MM Science Journal | www.mmscience.eu
ISSN 1803-1269 (Print) | ISSN 1805-0476 (On-line)
Special Issue | CUTTING TOOLS 2024
1st International Conference on Cutting Tools
November 20-22, 2024, Trnava, Slovakia
DOI: 10.17973/MMSJ.2025_06_2025035

CUTTINGTOOLS2024-00015

ANN-BASED TOOL LIFE PREDICTION IN MICRO-MILLING USING AN EXPERIMENTAL DATASET FROM CENTRAL COMPOSITE DESIGN

Peter Balint Mezo[1]*, Marton Takacs[1], Adam Jacso[1]
[1]Budapest University of Technology and Economics, Department of Manufacturing Science and Engineering, Budapest, Hungary
*Corresponding author; e-mail: mezopeterbalint@edu.bme.hu

Abstract
ANN-based predictive models are becoming increasingly popular in machining technologies. Our research focused on the potential applications of AI-based predictive models in micro-milling, using a dataset from a cutting experiment designed to analyze the tool life. This dataset was previously utilized solely to develop traditional regression models, so our goal was to create an Artificial Neural Network (ANN) that could more efficiently predict tool life based on this data. Given the small sample size of the dataset, leave-one-out cross-validation (LOOCV) was employed during validation. By experimenting with various network structures—modifying the numbers of layers and neurons, and types of activation functions—we determined an appropriate ANN model to outperform the original regression models. Fully-connected feed-forward neural networks were trained using the Adam optimizer for up to 200, 500, and 1000 epochs. The model complexity was adjusted by varying the number of hidden layers from 1 to 10 in steps of one, and the number of neurons per layer from 5 to 50 in increments of five. Each model's evaluation was based on the Mean Absolute Error (MAE) and the Coefficient of Determination (R^2) and Standard Deviation of repeated training. The optimized ANN structure outperformed the second-order linear regression method in terms of both evaluation metrics and monotonicity analysis between the data points.

Keywords:
Micro-milling, ANN, Tool life prediction, Central Composite Design, Leave-One-Out Cross-Validation

1 INTRODUCTION

In recent years, the development of Artificial Intelligence has been picking up pace. This trend could be largely attributed to the continuous and vast expansion of accessible and diverse data. The trend is not limited to any specific field, it will gain traction in more and more fields in the future, including industry and manufacturing. Data-driven methods provided smart manufacturing with unprecedented opportunities to facilitate the transition toward Industry 4.0-based production. Machine learning and deep learning play a critical role in developing intelligent systems for descriptive, diagnostic, and predictive analytics for machine tools and process monitoring (Vahid Nasir and Farrokh Sassani 2021). Recent studies pay particular attention to tool wear, the determination of cutting forces, variations in surface roughness and other process parameters during machining (Jacso et al. 2023). Fluctuations in these machining-related parameters significantly influence dimensional accuracy and productivity. Moreover, excessive tool wear growth can potentially lead to tool breakage and subpar quality of the machined part (Pimenov et al. 2023).

Adizue et al. focused on predictive modeling for ultra-precision hard-turning, leveraging AI-driven approaches to optimize manufacturing processes. A full factorial experiment was conducted on AISI D2 hardened steel using a CBN cutting insert, varying cutting speed, feed and depth of cut, with surface roughness as the key response variable. To assess machine learning models, ANFIS, ANN, SVM, GPR type models were developed and validated for predictive accuracy, with ANFIS and ANN models proving to be more reliable in predicting surface roughness. The study identified optimal machining parameters to enhance surface quality and reduce production costs and highlighted the artificial intelligence's (AI) potential in real-time machining adjustments and cloud-based predictive maintenance (Adizue et al. 2023). Balázs et al. conducted micro-milling experiments on hardened Böhler M303 martensitic corrosion-resistant steel to analyze the effects of feed per tooth, depth of cut, and milling strategies on cutting force, vibration and characteristic frequencies. A 500 µm diameter, two-fluted coated micro-milling tool was used. An experimental-based cutting force model was developed, achieving high accuracy. They concluded that the cutting force components increased with higher cutting parameters; however, milling strategies influenced vibration differently. Up-milling minimized acceleration compared to down-milling, while groove-milling resulted in the smallest amplitude vibrations. Frequency analysis (FFT) revealed that characteristic frequencies mainly stem from process kinematics, with additional harmonics and environmental influences affecting lower-frequency components (Balázs et al. 2021).

One of the main difficulties in tool wear modeling is the acquisition of a sufficiently large dataset. This process is both time and cost-intensive since gaining deeper insights into wear mechanisms requires repeated machining experiments until the cutting tool is completely worn. Therefore, choosing an appropriate experimental design method is vital (Jiju Antony and Cahyono St 2022).

Considering three experimental factors (depth of cut, cutting speed, cutting feed) and four levels for each factor, a full factorial experiment consists of 64 trials. However, if the Taguchi method is applied for the design of experiments (DoE), the minimum required number of trials decreases to 16 (Hisam et al. 2024). At the same time, reducing the number of experiments means that there is a smaller data set available for creating predictive models.

To deal with a challenge caused by the small data set, the leave-one-out cross-validation (LOOCV) method can be used. In this method, the ANN is trained on the entire dataset except for one instance, which is then used to evaluate the model's performance. This process is repeated for all instances, ensuring every data point serves as the test set once. This method is advantageous when the data set is small: due to the training set being almost the entire dataset, therefore low bias will be present, as every data point is used for training and testing the model. However, it’s disadvantageous when larger data sets are present: due to the high resource cost in memory usage for every instance (Lumumba et al. 2024).

In model development, the goal is to achieve a well-fitting model, but it is important to avoid overfitting. When a model overfits, it not only learns the essential patterns but also picks up on peculiarities specific to the training data. As a result, the predictions can become highly inaccurate when applied to new data. The opposite of this is an underfitted model, where the model fails to learn the underlying relationships present in the training data set (Aliferis and Simon 2024). In the case of a small data set, the most significant concern is the risk of overfitting.

Pimenov et al. discussed modern approaches to tool condition monitoring within the framework of online trends across various machining operations. They aimed to study and analyze the application of traditional sensor systems and various AI methods for monitoring tool conditions, as well as to identify the advantages and disadvantages of these methods in modern manufacturing. To achieve this goal, they examined newly developed and applicable sensors for tool wear monitoring, along with the effective implementation of AI. They summarized and introduced a sensor systems used for tool wear monitoring and stated that these systems facilitate the automation and modeling of machining process parameters for primary cutting operations such as turning, milling, drilling, and grinding. They also covered modern artificial intelligence methods in the review, ranging from neural networks and fuzzy logic to genetic algorithms (Pimenov et al. 2023).

2 DEVELOPING AN ANN-BASED MODEL FOR TOOL WEAR PREDICTION

This paper presents a case study in which an ANN-based predictive model was developed for tool life prediction. Despite utilizing a small dataset, the model demonstrated excellent accuracy and reliability. This section outlines the dataset employed and its experimental background, the process of creating the predictive model, and also provides a comparative evaluation of its performance.

2.1 The basis of this work

As mentioned in the Introduction, generating data through targeted experiments is time and cost-consuming. Therefore, the base of this work on predictive modeling has been derived from an experiment performed by J.B Saedon, et al., who developed traditional regression models using self-made, valid micro-milling experiment data. The cutting experiments were conducted on a 3-axis CNC milling machine using a TiAlN-coated solid carbide end mill with a diameter of 0.5 mm. The tool had a cutting length of 1.0 mm, a helix angle of 30°, an edge radius of 5–7 µm, and a shank diameter of 6 mm. The workpiece material was AISI D2 cold work tool steel, machined in the form of a block measuring 90 mm in length, 20 mm in width, and 20 mm in thickness. The tools used and their edge rounding were measured with an Alicona-type measuring device, while tool clamping compensation was ensured using a Renishaw NC3 measuring system. The milling machine was prepared before the start of the experiment to avoid errors related to heat expansion. All key parameters were documented during the experiment, including cutting speed, feed rate, depth of cut, tool life, and the amount of material removed, which are presented in Table 1 (Saedon et al. 2012).

2.2 The data set used

The data set was created by an experimental series based on a Central Composite Design, as seen in Figure 1. In the experiment, micro-milling was performed on tool steel with a hardness of 62 Rockwell, using a coated tool with four edges and a diameter of 0.5 mm. For interpreting tool life, there are several different criteria. In the presented experiment series, tool life refers to the time elapsed during machining until the tool diameter decreased by 30 µm, which is approximately 6% relatively (Saedon et al. 2012).

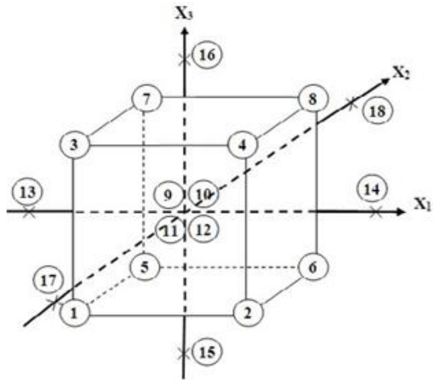

Fig. 1: The data set’s Central Composite Design from the experiment of J.B Saedon, et al. (Saedon et al. 2012)

Table 1: Micro-milling experimental data extract from J.B Saedon, et al. (Saedon et al. 2012)

Test No.	Cutting speed Vc (m/min)	Cutting Feed fc (μm/tooth)	Depth of Cut ap (μm)	Tool Life T (min)
1	20.00	1.00	15.00	4.71
2	50.00	1.00	15.00	1.42
3	20.00	2.00	15.00	2.77
4	50.00	2.00	15.00	1.10
5	20.00	1.00	55.00	3.14
6	50.00	1.00	55.00	0.95
7	20.00	2.00	55.00	1.19
8	50.00	2.00	55.00	0.55
9	32.00	1.42	29.00	1.90
10	32.00	1.42	29.00	1.70
11	32.00	1.42	29.00	1.70
12	32.00	1.42	29.00	1.90
13	14.17	1.42	29.00	2.64
14	68.30	1.42	29.00	0.68
15	32.00	0.78	29.00	2.16
16	32.00	2.5	29.00	0.92
17	32.00	1.42	10.00	3.12
18	32.00	1.42	85.00	1.56

Since there were differences in magnitude among the numerical values of the input data, all variables were min-max normalized between 0 and 1 (Shantal, Othman, and Bakar 2023) using the following formulas:

$$\hat{v}_{c,i} = \frac{v_{c,i} - \min(v_c)}{\max(v_c) - \min(v_c)} = \frac{v_{c,i} - 20}{68{,}2 - 20} \quad (1)$$

$$\hat{f}_{z,i} = \frac{f_{z,i} - \min(f_z)}{\max(f_z) - \min(f_z)} = \frac{f_{z,i} - 0{,}78}{2{,}5 - 0{,}78} \quad (2)$$

$$\hat{a}_{p,i} = \frac{a_{p,i} - \min(a_p)}{\max(a_p) - \min(a_p)} = \frac{a_{p,i} - 10}{85 - 10} \quad (3)$$

2.3 ANN models development and comparison

The ANN-based predictive models were implemented using Wolfram Mathematica 14.1. The input layer of the ANN had three neurons corresponding to the three input variables (cutting speed, feed rate, and depth of cut), and the output layer had one neuron for the tool life prediction. The activation function of the neurons was chosen to be hyperbolic tangent since during testing, the traditionally used RELU function showed a poorer fit. All networks were set to be Fully-Connected Feed-Forward neural networks using the Adam optimizer. The experiment consisted of three rounds, each with a maximum training round limit of 200, 500, and 1000 respectively. In each round, the number of layers varied from 1 to 10, increasing by one, and the number of neurons per layer varied from 5 to 50, with an increment of five. Therefore, 100 models were trained and tested each round.

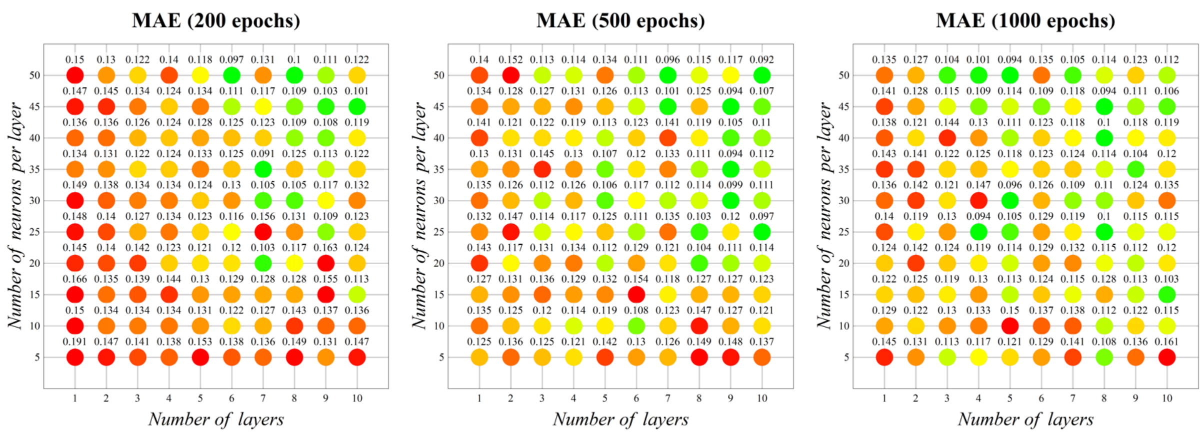

Fig. 2 The Maximum Absolute Error of the evaluated models

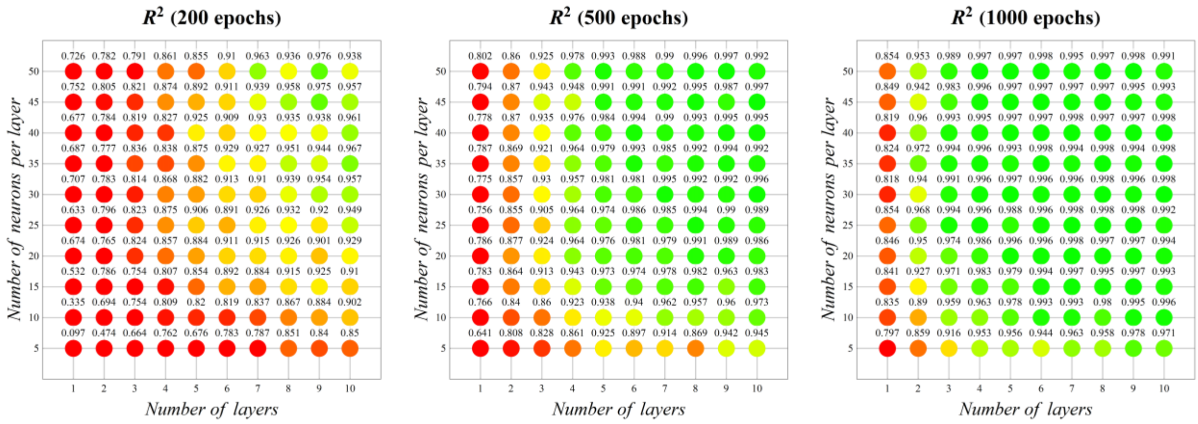

Fig. 3 The Coefficient of Determination of the evaluated models

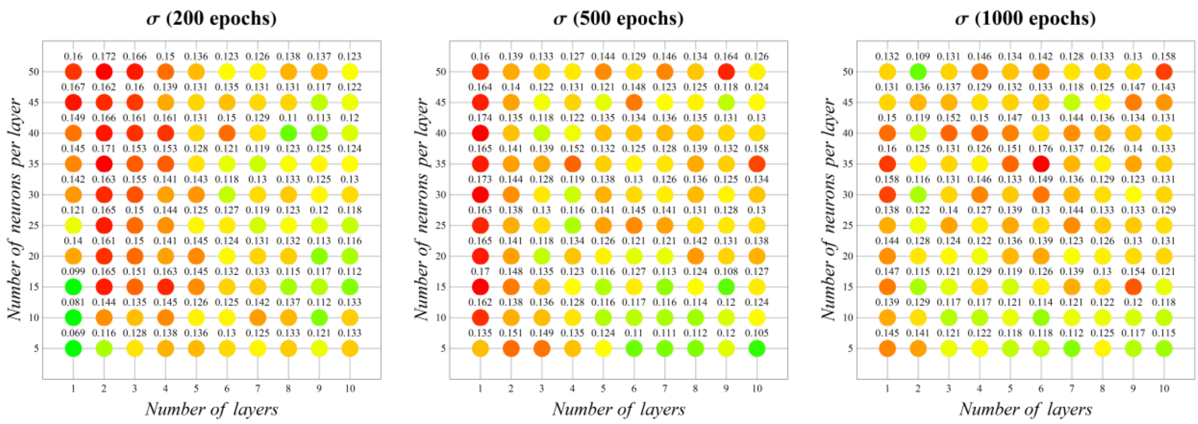

Fig. 4 The Standard Deviation of 10 training repetitions

To evaluate the algorithm's performance, the leave-one-out cross-validation (LOOCV) method was used and to measure the performance, two indicators were considered, namely the Mean Absolute Error (MAE, refer to Equation 4) and the Coefficient of Determination (R^2, refer to Equation 5). The former measures the average magnitude of absolute errors in a set of predictions and the latter is commonly used for regression tasks to assess the model quality. In Equation 4 and 5, the measured value is marked as y_i, the predicted value as $\hat{y}_i$, the mean of the actual values as $\bar{y}$ and the number of data points as n.

$$MAE = \frac{|\sum_{i=1}^{n}(y_i - \hat{y}_i)|}{n} \quad (4)$$

$$R^2 = 1 - \frac{SS_{red}}{SS_{ot}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (5)$$

In Equation 5, if $R^2 = 1$, then the model fits perfectly, if $R^2 = 0$, then the model performs no better than the mean value, and if $R^2 < 0$, the model is weaker predictor than the mean value. The MAE and R^2 results for each of the 100 network structures are shown in Figures 2 and 3 respectively, where the following trends can be noticed. The low MAE and R^2 values are concentrated on the left and the lower side, while the high MAE and R^2 values are in the upper right corner of the diagram. The first section indicates a lack of complexity that increases with more layers and neurons per layer while the latter area gets larger also by increasing the maximum epoch number. If we look at Figure 3, the graphs show that with more epochs, layers and neurons per layer, higher coefficient of determination can be achieved which indicates a better fitting. However, in Figure 4 where the standard deviation of the results of 10 training repetitions is shown, there are some sporadic red and orange dots even in the area declared for satisfactory models. The high variance indicates the unreliability of the network structures. In other words, using the LOOCV method, the appropriate size of the ANN structure to ensure sufficient complexity can be well limited based on MAE and R^2. Furthermore, uncertain network structures can be filtered out by repeating the training several times and analyzing the variance between the results.

However, checking for overfitting requires further examinations. To address this, we conducted a visual inspection of the predictive models' behavior across various structures between measurement points, while systematically fixing one input variable at a time, namely the axial depth of cut. Overfitting is indicated when the predictive model shows inconsistent behavior between measurement points (refer to Figure 5). We experienced this behavior in all cases with larger network structures. Based on this analysis, the model with 5 layers and 20 neurons per layer using 500 epochs was designated as the optimal network with simple structure, high accuracy and good reliability (refer to Figure 6).

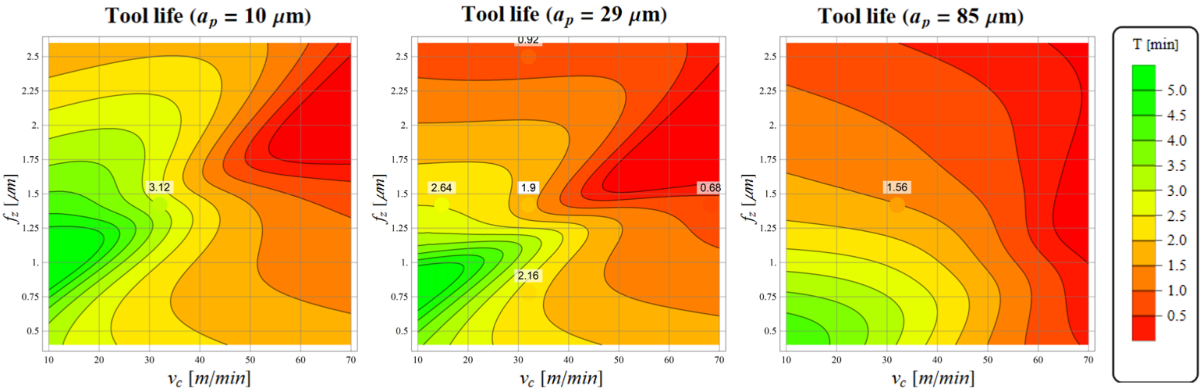

Fig. 5 Inconsistent behavior of the predictive model when overfitting between measurement points

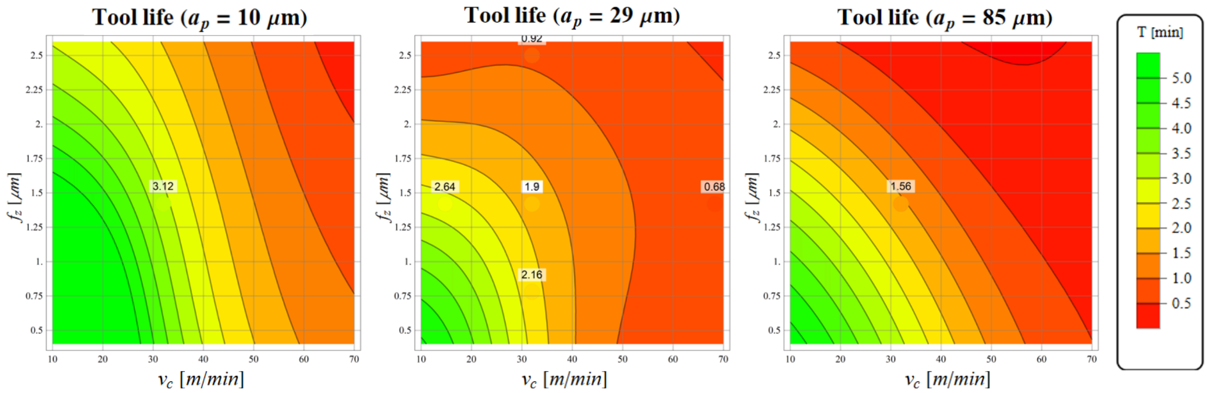

Fig. 6 The consistent behavior of the optimal model between measurement points

To summarize the key steps presented in this section, the models were trained using the Leave-One-Out Cross-Validation (LOOCV) method to reduce bias in the evaluation process. Models demonstrating favorable performance in terms of Mean Absolute Error (MAE) and Coefficient of Determination (R^2) were shortlisted; however, inconsistencies between individual data points were still observed. To investigate these regions of uncertainty, 27 intermediate data points were introduced. Each neural network configuration was trained ten times on these points, and their performance was evaluated based on the standard deviation of the predictions.

The optimal model, selected according to all three performance indicators (MAE, R^2, and standard deviation), was a Feed-Forward Fully-Connected Artificial Neural Network (ANN) with five hidden layers, 20 neurons per layer, and 500 training epochs.

2.4 Comparison with traditional regression methods

The optimized model was subsequently compared to two analytical models developed by J. B. Saedon et al., namely a first-order and a second-order regression model based on an extended form of Taylor's tool life equation. The comparison was conducted using the Coefficient of Determination as the primary evaluation metric. For the first-order model, this value was $R^2 = 0.882$, for the second-order model, it was $R^2 = 0.984$ (Saedon et al. 2012). These two models were surpassed in Coefficient of Determination by the optimal ANN-based model, which had a $R^2 = 0.989$ (refer to Figure 7). Furthermore, the ANN-based model did not exhibit monotonicity inconsistencies in tool life prediction, as observed in second-order linear regression (refer to Figure 6 and 8). So, with an ideal structure and training method, the ANN-based prediction can be considered more favorable in terms of both accuracy and reliability.

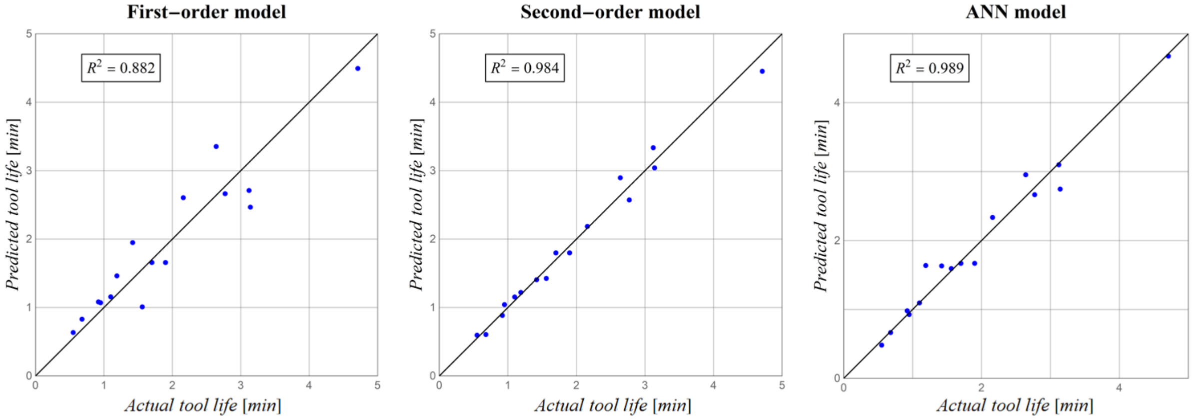

Fig. 7 The Coefficient of Determination of the different predictive models

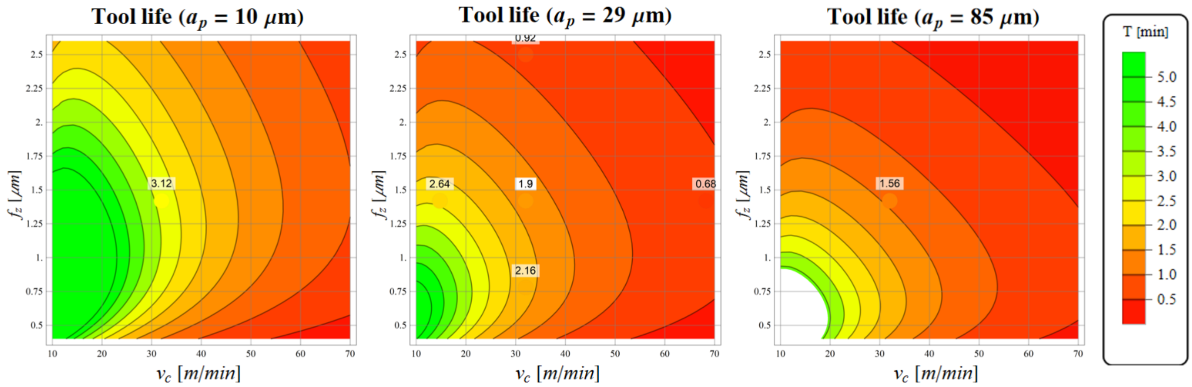

Fig. 8 The inconsistent behavior of the second-order regression between measurement points

3 SUMMARY

In this paper, an experimental dataset obtained from Central Composite Design (CCD) was used as the basis for developing Feed-Forward Fully-Connected Artificial Neural Network (ANN) models for tool life prediction. Three batches of training were performed with 200, 500, and 1000 epochs, respectively. In each epoch, the number of hidden layers varied from 1 to 10 (in increments of one), and the number of neurons per layer ranged from 5 to 50 (in increments of five). In total, 100 models were trained. These models were compared based on Mean Absolute Error (MAE) and Coefficient of Determination (R^2) to assess point-wise accuracy, as well as based on the standard deviation of repeated trainings. Finally, the optimal network was selected based on its overall performance across all three evaluation metrics.

In summary, the following statements can be made:

- increasing the number of layers, neurons and epochs generally result in a better Mean Absolute Error and Coefficient of Determination but carry the risk of overfitting
- the size of ANN structure that ensures the required complexity can be well limited by using the LOOCV method
- testing with intermediate points can show the phenomenon of overfitting
- in the current experiment, the Feed-Forward Fully-Connected ANN structure with 5 Layers, 20 Neurons per Layer and 500 epochs of training were more accurate and expressive compared to traditional regression models.

Further development opportunities could arise with more experiments and the involvement of other ANN structures, as more data is needed to verify the reliability of the model and to be a basis for further expansions. In cutting tool life analysis, even a full factorial experiment would not yield a large dataset, so LOOCV is expected to be an effective validation method in such cases as well. Although this work focused solely on the tools, the scope could be extended in the future to include other parameters such as vibration level or surface quality.

4 ACKNOWLEDGMENTS

The project was supported by the Doctoral Excellence Fellowship Programme (DCEP), the University Research Scholarship Programme (EKÖP) funded by the National Research Development and Innovation Fund of the Ministry of Culture and Innovation and the Budapest University of Technology and Economics, as well as by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, under the TKP2021-NVA funding scheme (project no. TKP-6-6/PALY-2021).

5 REFERENCES

Aliferis, Constantin, and Gyorgy Simon. 2024. 'Overfitting, Underfitting and General Model Overconfidence and Under-Performance Pitfalls and Best Practices in Machine Learning and AI'. In *Artificial Intelligence and Machine Learning in Health Care and Medical Sciences: Best Practices and Pitfalls*, edited by Gyorgy J. Simon and Constantin Aliferis, 477–524. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-39355-6_10.

Balázs, Barnabás Zoltán, Norbert Geier, Csongor Pereszlai, Dániel István Poór, and Márton Takács. 2021. 'Analysis of Cutting Force and Vibration at Micro-Milling of a Hardened Steel'. *Procedia CIRP*, 14th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 15-17 July 2020, 99 (January):177–82. https://doi.org/10.1016/j.procir.2021.03.025.

Hisam, Mohammed Wamique, Amir Ahmad Dar, Mohammed Osman Elrasheed, Mohammad Shahfaraz Khan, Rajat Gera, and Imran Azad. 2024. 'The Versatility of the Taguchi Method: Optimizing Experiments Across Diverse Disciplines'. *Journal of Statistical Theory and Applications* 23 (4): 365–89. https://doi.org/10.1007/s44199-024-00093-9.

Jacso, Adam, Tibor Szalay, Basant Singh Sikarwar, Rakesh Kumar Phanden, Rajeev Kumar Singh, and Janakarajan Ramkumar. 2023. 'Investigation of Conventional and ANN-Based Feed Rate Scheduling Methods in Trochoidal Milling with Cutting Force and Acceleration Constraints'. *The International Journal of Advanced Manufacturing Technology* 127 (1): 487–506. https://doi.org/10.1007/s00170-023-11506-x.

Jiju Antony and Cahyono St. 2022. *Design of Experiments for Engineers and Scientists SECOND EDITION*. 2022nd ed. Elsevier Ltd. https://www.researchgate.net/publication/362080058_Design_of_Experiments_for_Engineers_and_Scientists_SECOND_EDITION.

Lumumba, Victor Wandera, Dennis Kiprotich, Mary Lemasulani Mpaine, Njoka Grace Makena, and Musyimi Daniel Kavita. 2024. 'Comparative Analysis of Cross-Validation Techniques: LOOCV, K-Folds Cross-Validation, and Repeated K-Folds Cross-Validation in Machine Learning Models'. *American Journal of Theoretical and Applied Statistics* 13 (5): 127–37. https://doi.org/10.11648/j.ajtas.20241305.13.

Pimenov, Danil Yu, Andres Bustillo, Szymon Wojciechowski, Vishal S. Sharma, Munish K. Gupta, and Mustafa Kuntoğlu. 2023. 'Artificial Intelligence Systems for Tool Condition Monitoring in Machining: Analysis and Critical Review'. *Journal of Intelligent Manufacturing* 34 (5): 2079–2121. https://doi.org/10.1007/s10845-022-01923-2.

Saedon, J. B., S. L. Soo, D. K. Aspinwall, A. Barnacle, and N. H. Saad. 2012. 'Prediction and Optimization of Tool Life in Micromilling AISI D2 (~62 HRC) Hardened Steel'. *Procedia Engineering*, International Symposium on Robotics and Intelligent Sensors 2012 (IRIS 2012), 41 (January):1674–83. https://doi.org/10.1016/j.proeng.2012.07.367.

Shantal, Mohammed, Zalinda Othman, and Azuraliza Abu Bakar. 2023. 'A Novel Approach for Data Feature Weighting Using Correlation Coefficients and Min–Max Normalization'. *Symmetry* 15 (12): 2185. https://doi.org/10.3390/sym15122185.

Vahid Nasir and Farrokh Sassani. 2021. 'A review on deep learning in machining and tool monitoring: methods, opportunities, and challenges'. *The International Journal of Advanced Manufacturing Technology (2021) 115:2683–2709*. https://doi.org/10.1007/s00170-021-07325-7.