Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 46 No. 1
january 2025

Development of a Predictive Model for Tool Wear in Cutting and Drilling for Multipurpose Machining

Allahaddje Bonheur[1, *], James Kuria Kimotho[2], Mutuku Muvengei[3]

[1]Pan African University, Institute for Basic Sciences Technology and Innovation (PAUSTI), Nairobi, Kenya

[2, 3]Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya

Abstract: Accurate prediction of tool wear is important in machining to ensure product quality and operational efficiency. However, the conventional methods for monitoring the tool wear are generally slow, inefficient, and fail to deliver the real-time accuracy necessary for industrial environments, limiting their effectiveness in modern manufacturing environments. Therefore, this study introduces a novel framework for real-time flank wear prediction, leveraging high-frequency vibration data collected from an accelerometer mounted on a locally manufactured multipurpose machine. The collected dataset, covering cutting and drilling operations, was used to train and test two deep learning models: Long Short-Term Memory (LSTM) and transformer networks. The novelty of this work lies in the detailed comparative analysis and optimization of these models, designed specifically for machining environments. By using several calibration techniques, the optimized LSTM showed higher performance through experiments with a Root Mean Square Error (RMSE) of 0.1417, Mean Absolute Error (MAE) of 0.076, and coefficient of determination (R^2) of 0.987, outperforming the transformer model. This finding gives a predictive maintenance approach, addressing the challenges of real-time tool wear monitoring and offering insights into model selection for industrial applications. The proposed framework sets a new benchmark for integrating adaptive predictive techniques into manufacturing, enhancing tool life, minimizing downtime, and supporting the flexibility of predictive maintenance in manufacturing settings.

Keywords: Flank Wear, Deep Learning, Accuracy Prediction, Sensor placement, Operational Efficiency

1. Introduction

In the progressive framework of Industry 4.0, precision and manufacturing efficiency are the highest priority. A concern in the tool wear management impacts many factors such as operational efficiency, product quality, and tool longevity. Effective monitoring and management of tool wear are essential to maintain production quality and minimize costly downtimes, especially as manufacturing systems increasingly adopt multifunctional machines that offer both flexibility and economic advantages [1, 2]. However, these machines present unique challenges for tool wear monitoring, emphasizing the pressing necessity for resilient, flexible solutions to maintain optimal performance and operational integrity.

Monitoring tool wear in the machine, especially locally built ones, presents unique challenges that limit the applicability of traditional predictive models. The conventional approach for monitoring tool wear is time-consuming and lacks the real-time precision essential for adaptive maintenance, highlighting the urgent need for innovative, high-precision solutions [8]. Existing methods are usually adapted to specific machine types, or defined workpieces, reducing their effectiveness in the diverse and complex environments of machining [3, 4]. The most conventional designs lack real-time data processing capabilities, which are important for implementing predictive and proactive maintenance strategies in high-demand industrial manufacturing applications. Therefore, there is an immediate need for a predictive framework that not only provides real-time insights but also adapts to the specific requirements of machining environments, particularly those involving locally produced equipment [8].

The degradation of the tool occurs in two primary forms: crater wear (KT) and flank wear (VB), as shown in Figure 1. Flank wear (VB), is specially seen on the sides of the tooth, and crater wear (KT), occurs on the rake face of the tool. In the case of the hacksaw, the wear commonly referred to as flank

wear (VB) corresponds to wear on the rake face of the tooth, which is the predominant form of degradation, while wear on the sides of the tooth is less significant compared to other cutting tools. This analysis emphasizes flank wear (rake face in hacksaw), as it has a higher impact on cutting efficiency, dimensional accuracy, and overall performance of the tool than crater wear. Flank wear (VB), generally from 0.3 mm to 0.6 mm as stated by ISO norm 3685:1993 standards, with values exceeding this limit causing lower cutting efficiency, increased risk of tool failure, and degraded process reliability [6]. VB is fundamental for determining the durability of the tool. When flank wear exceeds these limits, tool effectiveness declines significantly, increasing the risk of failure and compromising process reliability [7].

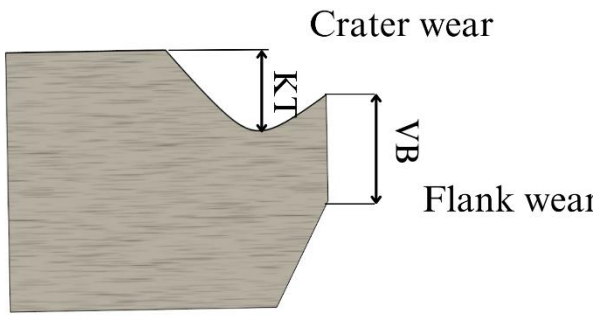


Figure 1 Representation of Flank Wear (VB) and Crater Wear (KT).

A predictive approach for real-time flank wear forecasting using deep learning models is presented in this study, specifically Long Short-Term Memory (LSTM) networks and transformer models. LSTM networks are highly effective for sequential data processing, making them ideal for real-time applications, while transformer models excel at capturing complex, long-range dependencies in time-series data. This capability makes it especially well-suited for the dynamic nature of machining environments. Few studies have compared LSTM and transformer models in the context of prediction. This work addresses this gap by evaluating and optimizing the best model for predictive maintenance in machining environments.

The main goal of this study is to develop an adaptive, real-time predictive maintenance solution that improves tool longevity and operational efficiency by accurately forecasting flank wear. By comparing LSTM and transformer models, the research aims to identify the most effective approach for monitoring tool wear in Industry 4.0 environments, focusing on locally manufactured, multifunctional machines.

The structure of this manuscript is organized into the following: Section 2 evaluates relevant related work, finding limitations in existing methods for tool wear. Section 3 describes the methodology for data collecting protocols, feature extraction, and building of the model and the regression metrics. Section 4 offers a comparison performance by experimental analysis, focusing on projected accuracy and computing efficiency. Finally, Section 5 summarizes the meaning of the study for real-time predictive maintenance and identifies future research topics for developing adaptive industrial systems.

2. Related work

The development of predictive models for flank wear in cutting and drilling operations is important for enhancing the efficiency and reliability of functional machining systems. Current developments in machine learning and deep learning algorithms have more contributed to this field, providing frameworks for real-time monitoring and predictive maintenance.

Initial research in tool wear prediction used mainly conventional machine learning techniques, including artificial neural networks (ANNs) and several other algorithms. Kilickap et al. revealed that Artificial Neural Networks could reach 85% precision in the prediction of the wear during the milling operation of Ti-6242S, but Sk et al. demonstrated an approach similar to 88% precision in predicting the wear of the tool in hard turning of EN8 steel material. Alajmi and Almeshal [1, 2] also applied the XGBoost-SDA algorithm for tool wear management in drilling, resulting in a mean squared error (MSE) of 0.0052, higher than the conventional approach such as Support Vector Machines (SVM) and Multi-Layer Perceptron (MLP). Despite these improvements, the level of detail and flexibility of real-world machining configurations remained major difficulties, requiring the creation of increasingly advanced prediction techniques.

Recent research has increasingly focused on

deep learning models, particularly Long-Short-Term Memory (LSTM) networks and convolutional neural networks (CNNs), which excel at capturing temporal dependencies and complex patterns in machining data (Yang et al.[12] introduced an algorithm for wear forecasting framework built around LSTM, which combined handmade features with automatic learning of features, getting a prediction accuracy of 92%. Similar to this, Huang et al. [13] utilized a CNN with a Short-Time Fourier Transform (STFT) for wear of tools surveillance, obtaining 91% accuracy. Duan et al. [14] Further proving the effectiveness of CNNs, a multi-frequency-band deep CNN model was created, which obtained an excellent 97% accuracy in forecasting tool wear utilizing characteristics from vibration and force data. These studies illustrate the potential of deep learning to increase predictive accuracy in complex data sets.

Combining deep learning architectures has also been researched to increase forecast accuracy. Huang et al.[15] created a multi-scale convolutional neural network with attention fusion for tool wear prediction, which displayed a considerable improvement in accuracy by efficiently extracting sensitive wear deterioration signals from monitoring data. This model points out the capabilities of CNNs and attention mechanisms and emphasizes the hybrid technique's advantage in solving the issues of tool wear prediction.

The introduction of transformer models has marked a significant advancement in predictive maintenance. These models leverage self-attention mechanisms to capture complex, long-range dependencies in time-series data. Yang et al. [16] used transformers for tool wear prediction, which face challenges in accomplishing forecast accuracies of up to 93% in milling. Their methodology effectively employed high-frequency vibration data, demonstrating the efficacy of transformers in real-time monitoring contexts. In addition, Lu et al. [17] implemented a transformer-based model to forecast flank wear in micro-milling processes, achieving an accuracy of 94%. This demonstrates the applicability of transformer models to several machining environments.

The integration of real-time data collecting systems and sensor fusion techniques has significantly progressed the field. Liu et al. [18] suggested a tool wear prediction model utilizing sensor data, which improved forecasting accuracy by integrating diverse data sources. Their methodology underlined the significance of amalgamating sensor inputs to deliver a thorough comprehension of tool wear dynamics, attaining an accuracy of 90%. Alam et al. [19] identified the value of combining wavelet transforms with machine learning approaches to boost the accuracy of tool wear forecasts. Their work has shown that utilizing wavelet transforms for feature extraction could improve the efficacy of prediction models, getting accuracies of 95% in specific applications.

Despite improvements, challenges continue in the implementation of these models to locally made equipment. Many existing models for prediction are targeted to certain machine types or operational scenarios, limiting their application in different machining environments. Also, the processing requirements of complex models, such as transformers, can limit their implementation in real-time systems. Therefore, establishing adaptive frameworks that can handle the particular requirements of locally produced equipment while maintaining high predicted accuracy is needed. The comparative analysis of LSTM and transformer models, as described in this paper, would provide useful insights into model selection and optimization for real-time predictive maintenance in complicated machining situations. The research shows a clear trend toward using deep learning algorithms for predicting tool wear, significantly enhancing reliability and accuracy. However, there is still an important need for adaptable and efficient models for predictive maintenance in Industry 4. 0.

3. Materials and Methods

This section defines an approach for constructing the suggested flank wear prediction system, including data collecting, signal processing, feature extraction, predictive modeling, and evaluation metrics. The methodology covers experimental obstacles and confirms every step, ensuring conformity with innovative techniques. Figure 2 illustrates the integration framework of sensor design, data preprocessing, model building, and continuous validation to provide precise and dependable predictions.

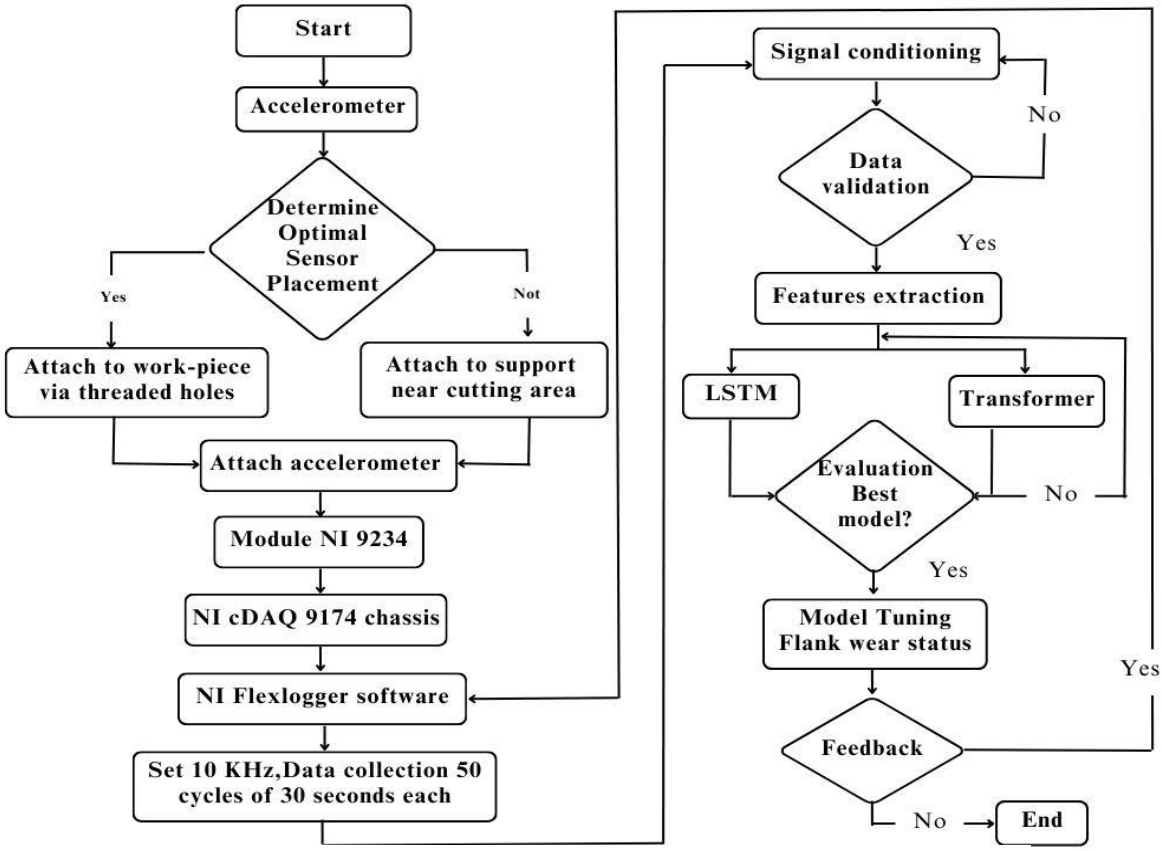


Figure 2 Framework for Flank Wear Prediction System

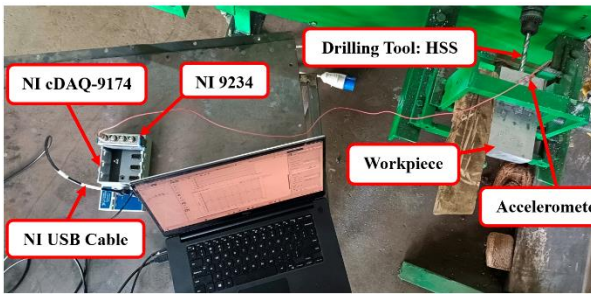
Data acquisition

The experiments were conducted using an accelerometer mounted near the cutting area to capture vibration signals. These signals served as indicators of flank wear on the drill bit and the rake face of the hacksaw during the cutting process. A sensor was interfaced through a NI 9234 module and NI cDAQ-9174 chassis, connected to a computer with NI FlexLogger software for real-time monitoring and data logging. The sampling rate was defined as 2 kHz to capture the high-frequency rate of vibrations associated with wear progression with high precision accuracy. The information was recorded in a CSV file format to simplify the procedure of action of pre-processing and analysis.

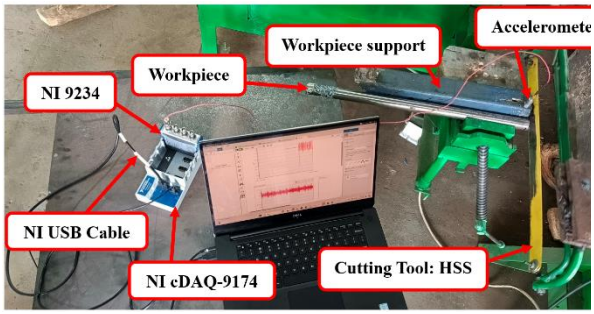
The tool wear data was collected during the drilling and cutting of the specimens where a custom-designed experimental setup was used to create realistic manufacturing conditions during the machining processes:

❖ Drilling Experiment (Fig. 31): A drill bit made of high-speed steel with a diameter of 9 mm has been employed on the surface made of aluminum. An accelerometer was mounted on the workpiece using threaded holes for the permanent acquisition of vibration information. The data was sent through cables to the National Instruments module NI 9234, then to the NI cDAQ-9174 chassis, and recorded on the laptop using the NI FlexLogger software with a sample rate of 2 kHz. Every 30-second cycle of each hole drilling was selected to enhance the collection of requisite data and at the same time ensure tool wear is accurately monitored. Such time intervals made it possible for enough flank wear to take place which made the vibration signals obtained interesting and also mimicked realistic conditions during the machining of tools whenever short periods were encountered. It also enhanced the signal-to-noise ratio and thus reliable data was guaranteed. Over 50 cycles were taken, and this arrangement made possible the creation of huge datasets hence wear pattern recognition plus model prediction accuracy improvement was facilitated.

❖ Cutting Experiment fig(3b): The mild steel rod has been machined with a hacksaw blade which is tooth-shaped and made of high-speed steel. To obtain a suitable dataset for comparative study purposes, data was collected for 30 seconds and repeated for 50 cycles at the sample rate of 2 kHz.



(a)



(b)

Figure 3 Experimental setup: (a) drilling and (b) cutting

According to the information obtained, the progress of flank wear was determined by the signals that were registered and these were used to comprehend the drilling and cutting operations. The penetration in tool wear was evidenced by the increase in the amplitude and frequency of the signals, which was mainly in the high-frequency areas and represented the indicators of wear. Besides, the data equally pointed out, the variability of the operations with cutting operations possessing larger amplitudes of vibration than drilling, because of the specifics of tool-material interactions. These features are well shown in the example of the cutting operation when the vibration signal was recorded during the procedure. Figure 4 indicates that high-frequency oscillations characterize particular cutting processes. It was seen from the figure that the frequency oscillations failed together with the increase of cutting processes which indicated the wear of the tool. These facts indicate that using vibration signals in monitoring and characterizing tool wear progress is possible.

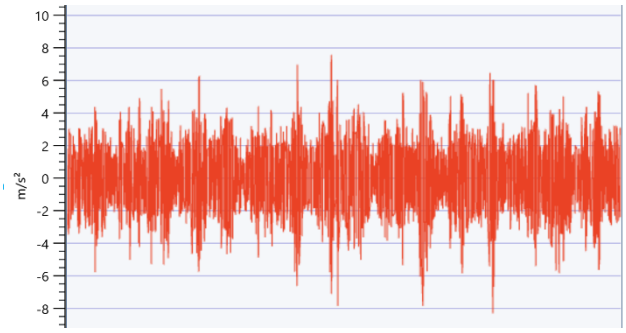


Figure 4 Vibration signal during cutting operation showing high-amplitude oscillations linked to wear progression

The beginning of the analysis has shown that the initial raw dataset was unreliable and required extensive data processing steps. Often, tool wear signals were present but were drowned out by other factors, such as environmental noise introduced by machine activity. Also, the alternation of the workpiece surface features produced noises. Those facets aggravated the importance of procedures to cudgel arrays that did not contain straightforward images and useful information.

A second-order Butterworth low-pass filter was employed to decrease high-frequency noise in the sensor data while maintaining critical low-frequency components. The filter's cut-off frequency was chosen using the Nyquist frequency, and zero-phase filtering was conducted to verify no phase distortion.

To facilitate processing, the timestamps were converted from string to numerical format. For normalization, the data were standardized using z-score scaling to ensure that each feature had a mean of 0 and standard deviation of 1, according to the formula in Eq. (1): where the representation of the original data is X, the representation of the mean is μ, and σ represents the standard deviation. This step prevented features with larger values from dominating the model, therefore ensuring a balanced contribution of the feature. These preprocessing steps, filtering, denoising, and normalization, effectively prepared the data for feature extraction and afterward model development.

$$Z = \frac{\mathrm{X} - \mu}{\sigma} \quad (1)$$

To assess tool wear during cutting and drilling, time-domain and frequency-domain data were extracted from acceleration sensor signals. These features help identify wear mechanisms affecting tool performance.

Time-domain features are necessary for taking statistical properties of vibration signals, such as mean (μ), which indicates the overall vibration level during machining Eq. (2); Root Mean Square (RMS), correlating with flank wear severity Eq. (3); standard deviation (σ), quantifying signal variability and process fluctuations Eq. 4; indication of the signal distribution imbalance and irregular tool behavior by skewness (γ_1), as represented in Eq. (5); and, the reflection of the sharpness or peaks in signal distribution, often indicating sudden impacts or failures by kurtosis (γ_2) in Eq. (6).

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (2)$$

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \quad (3)$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2} \quad (4)$$

$$\gamma_1 = \frac{1}{N}\sum_{i=1}^{N} (\frac{x_i - \mu}{\sigma})^3 \quad (5)$$

$$\gamma_2 = \frac{1}{N}\sum_{i=1}^{N} (\frac{x_i - \mu}{\sigma})^4 - 3 \quad (6)$$

Frequency-Domain Features: such characteristics take the basic vibration aspects and energy distribution, enabling an understanding of wear development as Eq. (7). The frequency with the largest intensity, indicating the primary vibration mode during machining and comparing with tool wear severity communicated by Peak Frequency (f_{peak}) in Eq. (8). Mean Frequency (f_{mean}): A balanced average of frequency components, representing changes in vibration behavior owing to tool wear over time. The measurement of total energy across all frequency components, offering insights into vibration signal strength and its association with wear is called Power Spectral Density Sum (PSD_{Sum}) shown in Eq. (9).

$$f_{peak} = argmax(|FFT_{(X)}|) \quad (7)$$

$$f_{mean} = \frac{\sum_{i=1}^{N} f_i.|FFT_{(X)}|}{\sum_{i=1}^{N} |FFT_{(X)}|} \quad (8)$$

$$PSD\ Sum = \sum_{i=1}^{N} PSD(x_i) \quad (9)$$

The mathematical model relating tool flank wear (VB) to the extracted features is described in Eq. (10). In this model, β_0 to β_6 indicate the coefficients characterizing the impact of each feature, but ϵ accounts for the error factor. This formulation illustrates the links between vibration data properties and their impact on wear mechanisms, offering an exhaustive structure for predicting flank wear. This model underlines the important relevance of vibration data in collecting and analyzing wear dynamics, giving helpful information for maintenance prediction purposes.

$$VB = \beta_0 + \beta_1.RMS + \beta_2.f_{peak} + \beta_3.f_{mean} + \beta_4.\sigma + \beta_5.\gamma_1 + \beta_6.\gamma_2 + \epsilon \quad (10)$$

The dataset used for this study is from vibration signals taken at a sample frequency of 2 kHz at 30-second intervals per recording. Each session gave 60,000 data points, and with 50 recordings conducted, a total of 3,000,000 data points were collected. For model building and evaluation, the dataset was assigned to training (80%) and testing (20%) subsets, resulting in 2,400,000 samples for training and 600,000 samples for testing. This rigorous technique guarantees the dataset is deep and authentic, allowing robust model performance evaluations.

Model development

As mentioned in the literature review, transformer and LSTM models are suitable for time-series analysis prediction. This section consists of both structures used to forecast flank wear, taking advantage of their respective strengths in handling sequential data. TensorFlow Keras was used to build the models, and the evaluation metrics were built on sklearn and ran in the Spyder IDE within Anaconda.

Transformer model

The transformer concept developed by Vaswani et al. in 2017 takes advantage of a self-attention system to appropriately break down consecutive inputs. Adapted to the regression evaluation environment, it effectively takes the temporal relationship in the tool data. The 2D tensor is the input of the model with a sequence duration of one, representing the data attributes. The model design began with a dense layer, which applied the ReLU activation function defined by the following equation:

$$H = ReLU(XW + b) \ (11)$$

W is the weight matrix and b is the bias vector. A rectified linear unit (ReLU) activation function includes nonlinearity in the model. To prevent overfitting, regularization of L2 was performed and formulated, as shown in Eq. (12). Where $L2(W)$ is the L2 regularization term, W_{ij} represents the weight values in the matrix W, λ is the regularization parameter, controlling the degree of penalization.

$$L2(W) = \lambda \sum_{i,j} W_{ij}^2 \ (12)$$

This regularization strategy penalizes big weights, aiding in model generalization. The most important characteristic of the transformer model is its multi-head self-attention mechanism, which provides ongoing concentration on various elements of the input sequence. This mechanism can be described by Eq. (13) as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{dk}}\right)V \ (13)$$

Q, K, and V denote the query, key, and value matrices, respectively. dk represents the dimensionality of the keys Each transformer block incorporates a Feed-Forward Network (FFN) with two thick layers and ReLU activations, thereby improving its capabilities for intricate representations.

Long Short-Term Memory (LSTM)

The LSTM approach has been recognized as the greatest useful for time-series data use as well as algorithms, obtaining links between time in sequences. The memory cells and gates forget, input, and output are known as the structures that regulate the transmission of data within the system. The forget gate discards unnecessary data from the prior cell state, whereas the input gate integrates new information. This characteristic allows LSTM to properly handle temporal data that are important for precise flank wear prediction.

The LSTM model in Figure 5 analyses sequential input, adding long-term dependencies.

The first 64-unit LSTM layer produces hidden states, but the additional 32-unit LSTM layer further analyzes and sends the final secret state to the outputs layer for prediction. Forget Gate (f_t) Eq. (14) determines the information that is eliminated from the cell state, and the input gate (i_t , $\tilde{C}_t$) Eq. (15,16) decides the information to add. The updated input cell state (C_t) in Eq. (17). The output Gate (o_t, h_t) Eq. (18, 19) control output. W and b denote weights and biases, h_t is the hidden state, x_t is the input at the time t, and σ is the sigmoid activation function. This architecture allows the LSTM to retain and utilize relevant temporal features, enhancing its ability to predict machining-induced flank wear.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \ (14)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \ (15)$$

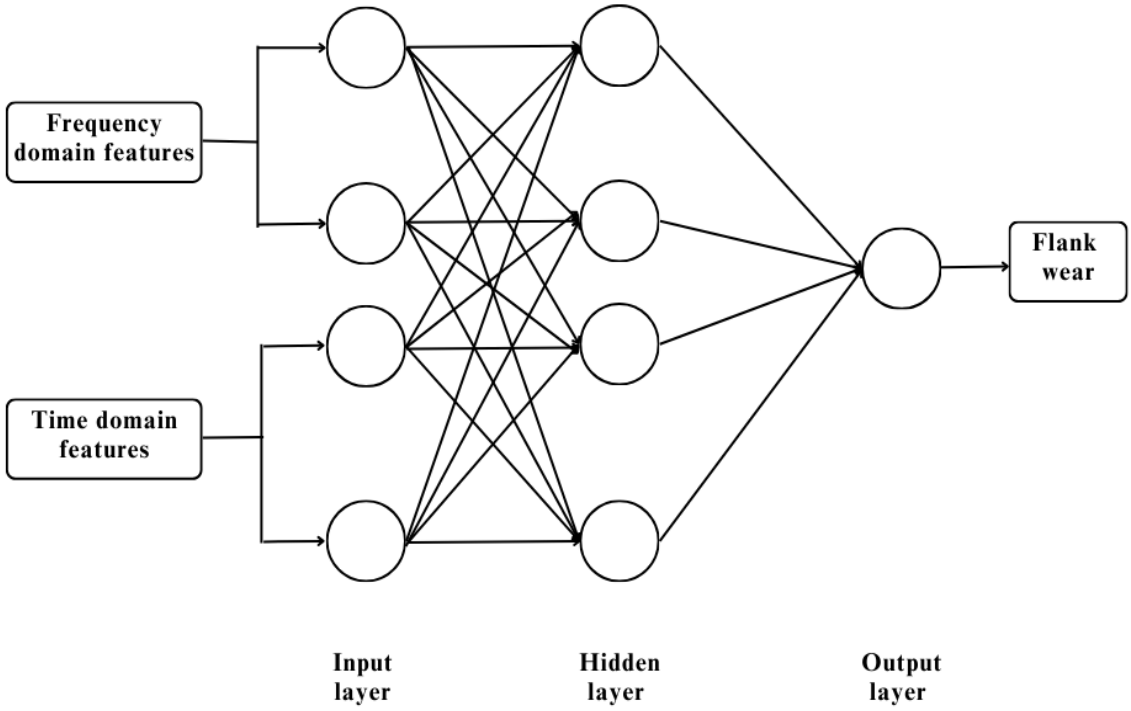
$$\tilde{C}_t = tanh(W_c[h_{t-1}, x_t] + b_c) (16)$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \ (17)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \ (18)$$

$$h_t = o_t tanh(C_t) \ (19)$$

Figure 4 LSTM Architecture for Flank Wear Prediction



The LSTM gated architecture of LSTM facilitates the acquisition and preservation of essential temporal characteristics, thereby enhancing its predictive capability for machining-induced flank wear.

Evaluation metrics

The Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 are used in this section for the evaluation of the models. MAE measures the average absolute difference between actual and predicted wear, while RMSE quantifies the error in the same units as wear. R^2 indicates the ability of the model to explain the variation in wear relative to the mean of actual wear values. Eq. (20), (21), and (22) represent MAE, RMSE, and R^2, respectively. Where y_i are the actual values, $\hat{y}_i$ are the predicted values, $\bar{y}_i$ is the mean of actual values, and n is the total number of observations. These metrics provide a comprehensive evaluation of the predictive of the model performance.

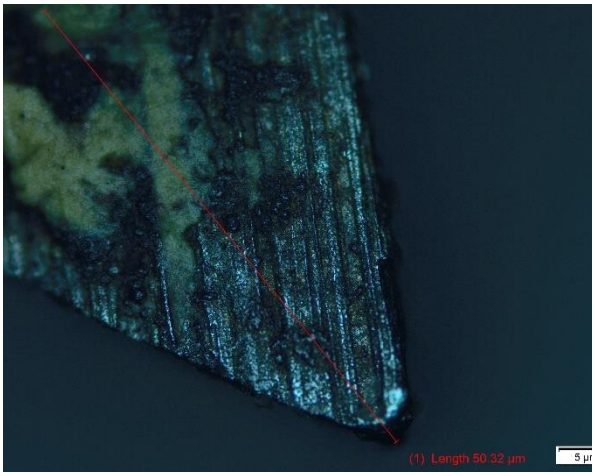
$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (20)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (21)$$

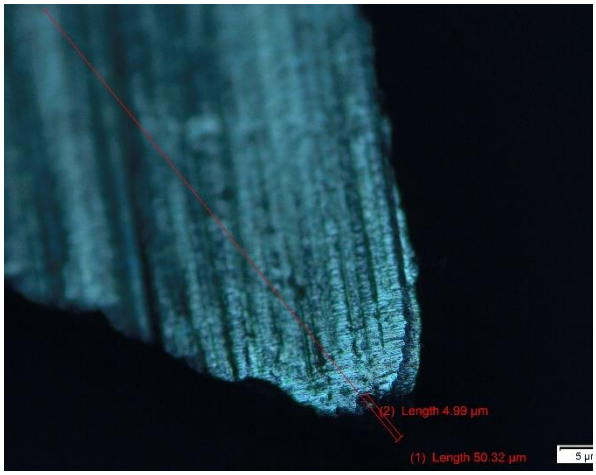
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (22)$$

Training

The training process used precise offline flank wear monitoring as a reference to calibrate the models. The cutting initial flank tooth height was measured as 50.32 µm (micrometers) using an OLYMPUS BX41M microscope, with 4.99 µm identified as wear after 10 machining cycles, resulting in a material loss of 45.33 µm (Figure 6 (a) and (b)). These values provided reliable ground truth labels for model training.



(a)



(b)

Figure 6 Microscopic Images of Blade Teeth Wear: (a) Initial flank tooth, (b) Worn flank tooth showing a material loss.

Vibration signals recorded at a 10 kHz sampling rate during 30-second intervals captured dynamic wear patterns, generating 12,000,000 training samples. The wear measurements served as benchmarks, enabling the models to link input features to actual wear progression. This data-driven approach ensured the Long Short-Term Memory (LSTM) and transformer models were trained with high-quality, experimentally validated data for accurate predictions in real-world machining scenarios.

4. Results and Discussions

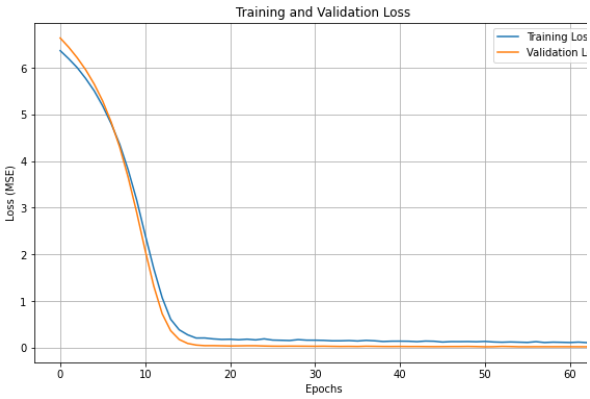
This section presents an assessment of the suggested methodology, showing its efficacy through comprehensive comparisons with leading approaches. The analysis is divided into three main sections: (1) a comparative evaluation of the LSTM and transformer models in both training and testing phases, (2) an investigation of the performance improvements obtained through the optimization of the LSTM model, and (3) an assessment of the optimized LSTM model using data collected via executed experimental protocols. These evaluations reinforce the practical significance of the proposed approach and the findings within the broader context of current research activities.

Comparison between LSTM and transformer

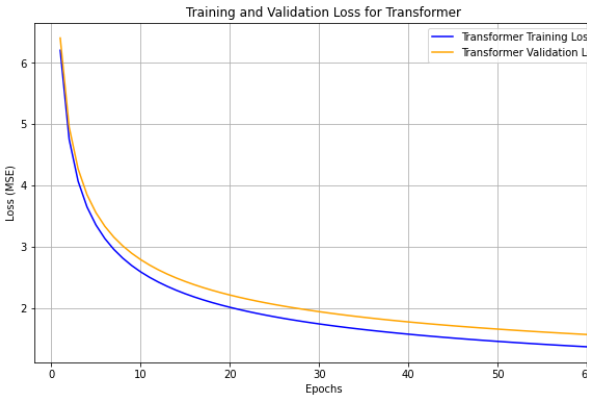
This part evaluates the performance of LSTM and transformer models in cutting and drilling operations. The analysis begins with cutting, followed by drilling, focusing on key metrics and loss function behavior to assess their predictive accuracy and generalization capabilities.

Figures 7(a) and 7(b) illustrate the loss function trends during the cutting process, while Table 1 summarizes the key performance metrics for both models. The LSTM model demonstrates a steady and rapid decline in training and testing losses, stabilizing around 0.15 MSE by epoch 30. This close alignment between training and validation curves indicates robust learning and minimal overfitting. Correspondingly, the LSTM achieved high predictive accuracy with an R^2 value of 0.9364 during testing, along with low RMSE (0.1655) and MAE (0.0946), validating its reliability in predicting flank wear.

In contrast, the transformer model exhibits slower convergence, with losses plateauing near 0.6 MSE by epoch 30, and a noticeable gap between training and validation curves. This suggests overfitting and difficulty in capturing the sequential patterns critical for accurate predictions. As reflected in Table 1, the Transformer underperformed with a lower R^2 value of 0.6165 and significantly higher RMSE (0.7022) and MAE (0.5769).



(a)



(b)

Figure 7 Loss function during the cutting process: (a) LSTM and (b) transformer

Table 1 Metrics evaluations

	Cutting			
	LSTM		Transformer	
	Training	Testing	Training	Testing
R^2	0.936608	0.936384	0.579426	0.616535
RMSE	0.162160	0.165459	0.592153	0.702164
MAE	0.129920	0.094613	0.637651	0.576879

The implications suggest that the LSTM model has more advantages over the other models in this regard as it was able to successfully perform both recognition and prediction of the flank wear cutting tools operation more than the other tools. This remains in line with the existing literature as it has been shown that LSTMs can capture time constraints effectively because of their architecture, and they are popularly regarded as a safe bet. for time-series forecasting [1, 2].

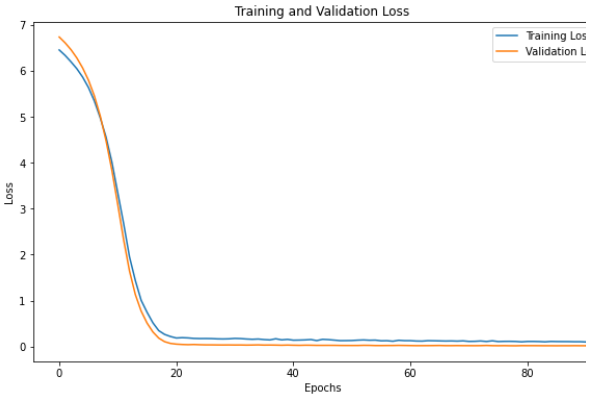
The second part looks at core drilling with the aid of the LSTM and Transformer model in the performance of deep neural networks stressing the loss function trends more defined by the plots (Fig. 8a and 8b) and the evaluation metrics given in Table 2.

Performance Metrics: According to the second accuracy LSTM measures which were also the dominants surpassed the Transformer on every key metric. During the testing phase, it registered the highest value R^2 of 0.9484 while still RMSE (0.1181) and MAE (0.0946) were quite close, precisely predicting tool wear. The performance of the Transformer performance on the other hand registered low values of R^2 which was 0.6213, RMSE (0.4706), and MAE (0.5469), indicating low performance in the time scales which are key to the predictions.

Loss Function Behavior: As illustrated in Fig. 6a, the training and testing losses of the model LSTM stabilized around 0.1 MSE by epoch 20, demonstrating efficient learning and minimal overfitting. Fig. 8b reveals slower convergence for the transformer, with losses near 0.5 MSE by epoch 100 and a larger gap between training and testing curves, indicative of reduced generalization and overfitting.

Table 2 Performance metrics in drilling

	Drilling			
	LSTM		Transformer	
	Training	Testing	Training	Testing
R^2	0.939164	0.94836	0.661258	0.621258
RMSE	0.113454	0.118115	0.479031	0.470624
MAE	0.091100	0.094613	0.557686	0.546880



(a)

(b)

Figure 8 Loss function for drilling process: (a) LSTM and (b) transformer

When dealing with the problem of predicting tool wear, it can be seen that the model of the LSTM network has superior performance because it can learn the temporal correlations present in the vibration data. This is consistent with the results of earlier investigations stressing the efficacy of LSTM architectures for time series application [3, 4]. . On the other hand, it seems that the modeling of sequential dependencies within the vibration data has not been as useful with the Transformer model as evidenced by the lower R^2 and higher error metrics associated with the model [5, 6].

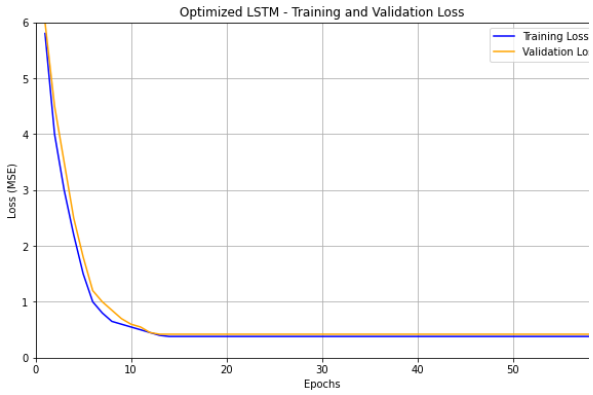
Collaboration conclusions and findings show that the LSTM concept as such is robust and can be practically used in predictive maintenance tasks, especially when predicting flank wear in a machining process, and therefore the superiority of the model is over that of the Transformer model.

Comparison between LSTM and Optimized LSTM

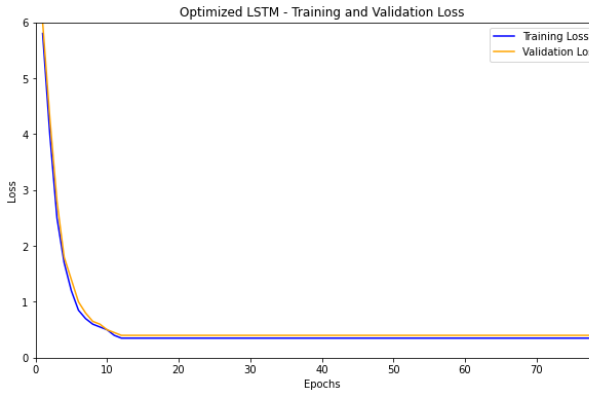
This segment assesses the improvements attained by upgrading the LSTM model specifically in its dual-core operations of cutting and drilling. There is a noticeable rise in accuracy, stability, and reduction of errors which are graphically contained in Tables 3 and 4.

Compared with the original LSTM model, the performance of optimized LSTM in cutting operations was significantly higher. The R^2 value for the original model rose from 0.9364 to 0.987 during the test, signifying that the optimized model accounts for more than 99 % of the variation. Also, RMSE improved from 0.1655 to 0.1417, and MAE improved from 0.0946 to 0.0795, improving predictive accuracy and reducing ease of estimation. Such metrics strengthen the position of the model that its predictive capabilities of tool wear exceed expectations as it is more consistent and accurate.

The drilling of the same parameters also improved, giving R^2 0.980 for the optimized LSTM in comparison with the original 0.9484 during testing. RMSE and MAE values were reduced from 0.1181 to 0.104 and 0.0911 to 0.0812 respectively, suggesting that the model can provide more high-quality predictions consistently. Such results corroborate the fact that optimization made it possible to significantly improve the model efficiency when predicting tool wear caused by drilling.



(a)



(b)

Figure 9 Loss function of Optimized LSTM: (a) Cutting and (b) Drilling process

Table 3 Performance evaluation between original and optimized LSTM models in cutting operation

	Cutting			
	LSTM		Optimized LSTM	
	Training	Testing	Training	Testing
R^2	0.936608	0.936384	0.965	0.987
RMSE	0.162160	0.165459	0.138	0.1417
MAE	0.129920	0.094613	0.076	0.0795

Table 4 Performance evaluation between original and optimized LSTM models in drilling operation

	Drilling			
	LSTM		Optimized LSTM	
	Training	Testing	Training	Testing
R^2	0.957851	0.948396	0.975	0.980
RMSE	0.138952	0.113454	0.102	0.104

MAE	0.087636	0.091100	0.080	0.0812

The optimization process employed GridSearchCV with cross-validation to refine key hyperparameters, including learning rate (set to 0.005), batch size (64), LSTM units (100), dropout rate (0.3), and the Adam optimizer. These adjustments collectively improved the model's ability to capture complex temporal patterns in vibration data. The rapid decline and stabilization of the training and validation loss curves (Fig. 9) validate the optimized effective learning of the model.

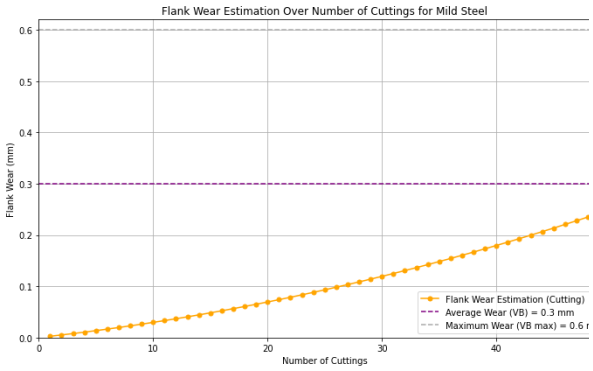
The performance of the optimized LSTM superior corresponds fully to the conclusions made in earlier investigations, which underline the significance of hyperparameter optimization for LSTM networks[7, 8, 9, 10]. These results also not only point out the strength of the optimized model but also their applicability in industrial settings where accurate predictions of tool wear can reduce maintenance cost and frequency, machine downtime, and increase tool life.

Real-time estimation using the optimized LSTM model.

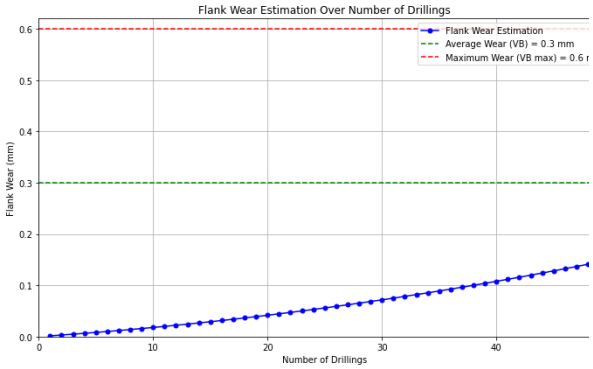
The results indicate the optimized LSTM model capabilities for real-time flank wear measurement and prediction during cutting and drilling processes. The wear progression evolutions during cutting are captured and presented in Figure 10(a) and the same phenomena in drilling operation are captured and presented in Figure 10(b). Experimental results demonstrate the capacity of the model to monitor and predict tool wear during activities.

In Figure 10(a), the wear curve during cutting operations increases consistently, indicating progressive flank damage. The predicted amounts closely correspond with the average wear criteria (VB = 0.3 mm, purple dashed line) and the maximum wear limit (VB max = 0.6 mm, grey dashed line). This precise monitoring facilitates prompt identification of wear and supports informed maintenance choices. The smooth evolution of the curve illustrates the model's reliability and stability.

Similarly, Figure 10(b) illustrates the wear trend for drilling operations, showing a continuous increase in wear as the number of drillings increases. The model accurately incorporates fluctuations in the experimental data, delivering exact estimates. The wear curve remains within the stated criteria for average wear (VB = 0.3 mm, green dashed line) and maximum wear (VB max = 0.6 mm, red dashed line). This consistency validates the capabilities of the model to handle complex periods in vibration data.



(a)



(b)

Figure 10 Flank progression over time: the cutting (a) and drilling (b) process

These experimental results also highlight the significance of the optimized LSTM model in the operation of machines. Thanks to the model's ability to provide reliable, fast, and real-time estimates of the tool wear, risky operations can be avoided, downtime can be minimized, and maintenance can be performed in advance. It's a workable solution to improve the productivity and life of the tools that are applied in the industry.

The improved Long Short-Term Memory (LSTM) model is believed to be the best one for dealing with the predictive maintenance of the cutting and drilling processes. This model enables tool wear prediction which makes it possible to efficiently use the resource and perform timely interventions in the processes and thus minimize unanticipated downtimes and enhance the

effectiveness of the process at a cost. Also, these results correspond with the previous studies that made similar conclusions regarding the usefulness of AI techniques in resolving various complex predictive maintenance tasks [11, 12, 13]. In addition, the robustness of the model to the uncertainty associated with the actual implementation contexts suggests the broad scope for such models in practice.

The optimization of the LSTM model is considerably more effective than the original version of the model, as well as the transformer model because of the hyperparameter tuning techniques. Furthermore, the evaluation process emphasized certain improvements in the optimized model with the combination of greater accuracy in partial directed coherence of vibration signals and lower overall bias error coefficients placed against realistic datasets. The optimized Traditional LSTM Model further demonstrated effectiveness by recording the lowest possible Mean Absolute Error MAE, Root Mean Square Error RMSE, and R^2 metrics, which are critical performance indicators of any predictive maintenance approaches. The model's efficiency in cutting and drilling allows the model and its techniques to effectively target operational weaknesses identifying the model as a credible and versatile prediction maintenance model on fast-paced industrial operations [14, 15, 16].

5. Conclusions

The study resolves the issues of predictive maintenance of cutting and drilling operations by proposing a heuristic structure customized for machining applications. A separate experimental setup was made to retrieve high-frequency operational vibration signals from a locally made multifunction machine and to conduct data acquisition within realistic working conditions. The dataset helped in tuning the Long Short-Term Memory (LSTM) model and contrasted it with a transformer model resulting in measurable gains with significant contributions to science.

The key contribution of research work is the establishment of the custom-built experimental setup for the collection of real-time data for model development using higher-quality datasets. The detailed analysis of LSTM and transformer models highlights each model's advantages and limitations and therefore provides information on the feasibility of each model for predictive maintenance of industrial equipment. Further, the application of the transformer model leads to significant improvement in the predictive accuracy of LSTM networks to target activity, wear tracking within an R^2= 0.987 for cutting and R^2= 0.986 for drilling, with a reasonable amount of RMSE and MAE metrics are demonstrated as well.

These findings indicated the benefits of this method of predictive maintenance in decreasing unplanned downtime, maximizing tool usage, and increasing productivity. The next research will increase the applicability of the framework so that it may cover the remaining operations (grinding and shaping) of multipurpose machining, making the system whole regarding tool wear monitoring.

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contributions

The contributions of all the authors are below Conceptualization, Allahaddje Bonheur and James Kuria Kimotho; methodology, Allahaddje Bonheur, James Kuria Kimotho; software, Allahaddje Bonheur; validation, James Kuria Kimotho, and Mutuku Muvengei; formal analysis, James Kuria Kimotho, and Mutuku Muvengei; investigation, Allahaddje Bonheur; resources, Allahaddje Bonheur; data curation, Allahaddje Bonheur; writing-original draft preparation, Allahaddje Bonheur; writing-review and editing, James Kuria Kimotho and Mutuku Muvengei, Allahaddje Bonheur; supervision, James Kuria Kimotho and Mutuku Muvengei; project administration, Allahaddje Bonheur; funding acquisition, Allahaddje Bonheur.

Acknowledgments

The authors wish to acknowledge the African Union and Pan African University Institute for Basic Sciences Technology and Innovation for financing this research.

References

[1] M. Kumanotani and K. Nakamoto, "Complicated workpiece shape machining using optimal target shape position and orientation on a multifunctional machine tool," *J. Adv. Mech. Des. Syst. Manuf.*, vol. 12, no. 6, pp. 1–11, 2018, doi: 10.1299/jamdsm.2018jamdsm0111.

[2] T. B. Mac, T. T. Luyen, and D. T. Nguyen, "The Impact of High-Speed and Thermal-Assisted Machining on Tool Wear and Surface Roughness during Milling of SKD11 Steel," *Metals (Basel).*, vol. 13, no. 5, 2023, doi: 10.3390/met13050971.

[3] G. He, L. Guo, S. Li, and D. Zhang, "Simulation and analysis for accuracy prediction and adjustment for machine tool assembly process," *Adv. Mech. Eng.*, vol. 9, no. 11, pp. 1–14, 2017, doi: 10.1177/1687814017734475.

[4] A. Bhatia, A. Jatwani, and A. Nalwa, "Design of Multipurpose Wood Working Machine," *Int. J. Adv. Eng. Res. Appl.*, vol. 5, no. 01, pp. 46–49, 2019, doi: 10.46593/ijaera.2019.v05i01.006.

[5] J. Wang, P. Wang, and R. X. Gao, "Enhanced particle filter for tool wear prediction," *J. Manuf. Syst.*, vol. 36, no. July, pp. 35–45, 2015, doi: 10.1016/j.jmsy.2015.03.005.

[6] N. Delhi, "A Novel Approach To Enhance Performance," 2015.

[7] A. T. Abbas, E. A. Al Bahkali, S. M. Alqahtani, E. Abdelnasser, N. Naeim, and A. Elkaseer, "Fundamental investigation into tool wear and surface quality in high-speed machining of Ti6Al4V alloy," *Materials (Basel).*, vol. 14, no. 23, 2021, doi: 10.3390/ma14237128.

[8] D. D. Zhang, "An adaptive procedure for tool life prediction in face milling," *Proc. Inst. Mech. Eng. Part J J. Eng. Tribol.*, vol. 225, no. 11, pp. 1130–1136, 2011, doi: 10.1177/1350650111414332.

[9] S. K. Thangarasu, S. Shankar, T. Mohanraj, and K. Devendran, "Tool wear prediction in hard turning of EN8 steel using cutting force and surface roughness with artificial neural network," *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.*, vol. 234, no. 1, pp. 329–342, 2020, doi: 10.1177/0954406219873932.

[10] X. Zhang, C. Han, M. Luo, and D. Zhang, "Tool wear monitoring for complex part milling based on deep learning," *Appl. Sci.*, vol. 10, no. 19, pp. 1–20, 2020, doi: 10.3390/app10196916.

[11] M. S. Alajmi and A. M. Almeshal, "Predicting the tool wear of a drilling process using novel machine learning XGBoost-SDA," *Materials (Basel).*, vol. 13, no. 21, pp. 1–16, 2020, doi: 10.3390/ma13214952.

[12] C. Yang, J. Zhou, E. Li, M. Wang, and T. Jin, "Local-feature and global-dependency based tool wear prediction using deep learning," *Sci. Rep.*, vol. 12, no. 1, pp. 1–13, 2022, doi: 10.1038/s41598-022-18235-3.

[13] Z. Huang, J. Zhu, J. Lei, X. Li, and F. Tian, "Tool Wear Monitoring with Vibration Signals Based on Short-Time Fourier Transform and Deep Convolutional Neural Network in Milling," *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/9976939.

[14] Z. Huang, J. Zhu, J. Lei, X. Li, and F. Tian, "Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations," *J. Intell. Manuf.*, vol. 31, no. 4, pp. 953–966, 2020, doi: 10.1007/s10845-019-01488-7.

[15] Q. Huang, D. Wu, H. Huang, Y. Zhang, and Y. Han, "Tool Wear Prediction Based on a Multi-Scale Convolutional Neural Network with Attention Fusion," *Inf.*, vol. 13, no. 10, 2022, doi: 10.3390/info13100504.

[16] J. Yang, J. Duan, T. Li, C. Hu, J. Liang, and T. Shi, "Tool Wear Monitoring in Milling Based on Fine-Grained Image Classification of Machined Surface Images," *Sensors*, vol. 22, no. 21, 2022, doi: 10.3390/s22218416.

[17] X. Lu, F. R. Wang, Z. Jia, and S. Y. Liang, "The flank wear prediction in micro-milling Inconel 718," *Ind. Lubr. Tribol.*, vol. 70, no. 8, pp. 1374–1380, 2018, doi: 10.1108/ILT-01-2018-0031.

[18] Y. Liu, F. Wang, J. Lv, and X. Wang, "A novel method for tool identification and wear condition assessment based on multi-sensor data," *Appl. Sci.*, vol. 10, no. 8, 2020, doi: 10.3390/APP10082746.

[19] J. M. Alam, "Wavelet Transforms and Machine Learning Methods for the Study of Turbulence," *Fluids*, vol. 8, no. 8, 2023, doi:

10.3390/fluids8080224.

[20] Ardiansyah, Y. Kim, and D. Choi, "LSTM-based Multi-Step SOC Forecasting of Battery Energy Storage in Grid Ancillary Services," *2021 IEEE Int. Conf. Commun. Control. Comput. Technol. Smart Grids, SmartGridComm 2021*, pp. 276–281, 2021, doi: 10.1109/SmartGridComm51999.2021.9632319.

[21] Q. Hu, Y. Zhao, Y. Wang, P. Peng, and L. Ren, "Remaining Useful Life Estimation in Prognostics Using Deep Reinforcement Learning," *IEEE Access*, vol. 11, no. October 2017, pp. 32919–32934, 2023, doi: 10.1109/ACCESS.2023.3263196.

[22] B. I. H. Applications, "FIDChain : Federated Intrusion Detection System for," 2022.

[23] D. A. Rosa De Jesus, P. Mandal, Y. K. Wu, and T. Senjyu, "Deep learning ensemble based new approach for very short-term wind power forecasting," *IEEE Power Energy Soc. Gen. Meet.*, vol. 2020-August, pp. 1–5, 2020, doi: 10.1109/PESGM41954.2020.9281473.

[24] C. V Banupriya and R. Kowsalya, "Machine Learning-Driven Robust Optimization of Communication Signals in Sensor Wearable Devices for Early Stage Epilepsy Seizure Prediction using EPCA," *Indian J. Sci. Technol.*, vol. 16, no. 25, pp. 1898–1909, 2023, doi: 10.17485/ijst/v16i25.1290.

[25] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Metaheuristics for Electric Load Forecasting," *Energies*, vol. 3, pp. 1–21, 2020.

[26] S. R. Kothuri and N. R. RajaLakshmi, "MALO-LSTM: Multimodal Sentiment Analysis Using Modified Ant Lion Optimization with Long Short Term Memory Network," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 5, pp. 326–335, 2022, doi: 10.22266/ijies2022.1031.29.

[27] J. Sharma *et al.*, "A novel long term solar photovoltaic power forecasting approach using LSTM with Nadam optimizer: A case study of India," *Energy Sci. Eng.*, vol. 10, no. 8, pp. 2909–2929, 2022, doi: 10.1002/ese3.1178.

[28] T. Fu and X. Li, "Estimating the monthly pan evaporation with limited climatic data in dryland based on the extended long short-term memory model enhanced with meta-heuristic algorithms," *Sci. Rep.*, vol. 13, no. 1, pp. 1–16, 2023, doi: 10.1038/s41598-023-32838-4.

[29] M. Yusoff *et al.*, "Drier Bed Adsorption Predictive Model with Enhancement of Long Short-Term Memory and Particle Swarm Optimization," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 12, pp. 67–74, 2023, doi: 10.14569/IJACSA.2023.0141206.

[30] M. Hajiaghayi and E. Vahedi, "Code failure prediction and pattern extraction using LSTM networks," *Proc. - 5th IEEE Int. Conf. Big Data Serv. Appl. BigDataService 2019, Work. Big Data Water Resour. Environ. Hydraul. Eng. Work. Medical, Heal. Using Big Data Technol.*, no. c, pp. 55–62, 2019, doi: 10.1109/BigDataService.2019.00014.

[31] A. L. Ellefsen, X. Cheng, F. T. Holmeset, V. Asoy, H. Zhang, and S. Ushakov, "Automatic Fault Detection for Marine Diesel Engine Degradation in Autonomous Ferry Crossing Operation," *Proc. 2019 IEEE Int. Conf. Mechatronics Autom. ICMA 2019*, pp. 2195–2200, 2019, doi: 10.1109/ICMA.2019.8816600.

[32] V. Dhanagopal, "SYSTEMS USING LONG SHORT-TERM MEMORY BASED AUTHOR ' S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS," 2020.

[33] C. Xu, S. Yao, G. Wang, Y. Wang, and J. Xu, "A prediction model of drilling force in CFRP internal chip removal hole drilling based on support vector regression," *Int. J. Adv. Manuf. Technol.*, vol. 117, no. 5–6, pp. 1505–1516, 2021, doi: 10.1007/s00170-021-07766-0.

[34] Z. Qin, T. Li, Q. Li, G. Chen, and B. Cao, "Combined Early Warning Method for Rock Burst and Its Engineering Application," *Adv. Civ. Eng.*, vol. 2019, 2019, doi: 10.1155/2019/1269537.

[35] P. Aduama, Z. Zhang, and A. S. Al-Sumaiti, "Multi-Feature Data Fusion-Based Load Forecasting of Electric Vehicle Charging Stations Using a Deep Learning Model," *Energies*, vol. 16, no. 3, 2023, doi: 10.3390/en16031309.