



Predictive modeling of surface roughness and resultant force in CNC turning of AISI H13 using optimized artificial neural networks

Sunardi Sunardi*, Ananda Nur Daffa Zain

Department of Mechanical Engineering, Universitas Muhammadiyah Yogyakarta, Yogyakarta 55183, Indonesia
*Corresponding author: sunardi@umy.ac.id

Abstract

Artificial Neural Networks (ANNs) have gained increasing attention as effective tools for modeling nonlinear and multivariate relationships in complex manufacturing processes, where conventional predictive approaches often exhibit limited accuracy. In this study, an ANN-based predictive framework was developed to estimate surface roughness (Ra) and resultant force (F) in CNC turning of hardened AISI H13 steel. The framework was constructed using an experimental dataset comprising 324 machining records, with cutting speed (vc), feed rate (f), and depth of cut (ap) as input parameters, all normalized using the Min-Max scaling method to ensure stable and efficient model training. To identify the optimal training configuration, eight optimization algorithms: Adam, RMSprop, Nadam, Adagrad, Adadelta, Adamax, FTRL, and Stochastic Gradient Descent (SGD) are systematically evaluated, and Nadam was selected as the most effective optimizer with a learning rate of 0.0001 and a batch size of 16. Two dedicated feed forward ANN models are designed separately for Ra and F prediction and validated using the Leave-One-Out Cross-Validation (LOOCV) technique to enhance generalization and minimize overfitting. The resulting models achieved excellent predictive accuracy for resultant force ($R^2 = 0.9939$, MAE = 4.3313 N, RMSE = 7.5955 N) and moderate accuracy for surface roughness ($R^2 = 0.6454$, MAE = 0.1440 μm , RMSE = 0.1960 μm). These results demonstrate that the proposed ANN-based framework provides a reliable decision-support tool for process optimization, monitoring, and surface quality control in high-performance machining environments.

Keywords:

Artificial neural network, CNC turning, surface roughness, resultant force, predictive modeling.

1 Introduction

The increasing demand for high-precision and durable components in modern manufacturing has intensified the machining of hardened materials across key industries, including mold making, automotive, and aerospace [1, 2]. Among these materials, AISI H13 tool steel is widely recognized for its outstanding mechanical strength, thermal resistance, and dimensional stability under high mechanical loads, making it ideal for high performance applications [3]. Nevertheless, machining hardened steels such as AISI H13 presents significant technical challenges, including elevated cutting forces, accelerated tool wear, and unstable surface finishes [4, 5]. These complications can negatively affect tool life, reduce productivity, and compromise part accuracy. Despite extensive research on machining response prediction, a clear and specific knowledge gap remains: surface roughness and cutting force are

predominantly modelled as independent outputs, limiting the ability of existing approaches to represent their coupled and nonlinear interactions during machining. Therefore, the objective of this study is to develop a unified dual-output ANN-based predictive framework capable of simultaneously estimating surface roughness (Ra) and resultant force (F) in CNC turning of hardened AISI H13 steel with high accuracy and strong generalization performance.

Historically, empirical equations and regression-based models have been utilized to delineate the relationships between machining parameters and output responses. While such approaches may yield acceptable predictions under simplified or linear conditions, they often struggle to capture the inherently nonlinear and multivariate behaviour of real-world machining operations. As machining systems become increasingly complex, these limitations have motivated a transition toward data-driven and computational intelligence-based modeling strategies. The advent of computational intelligence has introduced machine learning (ML) techniques as powerful tools for modeling these complex relationships [6, 7]. In particular, artificial neural networks (ANNs) have garnered significant interest due to their demonstrated ability to process high-dimensional data and uncover latent patterns between input and output variables that may not be readily apparent through conventional means [8]. This capacity to learn underlying relationships without the need for predefined mathematical formulations represents a key advantage of neural networks, distinguishing them from many traditional data modeling algorithms.

Despite their strengths, most existing ANN-based studies in machining focus on single-output prediction, either surface roughness or cutting force. Integrated frameworks capable of simultaneously predicting multiple machining responses remain limited. Additionally, many previous studies adopt suboptimal ANN architectures and basic validation techniques, increasing the risk of overfitting and reducing generalizability, especially when working with relatively small datasets [9]. These methodological limitations highlight the necessity for a more robust and comprehensive ANN approach that explicitly supports multi-output prediction while maintaining high predictive accuracy and strong generalization capability. To address these gaps, this study proposes a dual-output predictive modeling framework based on artificial neural networks for estimating both surface roughness, specifically root-mean-square roughness (Ra) and resultant force (F) in computer numerical control (CNC) turning of hardened AISI H13 steel. The model was trained using the dataset referred to as EXP1, which comprises 324 machining records derived from a full factorial experimental design. The input features cutting speed (vc), feed rate (f), and depth of cut (ap) were normalized using the Min-Max scaling method to ensure balanced data representation and facilitate effective training [10].

Model development was conducted using Python, with TensorFlow and Keras libraries facilitating the implementation of deep learning frameworks. Four different ANN architectures were systematically evaluated in combination with eight widely used optimization algorithms: Adam, RMSprop, Nadam, Adagrad, Adadelta, Adamax, FTRL, and Stochastic Gradient Descent (SGD) each tested using a fixed learning rate of 0.0001. The highest predictive accuracy for both surface roughness and resultant force was achieved using distinct ANN configurations, each trained with a batch size of 16 [11]. To further improve generalization and minimize the risk of overfitting, the models were validated using the Leave-One-Out Cross-Validation (LOOCV) approach [12]. The novelty of this research lies in the development of a scalable dual-output ANN framework capable of concurrently predicting two critical machining responses surface roughness and resultant force under consistent experimental conditions. In contrast to prior studies that predominantly focus on single-output modeling and limited hyperparameter exploration, this work presents a comprehensive comparative analysis of multiple ANN architectures and optimization algorithms. Moreover, the application of Leave-One-Out Cross-Validation (LOOCV) enhances the robustness and generalization capability of the proposed models, thereby

establishing a reliable foundation for data-driven optimization in intelligent machining and advanced manufacturing systems.

2 Research methods

This study adopts a systematic experimental methodology combined with quantitative modeling to develop predictive models for CNC turning processes. The main objective is to estimate two critical machining responses: surface roughness (Ra) and resultant force (F). The resultant force reflects the total cutting force exerted on the tool, indicating cutting stability and mechanical stress during machining. Both output variables are modeled as functions of three fundamental process parameters: cutting speed (vc), feed rate (f), and depth of cut (ap) [13].

The modeling workflow is structured into five sequential stages: dataset acquisition and preprocessing, exploratory data analysis (EDA), ANN modeling, model tuning, and performance evaluation. This workflow is visually presented in the block diagram shown in Fig. 1, enabling a clearer and more intuitive understanding of the overall modeling procedure. All computational procedures were executed in Python, leveraging the flexibility and efficiency of TensorFlow and Keras in deep learning implementations. In order to ensure robust generalization and prevent overfitting, Leave-One-Out Cross-Validation (LOOCV) was employed. A systematic comparison was conducted among eight optimization algorithms: Adam, RMSprop, Nadam, Adagrad, Adadelat, Adamax, FTRL, and SGD, all operating within a fixed learning rate of 0.0001. To identify the most suitable configuration for each prediction target, several batch sizes were evaluated, thus optimizing both model stability and accuracy [14, 15].

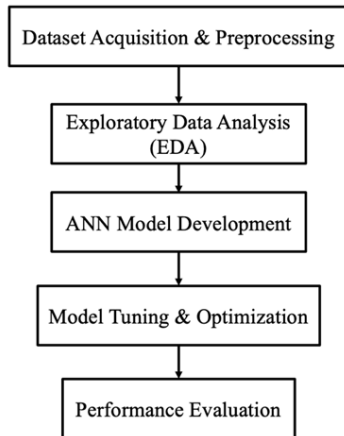


Fig. 1. Five-stage workflow of the proposed ANN-based predictive modeling framework for CNC turning

2.1 Dataset

The dataset applied in this investigation was retrieved from a publicly accessible source on Kaggle titled “CNC Turning: Roughness, Forces, and Tool Wear”. Specifically, this study focused on the EXP1 portion of the dataset, which contains 324 observations derived from dry CNC turning trials performed on AISI H13 hardened steel. All experimental procedures were conducted under controlled laboratory conditions using new carbide inserts to ensure consistency and repeatability.

Each record in the dataset consists of three independent machining variables: depth of cut (ap), cutting speed (vc), and feed rate (f). The associated output responses include surface roughness (Ra), measured in (μm) and resultant force (F), measured in (N), where the latter indicates the total mechanical load experienced by the cutting tool during the machining process. The experimental scheme employed a full factorial design of experiments (DoE), involving three discrete levels for each input parameter. This design ensured systematic exploration of the process parameter space, thus enhancing the dataset’s diversity and increasing its suitability for developing predictive models that can capture complex nonlinear input–output behaviors in CNC turning operations.

Moreover, the EXP1 dataset is particularly relevant to the objectives of this study because it simultaneously provides experimentally measured surface roughness and cutting force responses under identical machining conditions. This dual-response characteristic enables the development and validation of ANN-based predictive models that can accurately capture the coupled and nonlinear relationships between machining parameters and multiple output variables, thereby directly supporting the proposed dual-output modeling framework.

2.2 Exploratory data analysis (EDA)

Exploratory Data Analysis (EDA) serves as a foundational step in preparing data for predictive modeling, allowing for detailed assessment of its structure, completeness, and internal consistency prior to training [16, 17]. In this study, a thorough EDA was conducted on 324 data samples derived from dry CNC turning experiments involving AISI H13 steel, aiming to verify the dataset’s suitability for training artificial neural network (ANN) models. The initial statistical analysis entailed the calculation of descriptive statistics, with the purpose of characterizing the central tendency and variability of the three aforementioned input parameters: namely, depth of cut (ap), cutting speed (vc), and feed rate (f). Consequently, statistical descriptors such as mean, median, standard deviation, and interquartile range were employed to examine the distribution properties of the target variables: surface roughness (Ra) and resultant force (F).

To investigate linear associations between input and output features, a correlation analysis was carried out using the Pearson coefficient [18, 19]. This enabled quantification of the direction and strength of inter-variable relationships and provided key insights for refining feature selection and ANN architecture. Additionally, the correlation matrix helped detect any potential multicollinearity, which can impair model performance and generalization.

For numerical stability and consistent scaling during training, all input variables were normalized to a common range of 0 to 1 using the Min-Max normalization technique, as defined by the following equation:

$$Z_{norm} = \frac{Z - Z_{min}}{Z_{max} - Z_{min}} \quad (1)$$

Where Z is the original input value, Z_{min} and Z_{max} represent the minimum and maximum values observed within the dataset, and Z_{norm} denotes the normalized value. This transformation mitigates the potential for bias introduced by disparate feature scales and enhances convergence behavior during model training [13].

Following normalization, visual inspections were performed through histogram plots and statistical checks to evaluate the balance and integrity of data distributions. The dataset was confirmed to be free from anomalies such as extreme outliers and missing values, thereby validating its readiness for training. Finally, the data were divided into training and testing subsets to facilitate unbiased performance evaluation of the predictive models. This thorough EDA process ensured the statistical soundness and reliability of the dataset, thus supporting the development of accurate and reproducible ANN-based prediction frameworks.

2.3 Artificial neural network (ANN) modeling

Artificial neural networks (ANNs) were utilized in this work to represent the nonlinear and multivariable interactions between cutting parameters and machining responses in CNC turning processes. Given the intrinsic complexity in the relationships among depth of cut (ap), cutting speed (vc), and feed rate (f), ANNs offer a robust data-driven modeling approach that circumvents the need for explicit analytical formulations. Two independent neural networks were designed to predict surface roughness (Ra) and resultant force (F), each tailored for a specific output variable.

Both models were structured as multilayer perceptrons (MLPs) comprising fully connected layers. The input layer in each network received three features corresponding to the machining parameters (ap , vc , and f), followed by a series of hidden layers and a single output neuron for the prediction target. For estimating F , the model included five hidden layers configured with [256, 128, 64, 32, 16] neurons, whereas the Ra prediction model used a slightly reduced architecture of four hidden layers with [128, 64, 32, 16] neurons [11].

All hidden units employed the rectified linear unit (ReLU) activation function to facilitate efficient learning and avoid gradient vanishing issues [20]. The output layer adopted a linear activation function, appropriate for continuous variable regression tasks. Training was conducted via the backpropagation algorithm, with the Nadam optimizer selected due to its integration of Nesterov momentum, which enhances convergence speed and stability. A fixed learning rate of 0.0001 was maintained, and early stopping criteria were applied based on validation loss to prevent overfitting. After empirical tuning, a batch size of 16 was chosen for both models, offering a stable balance between training dynamics and predictive accuracy.

Model development was implemented in Python using TensorFlow and Keras frameworks, with GPU support enabled to accelerate training. All input features were normalized using the Min-Max scaling method within the range [0, 1] to ensure consistent numeric behavior across features. To robustly assess generalization performance, Leave-One-Out Cross-Validation (LOOCV) was adopted, which is particularly well-suited for smaller datasets such as the 324 experimental samples used in this study.

The architecture of the ANN model with five hidden layers for predicting resultant force is illustrated in Fig. 2.

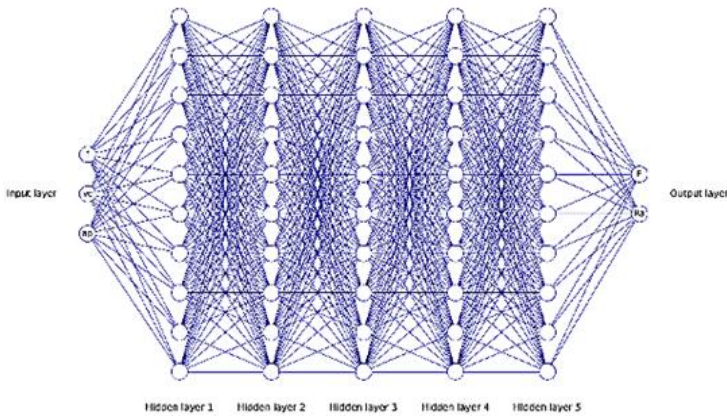


Fig. 2. Schematic Representation of the ANN Architecture Comprising Five Hidden Layers Used for Resultant Force Prediction

2.4 Hyperparameter tuning

To improve the predictive precision and generalization strength of the artificial neural network (ANN) models for forecasting surface roughness (Ra) and resultant force (F) in CNC turning of AISI H13 steel, an extensive process of hyperparameter optimization was undertaken. This involved systematic testing and refinement of several key configurations, including network architecture (depth and number of neurons), learning rate, batch size, and choice of optimization method. The tuning process aimed to ensure convergence stability while maximizing model accuracy for both output variables.

Eight popular optimizers were evaluated to determine their influence on training efficiency and prediction performance: Adam, RMSprop, Nadam, Adagrad, Adadelta, Adamax, FTRL, and stochastic gradient descent (SGD). Among these, Nadam, a hybrid of Adam's adaptive updates and Nesterov momentum, consistently yielded superior performance, evidenced by smoother learning curves and lower validation errors [21]. As a result, Nadam was adopted as the final optimizer for both ANN configurations.

The learning rate was tested at three different magnitudes: 0.01, 0.001, and 0.0001. The smallest rate (0.0001) facilitated the most stable training progression, reducing fluctuations and premature convergence [22]. Furthermore, several batch size values were compared to determine the optimal training behavior. A batch size of 16 offered the best compromise between convergence speed and generalization accuracy, proving effective for both Ra and F prediction tasks.

To verify the reliability of this configuration, Leave-One-Out Cross-Validation (LOOCV) was applied, ensuring performance robustness, particularly important given the limited dataset of 324 observations. The final model evaluation was based on three regression performance metrics: mean absolute error (MAE), root-mean-square error ($RMSE$), and the coefficient of determination (R^2). The MAE provides insight into the average prediction deviation and is mathematically expressed as:

$$MAE = \frac{1}{P} \sum_{j=1}^P |D_j - E_j| \quad (2)$$

Where P denotes the total number of observations, D_j is the actual output, and E_j is the predicted value for the J^{th} instance. A lower MAE indicates that the model's predictions are, on average, closer to the actual measurements.

$RMSE$ offers a more sensitive evaluation by penalizing larger errors more heavily than MAE . It computes the square root of the average of the squared differences between actual and predicted values, and is given by:

$$RMSE = \sqrt{\frac{1}{P} \sum_{j=1}^P (D_j - E_j)^2} \quad (3)$$

A reduced root-mean-square error ($RMSE$) value reflects higher predictive accuracy, especially in applications where minimizing large prediction errors is critical [23]. Meanwhile, the coefficient of determination (R^2) serves as an indicator of the model's goodness-of-fit by quantifying the proportion of variance in the observed data that is explained by the predictions. Mathematically, R^2 is defined as:

$$R^2 = 1 - \frac{\sum_{j=1}^P (D_j - E_j)^2}{\sum_{j=1}^P (D_j - \bar{D})^2} \quad (4)$$

where \bar{D} is the mean of the actual output values. An R^2 value close to 1 indicates strong explanatory power and a highly accurate fit.

To summarize the tuning process, the selected configuration comprising a deep ANN structure, a finely adjusted learning rate of 0.0001, a batch size of 16, and the use of the Nadam optimizer resulted in predictive models with excellent stability and performance [24]. This setup enabled the construction of a reliable data-driven system for accurately estimating machining outcomes in CNC turning of AISI H13 steel, thereby reducing reliance on manual trial-and-error methods and reinforcing data-centric process planning.

2.5 Model validation

Model validation is essential for ensuring that predictive frameworks are both reliable and capable of generalizing to unseen data, particularly in manufacturing studies that often rely on limited experimental observations. In the present work, the Leave-One-Out Cross-Validation (LOOCV) technique was implemented to thoroughly evaluate the predictive capability of artificial neural network (ANN) models developed for estimating surface roughness (Ra) and resultant force (F) during dry CNC turning of AISI H13 tool steel.

LOOCV was deemed appropriate due to its effectiveness in handling small datasets while yielding low-bias estimates of generalization error. This validation method involves isolating one sample at a time from the dataset to serve as a test case, with the remaining records used for model training. Given the total of 324 experimental entries, this procedure resulted in 324 individual training-validation cycles.

During each iteration, one observation was temporarily excluded from the training process and used as a validation point. The ANN model was retrained on the other 323 records, and the prediction was generated for the excluded sample. Once all data points had undergone this process, the performance metrics from each iteration were aggregated and averaged to derive an overall estimation of the model's predictive performance.

This exhaustive validation approach maximizes data utilization while minimizing variance introduced by arbitrary data splits. It is especially valuable in domains where collecting additional experimental data is resource-intensive or constrained by practical limitations.

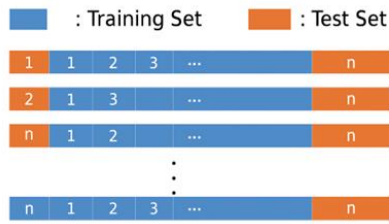


Fig. 3. Illustration of the Leave-One-Out Cross-Validation (LOOCV) procedure used in this study

3 Result and discussion

This section provides an in-depth examination of both the experimental dataset and the predictive behavior of the proposed ANN models for estimating surface roughness (Ra) and resultant force (F) in CNC turning of hardened AISI H13 steel. The analysis covers the statistical properties of the data, the correlation structure among variables, the design and assessment of the ANN architectures, and the outcomes of the hyperparameter tuning process. Model accuracy is evaluated quantitatively using three regression indicators: mean absolute error (MAE), root-mean-square error ($RMSE$), and the coefficient of determination (R^2). The ensuing discussion interprets these results, clarifies the observed discrepancy in prediction quality between force and surface roughness, and positions the findings in the context of related work reported in previous studies.

3.1 Dataset

As described previously in Section 2.1, the analysis is based on a dataset of 324 observations collected from dry CNC turning experiments on hardened AISI H13 tool steel. Each observation contains three main input parameters: cutting speed (vc), feed rate (f), and depth of cut (ap) together with two output responses: surface roughness (Ra , μm) and resultant force (F , N). The trials were arranged according to a full factorial experimental scheme and carried out in a controlled laboratory environment using fresh carbide cutting tools [25].

Before training the ANN models, all input variables were rescaled to the interval $[0,1]$ using Min-Max normalization to promote numerical stability and consistent feature magnitudes. The basic descriptive statistics of the dataset, namely the minimum, maximum, and mean for each variable, are reported in Table 1.

Table 1. Descriptive statistics of CNC turning dataset (EXP1)

Variables	Data Type	Min	Max	Mean
vc	Input	310	390	350
f	Input	0.07	0.13	0.1
ap	Input	0.25	0.8	0.517
Ra	Output	0.211	1.812	0.716

F	Output	69.338	377.573	187.948
-----	--------	--------	---------	---------

The wide distribution of machining parameters and the simultaneous availability of geometric and force-related outputs provide a strong basis for developing and validating ANN-based predictive models for CNC turning applications [13].

3.2 Exploratory data analysis (EDA)

To better understand the underlying relationships between the machining parameters and the resulting process outputs, an exploratory data analysis (EDA) was performed. The present analysis centered on the statistical interactions between three input variables: cutting speed (vc), feed rate (f), and depth of cut (ap) and the output responses of surface roughness (Ra) and resultant force (F). Identifying these interdependencies is essential for guiding the ANN model design and improving its generalization performance [26].

To quantify the strength and direction of linear associations among the variables, Pearson correlation coefficients were computed. The results of this study are presented in a correlation heatmap in Fig. 4. The Pearson coefficient, a measure of linear correlation, ranges from -1 to $+1$, where values approaching $+1$ indicate a strong positive linear relationship, and values close to zero indicate weak or negligible linear dependence [13].

As observed in Fig. 4, the depth of cut (ap) exhibits a strong positive correlation with resultant force (F) ($r = 0.90$), indicating that increased cutting depth substantially elevates the cutting force, a trend consistent with fundamental material removal mechanics. Similarly, feed rate (f) shows a strong positive correlation with surface roughness (Ra) ($r = 0.76$), implying that higher feed rates generally lead to rougher surface textures. In contrast, cutting speed (vc) demonstrates only weak linear correlations with both Ra and F , suggesting that its influence is predominantly nonlinear and not fully captured by simple pairwise correlation analysis [27].

These observations confirm that ap and f are dominant predictors in CNC turning, while the effect of vc requires nonlinear modeling approaches. Moreover, the correlation matrix provides valuable guidance for feature prioritization and multicollinearity control, both of which contribute to stable training and robust generalization of the ANN models.

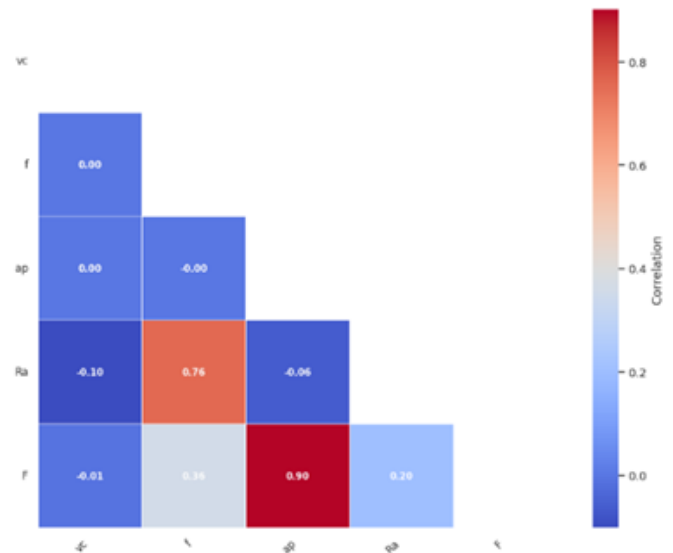


Fig. 4. Pearson correlation heatmap illustrating relationships between input and output variables in the CNC turning dataset

3.3 Artificial neural network (ANN) modeling

The artificial neural network (ANN) models developed in this study were implemented using Python, with the TensorFlow library serving as the core platform for deep learning execution. Each model was constructed to learn the mapping between three machining inputs cutting speed (vc), feed rate (f), and depth of cut (ap) and a

target machining response, either surface roughness (Ra) or resultant resulforce (F).

The architecture of each ANN model featured a single output neuron, designed to produce a continuous prediction for one of the output variables. To investigate the influence of architectural complexity, four distinct network configurations referred to as Arch 1 through Arch 4 were formulated, each varying in depth and number of neurons per hidden layer. The term "Arch" denotes "Architecture" and is used throughout this study to describe the specific structural arrangement of the neural networks. Details of these configurations are presented in Table 2.

Table 2. ANN architectures for CNC turning prediction

Name	Input Layer	Hidden layer	Neuron	Output Layer
Arch 1	1	2	32*16	2
Arch 2	1	3	64*32*16	2
Arch 3	1	4	128*64*32*16	2
Arch 4	1	5	256*128*64*32*16	2

Prior to training, all input features were normalized using Min-Max scaling to a range of [0, 1]. This step enhances numerical stability and accelerates model convergence. Fig. 5 illustrates the comparison between original and normalized values for Ra and F .

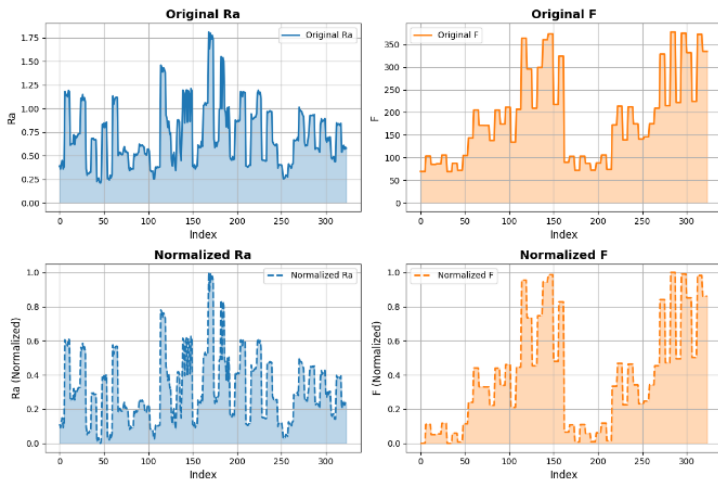


Fig. 5. Comparison of original and normalized data for surface roughness (Ra) and resultant force (F)

As shown in Fig. 5, the input features and corresponding output responses Ra and F were normalized prior to training. All ANN configurations were trained over 500 epochs using a consistent batch size of 16 for both prediction tasks. Although preliminary trials revealed that a batch size of 8 marginally improved predictions for F , adopting 16 across both models ensured better consistency, convergence stability, and computational efficiency with no significant trade-off in accuracy [28].

The training routine was executed using the RMSprop optimizer, selected for its effectiveness in handling non-stationary objectives, with a fixed learning rate of 0.0001. Additionally, an early-stopping mechanism was integrated, halting training automatically if no further reduction in validation loss was observed over 20 consecutive epochs. To evaluate model performance, the dataset was partitioned into an 80% training set and a 20% testing set [29].

Prediction accuracy was quantified using three standard regression metrics: mean absolute error (MAE), root-mean-square error ($RMSE$), and the coefficient of determination (R^2). A comparison of performance across architectures for predicting F is presented in Table 3, showing that Architecture 4 achieved the highest predictive accuracy, with an R^2 of 0.9951 and substantially lower error values than the other configurations.

Table 3. Prediction performance arch for resultant force (F)

Name	MAE	RMSE	R-squared
Arch 1	49.6552	60.6188	0.62310

Arch 2	18.6027	22.1994	0.94970
Arch 3	13.0806	16.3170	0.9727
Arch 4	4.5620	6.9441	0.9951

The comparative model performance for surface roughness (Ra) is shown in Table 4. Architecture 3 produced the most favourable results, with the lowest MAE and $RMSE$ values and the highest coefficient of determination (R^2) of 0.8201. Although Architecture 4 yielded comparable accuracy, the marginal increase in architectural complexity did not result in a significant improvement, making Arch 3 a more computationally efficient choice.

Table 4. Prediction performance arch for surface roughness (Ra)

Name	MAE	RMSE	R-squared
Arch 1	0.1306	0.1702	0.7571
Arch 2	0.1223	0.1510	0.8119
Arch 3	0.1194	0.1476	0.8201
Arch 4	0.1195	0.1480	0.8193

These results indicate that increasing the depth and number of neurons in the ANN generally improves predictive performance [30], particularly for more complex targets such as resultant force (F). However, beyond a certain point, additional complexity yields diminishing returns. Consequently, Arch 4 was selected as the optimal configuration for predicting force (F), while Arch 3 was chosen for surface roughness (Ra), balancing accuracy and computational cost.

3.4 Hyperparameter tuning

After selecting the optimal ANN Architecture 4 for predicting resultant force (F) and Architecture 3 for surface roughness (Ra), a thorough hyperparameter tuning process was conducted to further improve model generalization and predictive accuracy. The primary objective was to determine the best combination of training configurations that minimized prediction error on unseen data. Three key hyperparameters were considered during this process: the type of optimization algorithm, learning rate, and batch size. These hyperparameters were evaluated empirically through multiple training experiments, and the model performances were quantitatively assessed using standard regression metrics, namely Mean Absolute Error (MAE), Root Mean Square Error ($RMSE$), and the coefficient of determination (R^2) [13, 23].

3.4.1 Optimizer search

The first stage of the tuning process involved investigating the influence of different optimisation algorithms on the learning behaviour and predictive quality of the ANN models. Eight optimizers were tested: Adam, RMSprop, Stochastic Gradient Descent (SGD), Adadelta, Adamax, Nadam, FTRL, and Adagrad. For consistency across comparisons, each optimizer was applied to both output prediction models using fixed configurations: a learning rate of 0.0001, a batch size of 16, and the respective best-performing Architecture 4 for F and Architecture 3 for Ra .

The results for the resultant force (F) prediction task are summarized in Table 5. Among all optimizers, Nadam delivered the most favourable performance with an MAE of 4.6075, $RMSE$ of 6.3980, and R^2 of 0.9951, matching the best result from Adam but exhibiting slightly improved numerical stability. In contrast, classical optimizers such as SGD, Adadelta, FTRL, and Adagrad yielded significantly higher errors and poor R^2 values, indicating unstable convergence and underfitting behaviour.

Table 5. Testing results of various optimizers for resultant force (F)

Optimizer	MAE	RMSE	R-squared
Adam	4.563	6.9441	0.9951
RMSprop	4.6807	7.3815	0.9945
SGD	108.1284	115.083	-0.9634
Adadelta	187.4357	211.586	-3.585
Adamax	9.0818	12.4242	0.9847

Nadam	4.6075	6.398	0.9951
FTRL	180.4367	207.1351	-3.3264
Adagrad	180.4967	205.1761	-3.2964

A similar trend was observed in the surface roughness (Ra) prediction task, as shown in Table 6. Nadam again produced the best overall performance, recording an MAE of 0.1199, $RMSE$ of 0.1479, and an R^2 of 0.8196. Although the performance of Adam was closely comparable, Nadam's slightly lower $RMSE$ and smoother training curves justified its selection as the final optimizer. In contrast, optimizers like SGD, Adadelta, and FTRL resulted in much poorer accuracy and even negative R^2 values, reflecting their unsuitability for the dataset and ANN configuration used in this study.

Table 6. Testing results of various optimizers for surface roughness

Optimizer	MAE	RMSE	R-squared
Adam	0.1195	0.1480	0.8193
RMSprop	0.1265	0.1643	0.8164
SGD	0.2322	0.2964	-0.2099
Adadelta	0.6049	0.7338	-3.2438
Adamax	0.1330	0.1715	0.7385
Nadam	0.1199	0.1479	0.8196
FTRL	0.1690	0.2090	0.4037
Adagrad	0.1299	0.1626	0.8046

These results validate the superiority of adaptive optimizer especially Nadam in training deep neural networks for complex regression tasks in machining applications. Nadam's integration of Adam's adaptive learning with Nesterov accelerated gradients contributed to stable learning dynamics and high prediction fidelity. As such, Nadam was selected as the final optimizer configuration for subsequent training and validation procedures.

3.4.2 Learning rate search

Following the identification of Nadam as the most effective optimisation algorithm, the next stage of hyperparameter tuning focused on determining the optimal learning rate to further enhance model performance. Three learning rates 0.01, 0.001, and 0.0001 were evaluated while holding other training settings constant. These included the use of Nadam as the optimizer, a batch size of 16, and the best-performing ANN architectures (Architecture 4 for predicting resultant force and Architecture 3 for predicting surface roughness).

The impact of learning rate variations on the prediction accuracy of the resultant force (F) model is summarized in Table 7. The learning rate of 0.0001 achieved the best performance, with an MAE of 4.609, $RMSE$ of 6.913, and an R^2 of 0.9951, indicating excellent agreement between the predicted and actual values. A similar pattern was observed for the surface roughness (Ra) model, where the same learning rate (0.0001) produced superior results with an MAE of 0.119, $RMSE$ of 0.148, and R^2 of 0.8198, as shown in Table 8.

Table 7. Effect of learning rate variations on resultant force (F) prediction

Learning Rate	MAE	RMSE	R-squared
Lr = 0.01	6.825	9.003	0.9915
Lr = 0.001	4.863	7.319	0.9945
Lr = 0.0001	4.609	6.913	0.9951

These findings demonstrate that using a smaller learning rate significantly improves training stability and predictive accuracy [22]. In regression tasks involving subtle nonlinearity in CNC machining response modeling fine-tuned parameter updates play a critical role in achieving high model fidelity. The consistent superiority of the 0.0001 learning rate across both prediction tasks confirms its suitability for precise convergence and error minimisation.

Table 8. Effect of learning rate variations on surface roughness (Ra) prediction

Learning Rate	MAE	RMSE	R-squared
---------------	-----	------	-----------

Lr = 0.01	0.127	0.159	0.7911
Lr = 0.001	0.121	0.149	0.8162
Lr = 0.0001	0.119	0.148	0.8198

3.4.3 Batch size search

The final stage of hyperparameter tuning evaluated the impact of batch size on model convergence, prediction stability, and computational efficiency. Using Architecture 4 for resultant force (F) and Architecture 3 for surface roughness (Ra), four batch sizes 8, 16, 32, and 64 were tested. The Nadam optimizer and a fixed learning rate of 0.0001 were consistently applied across all configurations.

For the resultant force prediction, although batch size 8 yielded the lowest MAE (4.5483), batch size 16 produced nearly equivalent results ($MAE = 4.609$, $RMSE = 6.9126$, $R^2 = 0.9951$), while offering smoother convergence and lower computational overhead [28]. Table 9 summarises the model's performance at different batch sizes.

Table 9. Performance of resultant force (F) model at different batch sizes

Batch Sizes	MAE	RMSE	R-squared
8	4.5483	6.9586	0.9950
16	4.609	6.9126	0.9951
32	9.4998	12.4314	0.9840
64	15.7138	19.0417	0.9629

In the surface roughness (Ra) prediction task, batch size 16 also outperformed other configurations, achieving the lowest $RMSE$ (0.1478) and the highest R^2 (0.8198), as detailed in Table 10. As batch size increased beyond 16, predictive performance declined noticeably, indicating reduced model sensitivity to subtle patterns within the data.

Table 10. Performance of Surface Roughness (Ra) Model at Different Batch Sizes

Batch Sizes	MAE	RMSE	R-squared
8	0.1194	0.1480	0.8193
16	0.1194	0.1478	0.8198
32	0.1246	0.1607	0.7818
64	0.1426	0.1995	0.6655

Considering the comparable performance of batch sizes 8 and 16 for force prediction, and the superior results of batch size 16 for surface roughness, a unified batch size of 16 was adopted as the final configuration for both ANN models. This choice ensures modeling consistency, training efficiency, and ease of deployment without compromising prediction quality.

The final model performance is visually summarized in Fig. 6, which compares the MAE , $RMSE$, and R^2 values for both prediction targets. The results show that the model achieved high accuracy in predicting the resultant force (F), indicated by low error metrics and an R^2 of 0.9931. For surface roughness (Ra), the model attained reasonable predictive performance, with an MAE of 0.1194, an $RMSE$ of 0.1478, and an R^2 of 0.6198. These outcomes confirm the effectiveness of the selected hyperparameters in enabling the ANN to generalize well across both outputs.

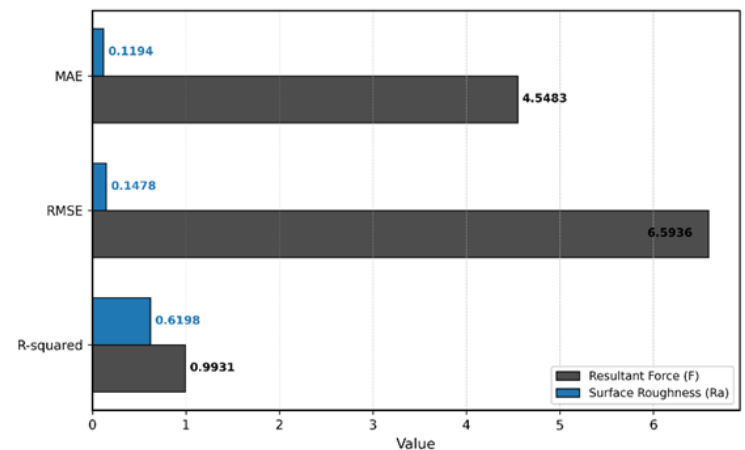


Fig. 6. Evaluation metrics for ANN model.

3.5 Model validation

To evaluate how well the proposed ANN models generalize to assess the generalization capability of the proposed ANN models on unseen data, the Leave-One-Out Cross-Validation (LOOCV) technique was employed [12]. This validation strategy is particularly suitable for small-sized datasets, as each data point is successively used once as a testing sample while the remaining observations form the training set. Although computationally intensive, LOOCV provides a comprehensive and unbiased estimate of model robustness and predictive reliability. In this study, a total of 324 validation iterations were performed, corresponding to the number of available experimental samples. The ANN configurations used in this stage followed the optimal architectures identified earlier, namely Architecture 4 for resultant force (F) prediction and Architecture 3 for surface roughness (Ra) prediction, with a batch size of 16, Nadam optimizer, and a learning rate of 0.0001. Training was conducted for a maximum of 500 epochs with an early-stopping criterion to prevent overfitting.

As illustrated in Fig. 7, both models exhibited smooth and stable convergence of training and validation loss curves, indicating well-controlled learning behaviour. Fig. 8 presents the model predictions obtained without LOOCV, while Fig. 9 shows the corresponding scatter plots after LOOCV implementation. The resultant force model produced predictions that were tightly clustered around the reference line, confirming excellent prediction accuracy. In contrast, surface roughness predictions showed a slightly higher dispersion; however, a clear linear trend was still observed, reflecting reasonable predictive capability despite the inherently higher variability of Ra .

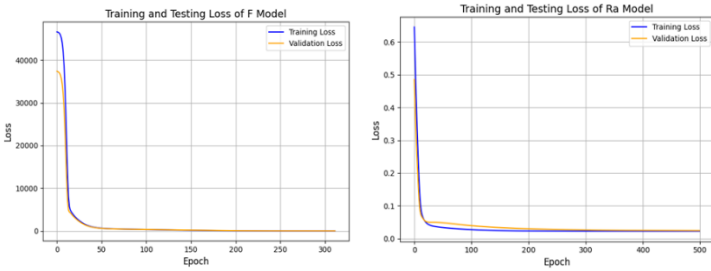


Fig. 7. Training and validation loss curves of ANN models.

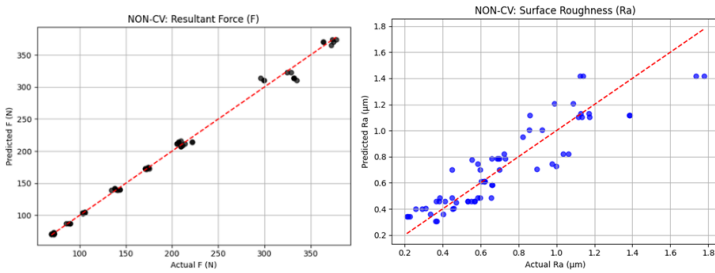


Fig. 8. Prediction results without LOOCV for force and roughness.

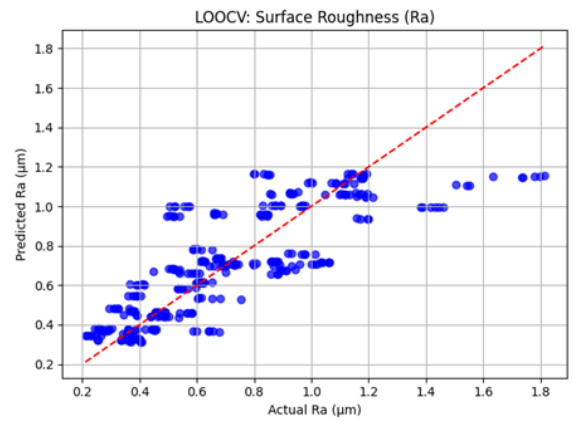


Fig. 9. Prediction results with LOOCV for force and roughness.

These results validate the robustness of the proposed ANN models, demonstrating their effectiveness in predicting machining responses and laying the groundwork for future integration into intelligent manufacturing systems.

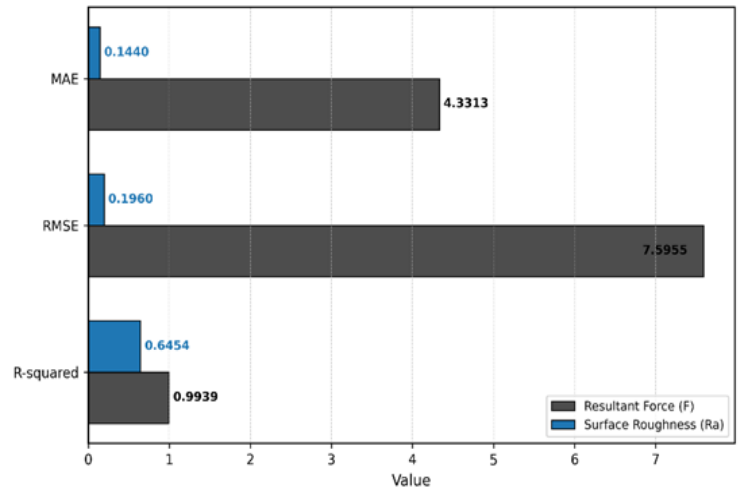


Fig. 10. Model evaluation metrics with LOOCV

The quantitative performance metrics obtained from the LOOCV procedure are summarized in Fig. 10. For the resultant force (F), the ANN model achieved an MAE of 4.3313, an RMSE of 7.5955, and an R^2 value of 0.9939, indicating that the predicted forces closely match the experimental measurements. For surface roughness (Ra), the model obtained an MAE of 0.1440, an RMSE of 0.1960, and an R^2 value of 0.6454, which can be regarded as acceptable given the complex and partly stochastic mechanism of surface finish formation in machining [10, 31].

Overall, these LOOCV results show that the proposed ANN-based framework is robust and generalizes well to unseen data. The very high accuracy for resultant force and the reasonable performance for surface roughness confirm that the models are able to capture the nonlinear relationships between machining parameters and responses. In addition, the validated framework has strong potential for future integration into intelligent manufacturing systems and data-driven strategies for machining process optimization.

4 Conclusion

This study confirms the effectiveness of artificial neural networks for modeling machining responses in CNC turning of hardened AISI H13 steel. The developed ANN models were able to capture the complex nonlinear relationships between machining parameters and output responses with a high level of predictive reliability. The cutting force model exhibited excellent predictive performance, while the surface roughness model achieved satisfactory performance despite the inherent variability of surface integrity phenomena. From a practical standpoint, the proposed ANN-based framework offers a valuable decision-support tool for

machining process planning and optimization. By enabling reliable prediction of cutting force and surface quality prior to actual machining, the model can assist engineers in selecting optimal cutting conditions, improving process stability, reducing tool wear, and minimizing trial-and-error experimentation. This capability is particularly beneficial for high-performance and precision manufacturing environments where productivity and surface integrity are critical. The robustness of the proposed approach was further confirmed through Leave-One-Out Cross-Validation (LOOCV), indicating strong generalization performance even under limited data availability. Future work may focus on extending the model by incorporating dynamic process variables such as tool wear, vibration, and temperature to enhance real-time adaptability.

References

- [1] R. Mallick, R. Kumar, A. Panda, and A. K. Sahoo, "Current status of hard turning in manufacturing: Aspects of cooling strategy and sustainability," *Lubricants*, vol. 11, no. 3, p. 108, 2023.
- [2] A. Siahsharani, M. Paknejad, and B. Azarhoushang, "Deburring of micro-milled hardened steel: influence of milling strategies and CNC-based post-polishing," *The International Journal of Advanced Manufacturing Technology*, pp. 1-10, 2025.
- [3] M. M. Monjez, N. Omid, P. Farhadipour, A. El Ouafi, and N. Barka, "Influence of Different Heat Treatments on Microstructure Evolution and High-Temperature Tensile Properties of LPBF-Fabricated H13 Hot Work Steel," *Metals*, vol. 15, no. 9, p. 1003, 2025.
- [4] S. Kolomy *et al.*, "Machinability of extruded H13 tool steel: Effect of cutting parameters on cutting forces, surface roughness, microstructure, and residual stresses," *Alexandria Engineering Journal*, vol. 99, pp. 394-407, 2024.
- [5] M. Akgün, B. Özlü, and F. Kara, "Effect of PVD-TiN and CVD-Al₂O₃ coatings on cutting force, surface roughness, cutting power, and temperature in hard turning of AISI H13 steel," *Journal of Materials Engineering and Performance*, vol. 32, no. 3, pp. 1390-1401, 2023.
- [6] J. H. Ko and C. Yin, "A review of artificial intelligence application for machining surface quality prediction: From key factors to model development," *Journal of Intelligent Manufacturing*, pp. 1-24, 2025.
- [7] B. P. Kamiel, A. D. Saputri, Z. H. Muizza, and A. Yobioktabera, "Smart Harvest: Web-Integrated Ripeness Detection for Apples with CNN Algorithm," *Ingénierie des Systèmes d'Information*, vol. 29, no. 6, 2024.
- [8] C. Mücher, "Artificial neural network based non-linear transformation of high-frequency returns for volatility forecasting," *Frontiers in Artificial Intelligence*, vol. 4, p. 787534, 2022.
- [9] P. Charilaou and R. Battat, "Machine learning models and overfitting considerations," *World Journal of Gastroenterology*, vol. 28, no. 5, p. 605, 2022.
- [10] M.-H. Tsai, J.-N. Lee, H.-D. Tsai, M.-J. Shie, T.-L. Hsu, and H.-S. Chen, "Applying a neural network to predict surface roughness and machining accuracy in the milling of SUS304," *Electronics*, vol. 12, no. 4, p. 981, 2023.
- [11] S. Chinchankar, S. Shinde, V. Gaikwad, A. Shaikh, M. Rondhe, and M. Naik, "ANN modelling of surface roughness of FDM parts considering the effect of hidden layers, neurons, and process parameters," *Advances in materials and processing technologies*, vol. 10, no. 1, pp. 22-32, 2024.
- [12] V. W. Lumumba, D. Kiprotich, M. Lemasulani Mpaine, N. Grace Makena, and M. Daniel Kavita, "Comparative analysis of Cross-Validation techniques: LOOCV, K-folds Cross-Validation, and repeated K-folds Cross-Validation in machine learning models," *K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models (June 01, 2024)*, 2024.
- [13] D. Leni, "Prediction modeling of low alloy steel based on chemical composition and heat treatment using artificial neural network," *Jurnal Polimesin*, vol. 21, no. 5, pp. 530-537, 2023.
- [14] W. Li, L. Zhang, C. Wu, Z. Cui, and C. Niu, "A new lightweight deep neural network for surface scratch detection," *The International Journal of Advanced Manufacturing Technology*, vol. 123, no. 5, pp. 1999-2015, 2022.
- [15] J.-S. Hwang, S.-S. Lee, J.-W. Gil, and C.-K. Lee, "Determination of optimal batch size of deep learning models with time series data," *Sustainability*, vol. 16, no. 14, p. 5936, 2024.
- [16] E. Marevac, E. Kadušić, N. Živić, D. Hamzić, and N. Hadžajlić, "Data-Driven Predictive Analytics for Dynamic Aviation Systems: Optimising Fleet Maintenance and Flight Operations Through Machine Learning," *Future Internet*, vol. 17, no. 11, p. 508, 2025.
- [17] D. Leni, A. Karudin, M. R. Abbas, J. K. Sharma, and A. Adriansyah, "Optimizing stainless steel tensile strength analysis: through data exploration and machine learning design with Streamlit," *EUREKA: Physics and Engineering*, no. 5, pp. 73-88, 2024.
- [18] P.-M. Huang and C.-H. Lee, "Estimation of tool wear and surface roughness development using deep learning and sensors fusion," *Sensors*, vol. 21, no. 16, p. 5338, 2021.
- [19] M. Sana, A. Khan, M. U. Farooq, and S. Anwar, "Artificial neural networks-based modelling of effects of cryogenic electrode treatment, nano-powder, and surfactant-mixed dielectrics on wear performance and dimensional errors on superalloy machining," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 46, no. 9, p. 539, 2024.
- [20] A. G. Ganie and S. Dadvandipour, "From big data to smart data: a sample gradient descent approach for machine learning," *Journal of Big Data*, vol. 10, no. 1, p. 162, 2023.
- [21] U. Saray and U. Çavdar, "Comparison of Different Optimization Algorithms in the Fashion MNIST Dataset," *International Journal of Multidisciplinary Studies and Innovative Technologies*, vol. 8, no. 2, pp. 52-58, 2024.
- [22] A. Sahraei, A. Chamorro, P. Kraft, and L. Breuer, "Application of machine learning models to predict maximum event water fractions in streamflow," *Frontiers in Water*, vol. 3, p. 652100, 2021.
- [23] T. O. Hodson, "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not," *Geoscientific Model Development Discussions*, vol. 2022, pp. 1-10, 2022.
- [24] V. Dubey, A. K. Sharma, and D. Y. Pimenov, "Prediction of surface roughness using machine learning approach in MQL turning of AISI 304 steel by varying nanoparticle size in the cutting fluid," *Lubricants*, vol. 10, no. 5, p. 81, 2022.
- [25] S. S. Patil, S. S. Pardeshi, N. Pradhan, and A. D. Patange, "Cutting tool condition monitoring using a deep learning-based artificial neural network," *International Journal of Performability Engineering*, vol. 18, no. 1, p. 37, 2022.
- [26] J. Sembiring, A. Amanov, and Y. S. Pyun, "Artificial neural network-based prediction model of residual stress and hardness of nickel-based alloys for UNSM parameters optimization," *Materials Today Communications*, vol. 25, p. 101391, 2020.
- [27] S. Mane, R. B. Patil, and S. Al-Dahidi, "Predictive modeling of surface roughness and cutting temperature using response surface methodology and artificial neural network in hard turning of AISI 52100 steel with minimal cutting fluid application," *Machines*, vol. 13, no. 4, p. 266, 2025.
- [28] D. Granzol, S. Zohren, and S. Roberts, "Learning rates as a function of batch size: A random matrix theory approach to neural network training," *Journal of Machine Learning Research*, vol. 23, no. 173, pp. 1-65, 2022.
- [29] H.-C. Chen *et al.*, "Estimation of various walking intensities

based on wearable plantar pressure sensors using artificial neural networks," *Sensors*, vol. 21, no. 19, p. 6513, 2021.

- [30] A. Sharma *et al.*, "Machine learning based approach for surface roughness prediction in precision dental prototyping," *Scientific Reports*, vol. 15, no. 1, p. 32239, 2025.
- [31] K. Antosz, E. Kozłowski, J. Sęp, and S. Prucnal, "Application of Machine Learning to the Prediction of Surface Roughness in the Milling Process on the Basis of Sensor Signals," *Materials*, vol. 18, no. 1, p. 148, 2025.