonstrates better performances than random sampling and grid search (Shin et al., 2020).

Consider a function $f(x)$ modelled as a Gaussian Process (GP) with mean function $m(x)$ and covariance function $k(xi, xj)$ (Brochu et al., 2010), and summarized as follows.

$$f(x) \sim GP\left(m(x), k\left(x_i, x_j\right)\right) \quad (6)$$

Given the prior observations $D_{1:t} = \{x_{1:t}, f(x_{1:t})\}$ we can compute the likelihood $P(D_{1:t}|f)$ and formulate a selection criterion for a new sampling point by minimizing the uncertainty function $f(x)$, as summarized in following Eq. (7).

$$x_{t+1} = argmin\, E\left(\left||f_{t+1}(x) - f(x)^+|\right| D_{1:t}\right) = argmin \int ||f_{t+1}(x) - f(x)^+|| P(f_{t+1}|D_{1:t}) df_{t+1} \quad (7)$$

Eq. (8), represents the improvement function proposed by Mockus et al., 1978, with respect to the expected improvement $f(x)^+$

$$I(x) = max\,(0, f_{t+1}(x) - f(x)^+)) \quad (8)$$

The results of the expected improvement are summarized in the following Eq. (9).

$$EI(x) = \begin{cases} (\mu(x) - x(x)^+))\Phi(Z) + \sigma(x)\Phi(Z) & if\ \sigma(x) > 0 \\ 0 \ \ if\ \sigma(x) = 0 \end{cases} \quad (9)$$

Where $Z = \frac{\mu(x) - f(x)^+}{\sigma(x)}$ and $\Phi(\cdot)$ is the standard normal distribution of a cumulative function.

In the present study, Bayesian optimization is used to minimize the RMSE of prediction, finding the optimal values of the following hyperparameters of the model:

- Initial learning rate
- Learning rate decay
- Number of hidden units
- Mini batch size.

Bayesian optimization is adopted to find the optimal values of GRU hyperparameters, ranging from a minimum to a maximum value. In practice, the GRU implementation consist in the following eigth-step algorithm:

1. Start with a set of random hyperparameters inside the range, associated to z index equal to 1. Then, *Leave One Out Cross Validation* is applied:
2. Start with j position index equal to 1
3. GRU network is trained over $M - 1$ signals
4. GRU network is validated on the remaining M-th signal
5. Calculate a first metric representing the prediction error of the globally optimized network: mean of the Root Mean Squared Error over the P channels $E_{1,GRU}(z,j) = \overline{RMSE_{z,j}} = \frac{(\sum_{p=1}^{P} RMSE_{p,z,j})}{P}$. Where $RMSE_{p,z,j}$ is the Root Mean Squared Error of the p-th channel of the M-th signal. j is the position index and z is the hyperparameters set index. Prediction error is the difference between the predicted M-th signal and the M-th target.
6. Increase the position index j by 1 and repeat the validation procedure until $j = M$. Every time a network is trained, initial and recurrent coefficients are randomized.
7. Select the minimum of the $E_{1,GRU}(z,:)$ row vector (length equal to $1\ x\ M$)
   a. $E_{2,GRU}(z) = \min\left(E_{1,GRU}(z,:)\right)$ (10)

  b. Thus, the z-th set of hyperparameter is associated to a single $E_{2,GRU}(z)$.

8. Bayesian optimization automatically changes the values of the hyperparameters and repeats the procedure along with the entire range (increasing z by 1), fixing a maximum optimization time.

The optimal set of hyperparameters should be found through the minimization of the $E_{2,GRU}$ vector. Moreover, since the LOOCV is applied inside the optimization procedure and the Eq. (10) selects the minimum of $E_{1,GRU}(z,:)$ over the M iterations, network coefficients are defined.

The training phase ends by freezing both hyperparameters and coefficients.

Testing. Testing phase has the role to test the predefined GRU network on new N signals.

Prediction. The prediction acts as a *one-signal-ahead* prediction (or "*one-cycle-ahead*" prediction). The prediction of the $M+n$ signal depends on

- Internal memory of the GRU network
- $M+n-1$ signals.

The adoption of a sliding window allows the inclusion of the $M+n-1$ signal in the prediction. The length l of the sliding window is fixed and equal to the signal length. Additional information about the usage of the sliding window in RNNs could be found in the work of Khandelwal et al., 2020.

3. Case Study

The proposed approach was tested and validated on a real production case study. It is composed of non-stationary signal patterns acquired from CNC embedded sensors during the milling of a vacuum pump.

The production process was performed on a MCM Clock Auto machine tool. All signals were acquired from machine embedded sensors through the PCL of the system. The acquisition frequency of the signals was about 20 Hz. Figure 2 shows a scheme of the axes of the CNC machine tool used in our case study and Table 1 shows a recap of the acquired signals.

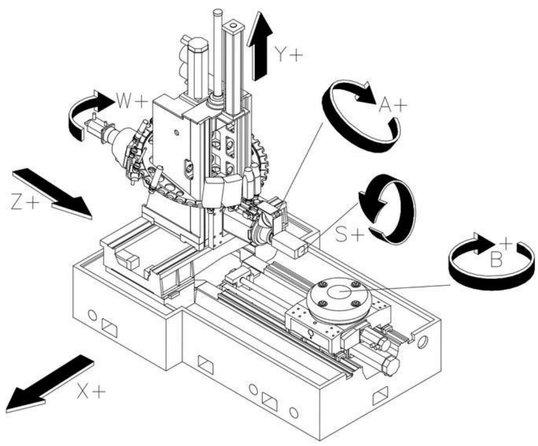

Fig. 2. MCM Clock Auto machine tool.

Table 1. Signals acquired during the manufacturing process.

ID channel	Signal
1	% Spindle power absorption
2	X axis current
3	Y axis current
4	Z axis current

Figure 3 shows the signals patterns divided per channel, acquired during the consecutive machining of 46 vacuum pumps by tool copy n.1 (i.e., the same cutting tool was used to mill 46 consecutive components before reaching its predetermined end of life based on total cutting time). These profiles are characterized by:

- Non-stationary mean, variance and autocorrelation structure over time
- Drift over time caused by tool wear
- High between-profiles variability.

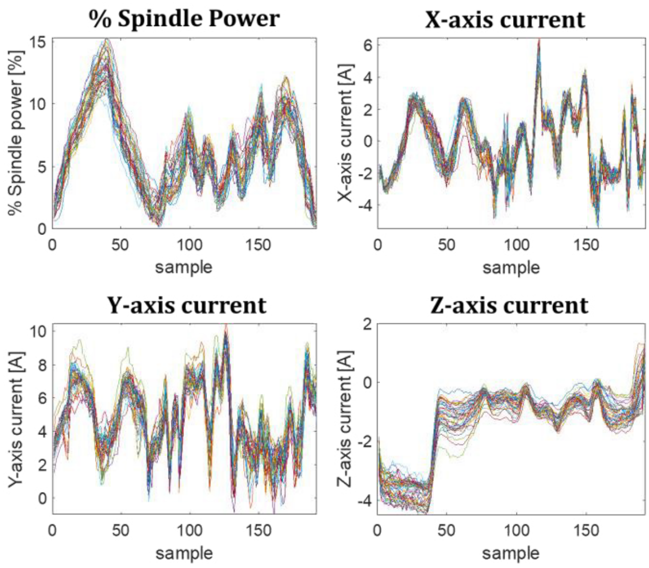

Fig. 3. Acquired signals divided per channel.

Figure 4 shows the drift induced by tool wear on the spindle power absorption and on the current absorbed by the Z-axis during the production of the 46 consecutive parts with the same tool. This drift caused a variation in terms of profile mean and profile shape along time, as the tool approached its end of life.

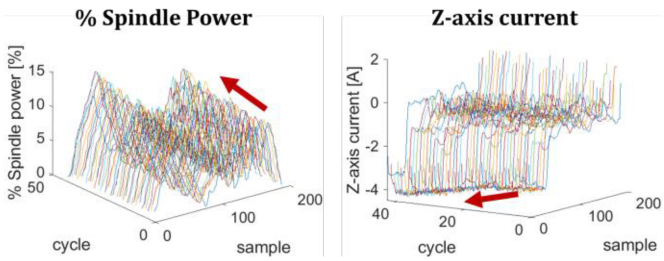

Fig. 4. Drift of the spindle power signals and z axis current.

4. Results

This section describes the results related to the training and testing phases.

Training. The training phase was composed of 46 cycles (one cycle consists of one milled part) acquired during the milling operations realized with one single copy of the tool (copy n.1). The overall duration of the training phase was limited to few minutes, leading the GRU suitable for real-time applications. Moreover, thanks to the specific nature of the GRU, the same network can be used to monitor and predict the signal patterns during the production of any other copy of the same product with different copies of the same tool. Between input and output layers, we adopted a GRU network with two GRU layers and a fully connected layer. The length l of the sliding window was fixed to 192 datapoints (equal to the duration of a milled part).

Following the guidelines provided in Mishkin et al., 2017 and Goodfellow et al., 2016, the listed Bayesian optimization ranges were used:

- Initial learning rate = [0.008 0.02]
- Learning rate decay = [0.2 0.5]
- Number of hidden units = [200 500]
- Mini batch size = [12 32]

The Bayesian optimization combined with a leave-one-out cross validation identified the optimal GRU with initial learning rate = 0,011, learning rate decay = 0,251, number of hidden units = 217 and mini batch size = 25. Network's coefficients are not reported.

Testing.
The test set was composed of 46 cycles acquired during the milling operations realized with tool copy n.2 (realizing the same operations of tool copy n.1).

The metric adopted to assess the performances of the network is the Root Mean Squared Error of the prediction, $RSME_p$, where p is the ID channel and the prediction error stands for the difference between the n-th predicted signal and the n-th target signal.

Figure 5 represents the 95% interval plot of the RMSE per channel. Since the network was globally optimized over the four data streams, the prediction accuracy of the next run process patterns changes according to the channel under evaluation. The result in Figure 5 reflects the dependence of the prediction accuracy on some intrinsic properties of the signal pattern. Indeed, the X- and Y-current signals are characterized by sudden changes of the first derivative of the signal over time, with high amplitude discontinuities. In correspondence of these features of the signal, the prediction error was systematically higher, as also shown in Figure 6. It is known that abrupt shifts introduce higher prediction uncertainty. However, the present study showed that such signal patterns had a small impact on the overall prediction performances, as the final prediction accuracy of the GRU was:

- ~ 96,25 % for % of the spindle power
- ~ 95 % for the X-axis current
- ~ 95 % for the Y-axis current
- ~ 95,25 % for the Z-axis current

The mean accuracy of the multichannel network over the four channels was ~ 95,4 %. The decrease of prediction accuracy observed in the two channels with abrupt shifts, namely X-axis and Y-axis current, was of 1 % only.

Moreover, the prediction uncertainty was stable over time, from initial cycles where the tool was in its early life state to the last few cycles, where the tool was approaching its end of life. This is another relevant aspect that makes the method suitable to monitor the process with a natural prediction uncertainty that remains stable over time and from run to run.

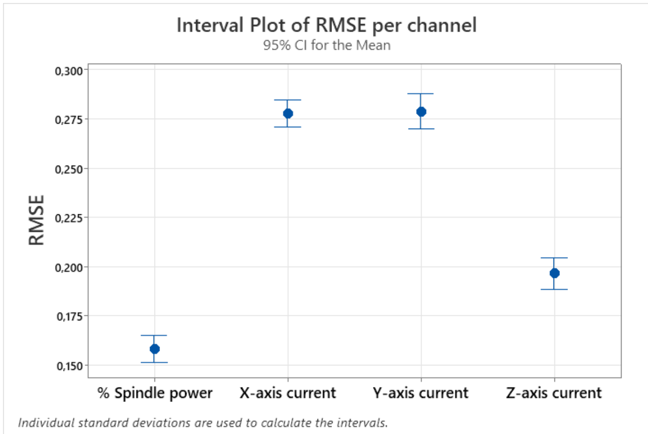

Fig. 5. Interval plot of the RMSE per channel.

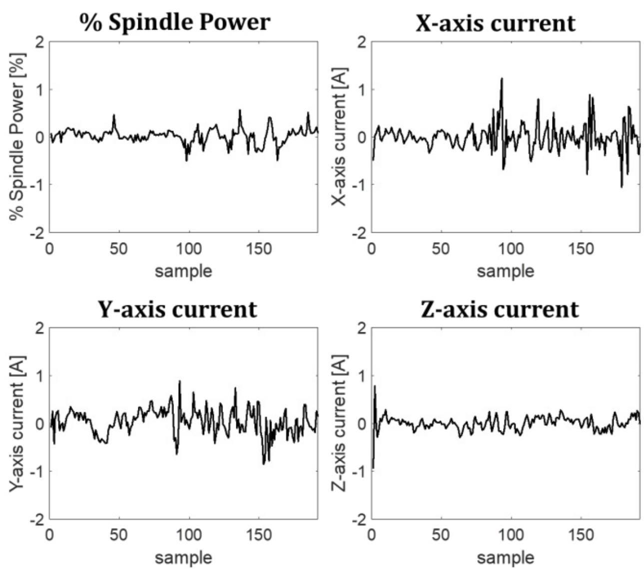

Fig. 6. Prediction errors for a reference cycle divided per channel.

5. Conclusions

In Industry 4.0, big data streams can be acquired from multiple sensors during every stage of the production chain. Thanks to proper data mining and modeling methodologies, value can be extracted from sensor data in the form of technological knowledge and novel smart manufacturing solutions. The increasing availability of complex data signals coming from multiple sources creates various challenges and stimulates the investigation and development of novel methodologies.

This study represents a preliminary exploration of the actual potentials and limitations of adaptive ANNs methods, like the GRU, in machining operations, being a first step of a broader study to identify their application field and their associated opportunities for implementation in real production environments.

In the current work, we evaluated the performances of a category of ANNs, i.e., the Gated-Recurrent-Unit-Network, in a multi-sensor prediction task of the next run process patterns. The proposed solution was implemented on a family of non-stationary signals, containing degradation and tool wear signatures. The GRU demonstrates high performances in modeling wild and complex patterns, reaching a prediction accuracy from 95% to 96% for all the channels. The capability of incorporating drifts and trends over time caused by the degradation of the tool enhances the capability of detecting abrupt changes in the signal, e.g., due to tool breakage, chipping, or external impacts.

Future extensions will be dedicated to incorporating the GRU prediction capabilities with an automated monitoring method to quickly detect rapid anomalous patterns, keeping into account the prediction uncertainty. Furthermore, a comparison between the proposed GRU network, other ANN schemes and parametric models would aim to highlight when and to what extent adaptive deep learning can be more convenient and effective than other methods and when, instead, such solutions may still represent the preferred approach.

References

Balsamo, V., Caggiano, A., Jemielniak, K., Kossakowska, J., Nejman, M., & Teti, R. (2016). Multi sensor signal processing for catastrophic tool failure detection in turning. *Procedia Cirp*, *41*, 939-944.

Brochu, E., Cora, V. M., & De Freitas, N. (2010). A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Denkena, B., Bergmann, B., & Stoppel, D. (2020). Reconstruction of process forces in a five-axis milling center with a LSTM neural network in comparison to a model-based approach. *Journal of manufacturing and materials processing*, *4*(3), 62.

Fogliazza, G., Arvedi, C., Spoto, C., Trappa, L., Garghetti, F., Grasso, M., & Colosimo, B. M. (2021). Fingerprint analysis for machine tool health condition monitoring. *IFAC-PapersOnLine*, *54*(1), 1212-1217.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Haili, W., Hua, S., Ming, C., & Dejin, H. (2003). On-line tool breakage monitoring in turning. *Journal of Materials Processing Technology*, *139*(1-3), 237-242.

Heinemann, R., & Hinduja, S. (2012). A new strategy for tool condition monitoring of small diameter twist drills in deep-hole drilling. *International Journal of Machine Tools and Manufacture*, *52*(1), 69-76.

Kang, G. S., Kim, S. G., Yang, G. D., Park, K. H., & Lee, D. Y. (2019). Tool chipping detection using peak period of spindle vibration during end-milling of inconel 718. *International Journal of Precision Engineering and Manufacturing*, *20*(11), 1851-1859.

Kerboua, A., Metatla, A., Kelaiaia, R., & Batouche, M. (2018). Real-time safety monitoring in the induction motor using deep hierarchic long short-term memory. *The International Journal of Advanced Manufacturing Technology*, *99*(9), 2245-2255.

Khandelwal, P., Konar, J., & Brahma, B. (2020, February). Training RNN and it's variants using sliding window technique. In *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)* (pp. 1-5). IEEE.

Lee, W. J., Wu, H., Yun, H., Kim, H., Jun, M. B., & Sutherland, J. W. (2019). Predictive maintenance of machine tool systems using artificial intelligence techniques applied to machine condition data. Procedia Cirp, 80, 506-511.

Lee, W. K., Abdullah, M. D., Ong, P., Abdullah, H., & Teo, W. K. (2021). Prediction of flank wear and surface roughness by recurrent neural network in turning process. *Journal of Advanced Manufacturing Technology (JAMT)*, *15*(1).

Li, E., Zhou, J., Yang, C., Wang, M., Li, Z., Zhang, H., & Jiang, T. (2022). CNN-GRU network-based force prediction approach for variable working condition milling clamping points of deformable parts. *The International Journal of Advanced Manufacturing Technology*, 1-21.

Li, X., Dong, S., & Yuan, Z. (1999). Discrete wavelet transform for tool breakage monitoring. *International Journal of Machine Tools and Manufacture*, *39*(12), 1935-1944.

Li, X., Liu, X., Yue, C., Liang, S. Y., & Wang, L. (2022). Systematic review on tool breakage monitoring techniques in machining operations. *International Journal of Machine Tools and Manufacture*, 103882.

Lo, S. P. (2002). The application of an ANFIS and grey system method in turning tool-failure detection. *The International Journal of Advanced Manufacturing Technology*, *19*(8), 564-572.

Marani, M., Zeinali, M., Songmene, V., & Mechefske, C. K. (2021). Tool wear prediction in high-speed turning of a steel alloy using long short-term memory modelling. *Measurement*, *177*, 109329.

Mishkin, D., Sergievskiy, N., & Matas, J. (2017). Systematic evaluation of convolution neural network advances on the imagenet. *Computer vision and image understanding*, *161*, 11-19

Moćkus, J., Tiesis, V., & Źilinskas, A. (1978). The Application of Bayesian Methods for Seeking the Extremum. Vol. 2.

Mohanraj, T., Shankar, S., Rajasekar, R., Sakthivel, N. R., & Pramanik, A. (2020). Tool condition monitoringtechniques in milling process—a review. Journal of Materials Research and Technology, 9(1), 1032-1042.

Nasir, V., & Sassani, F. (2021). A review on deep learning in machining and tool monitoring: methods, opportunities, and challenges. *The International Journal of Advanced Manufacturing Technology*, *115*(9), 2683-2709.

Pian, J., Tamanna, M. R., & Abdulmajid, A. U. (2021, May). Study on HS-RNN in Vibration Prediction of Mechanical Spindle. In *2021 33rd Chinese Control and Decision Conference (CCDC)* (pp. 630-633). IEEE.

Proteau, A., Tahan, A., & Thomas, M. (2019). Specific cutting energy: a physical measurement for representing tool wear. *The International Journal of Advanced Manufacturing Technology*, *103*(1), 101-110.

Serin, G., Sener, B., Ozbayoglu, A. M., & Unver, H. O. (2020). Review of tool condition monitoring in machining and opportunities for deep learning. *The International Journal of Advanced Manufacturing Technology*, *109*(3), 953-974.

Shin, S., Lee, Y., Kim, M., Park, J., Lee, S., & Min, K. (2020). Deep neural network model with Bayesian hyperparameter optimization for prediction of NOx at transient conditions in a diesel engine. *Engineering Applications of Artificial Intelligence*, *94*, 103761.

Tuttle, J. F., Blackburn, L. D., Andersson, K., & Powell, K. M. (2021). A systematic comparison of machine learning methods for modeling of dynamic processes applied to combustion emission rate modeling. *Applied Energy*, *292*, 116886.

Wang, J., Yan, J., Li, C., Gao, R. X., & Zhao, R. (2019). Deep heterogeneous GRU model for predictive analytics in smart manufacturing: Application to tool wear prediction. *Computers in Industry*, *111*, 1-14.

Wang, X., & Huang, Y. (2009). Optimized recurrent neural network-based tool wear modeling in hard turning. *Trans. NAMRI/SME*, *37*, 213-220.

Wong, S. Y., Chuah, J. H., & Yap, H. J. (2020). Technical data-driven tool condition monitoring challenges for CNC milling: A review. International Journal of Advanced Manufacturing Technology, 107(11-12), 4837-4857. doi:10.1007/s00170-020-05303-z

Wu, X., Li, J., Jin, Y., & Zheng, S. (2020). Modeling and analysis of tool wear prediction based on SVD and BiLSTM. *The International Journal of Advanced Manufacturing Technology*, *106*(9), 4391-4399.

Xu, S., Li, J., Liu, K., & Wu, L. (2019). A parallel GRU recurrent network model and its application to multi-channel time-varying signal classification. *IEEE Access*, *7*, 118739-118748.

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, *415*, 295-316.

Zhang, J., Zeng, Y., & Starly, B. (2021). Recurrent neural networks with long term temporal dependencies in machine tool wear diagnosis and prognosis. *SN Applied Sciences*, *3*(4), 1-13.

Zhao, R., Wang, J., Yan, R., & Mao, K. (2016, November). Machine health monitoring with LSTM networks. In *2016 10th international conference on sensing technology (ICST)* (pp. 1-6). IEEE.

Zhao, Z., Li, Y., Liu, C., & Gao, J. (2020). On-line part deformation prediction based on deep learning. *Journal of Intelligent Manufacturing*, *31*(3), 561-574.

Zhou, Y., & Xue, W. (2018). Review of tool condition monitoring methods in milling processes. The International Journal of Advanced Manufacturing Technology, 96(5), 2509-2523.

Zhou, J. T., Zhao, X., & Gao, J. (2019). Tool remaining useful life prediction method based on LSTM under variable working conditions. *The International Journal of Advanced Manufacturing Technology*, *104*(9), 4715-4726.