Article citation info:
Wu Y, Chen B. Real time diagnosis and prediction of tool wear in mechanical machining based on deep learning. Diagnostyka. 2026;27(1):2026113. https://doi.org/10.29354/diag/219836.

DIAGNOSTYKA, 2026, Vol. 27, No. 1

e-ISSN 2449-5220
DOI: 10.29354/diag/219836

REAL TIME DIAGNOSIS AND PREDICTION OF TOOL WEAR IN MECHANICAL MACHINING BASED ON DEEP LEARNING

Yuegui WU, Bailiang CHEN *

School of Mechatronics Engineering, Yangjiang Polytechnic, Yangjiang 529566, Guangdong, China
* Corresponding author's E-mail address: chenbailiang86@outlook.com

Abstract

Tool wear has a substantial impression on machining precision, operational stability, and production efficiency in mechanical manufacturing. Even though tool wear status detection and forecasting have been the subject of lots of research, the majority of current methods lack a single, real-time diagnostic capacity that combines both prediction and instantaneous recognition. To address this limitation, a novel deep learning (DL)-based framework is proposed for real-time diagnosis and prediction of tool wear, utilizing a customized architecture named Dynamic Gravitational-tuned Gate Adjusted Long Short-Term Memory (DG-GA-LSTM). Experimental datasets were acquired from a controlled Computer numerical control (CNC) milling setup, gathering multi-channel sensory data from cutting force, vibration, and acoustic emission sensors under varying operational conditions. The acquired signals underwent noise reduction and normalization using Robust Locally Weighted Regression (RLWR) to ensure consistent input quality. The proposed model was implemented in Python and achieved high performance across various evaluation metrics, including MAE (0.0068mm), RMSE (0.0094 mm), R^2 (0.9967), Precision (94.68%), and Recall (90.47%) when compared with other existing approaches. This model demonstrates the feasibility and effectiveness of implementing real-time, intelligent diagnostics and prediction for tool wear, contributing to the advancement of predictive maintenance in modern manufacturing environments.

Keywords: tool wear prediction, real-time diagnosis, dynamic gravitational-tuned gate adjusted long short-term memory (DG-GA-LSTM), smart manufacturing, mechanical machining.

1. INTRODUCTION

The worldwide manufacturing industry is moving toward intelligent and digital growth, which is crucial for boosting the industry's financial gains [1]. The amount of tool wear has a significant effect on the component's external cleanliness and production performance since tools are the last operational elements in machining processes. The tool wears down excessively before failing due to the pressure, resonance, thermal stress, and other aspects of the production process [2]. The main focus of tool condition monitoring has been tool wear monitoring. There have been several monitoring models set out that are appropriate for various production situations; these individuals can be brought up as data-driven and physics-based approaches [3]. By utilizing previous mechanical and physics knowledge of the cutting processes, the physics-based approach allows for the establishment of the relationship between tool wear and physical characteristics for tool condition monitoring [4]. Information technology and the contemporary industrial sector are closely intertwined to produce high-quality parts. The manufacturing sector requires metal cutting, and to manage the quality of the machined surface, there is a growing need to track and forecast machining parameters [5]. Surface roughness, a crucial metric in a system for assessing surface quality, is influenced by a number of variables, including both controllable ones like fundamental cutting settings and some uncontrolled ones like material heterogeneity, cutting vibration, and tool wear [6].

As the primary key subsystems of CNC machine tools, which are now a crucial piece of fundamental processing equipment in contemporary industrial production, machining centers are widely utilized in the aerospace, automotive, aviation, and other technologically advanced manufacturing sectors [7]. The mechanical construction of CNC machine tools is growing increasingly complicated to fulfil the production needs of the contemporary manufacturing industry. Considerable resources have been consumed due to the reliance on periodic maintenance and replacement due to the challenges in diagnosing CNC machine equipment [8]. Maintaining machining accuracy, creating high-

Received 2025-08-06; Accepted 2026-03-20; Available online 2026-03-27

quality products, and reducing production costs all depend on the ability to predict tool wear. In nonlinear wear mechanisms, such as those in materials such as aluminum alloys, recent developments have put a heavy emphasis on combining both image processing and machine learning methods to forecast tool wear mechanisms [9]. The supervisory systems make possible more flexible and intelligent production tactics, and have demonstrated positive results in real time predicting tool deterioration [10].

1.1. Research Objective

The research presents a DL-based system that can be both used to diagnose and predict tool wear in real time. It uses the DG-GA-LSTM architecture which overcomes the shortcomings of other current approaches which process these tasks independently. Multi-sensor data integration, powerful multi-sensor preprocessing, and feature extraction are used to provide correct low-latency predictions needed in intelligent machining and predictive maintenance.

1.2. Research Contribution

This research suggested a new hybrid DL model that enhances the predictability of tool wear prediction, and is computationally proficient. The addition of dynamic optimization and adaptive gating programs contributes to an increased number of strong performance in the real-time machining environments.

DG-GA-LSTM Framework: Proposed a dynamic Gravitational-tuned Gate Adjusted LSTM model which dynamically alters the flow of memory to provide the correct estimation of the tool wear when a system is operating under different conditions.

Hybrid Optimization Mechanism: Added a DG Algorithm to optimization of learning sensitivity and faster convergence with less time spent on execution and better generalization.

Multi-Sensor Fusion: used multi-Sensor (leveraged acoustic emission, vibration, cutting force) feature integration to classify the wear state in a comprehensive way and exhibit better performance results compared to traditional DL models.

2. RELATED WORK

The research [11] suggested a technique that would be applicable in the accurate prediction of the hob wear status based on real-time CNC monitoring data of machining of worm gears. The method employs an orthogonal test design to reflect the tool wear course. The transfer learning based augmented deep belief network is continuously modified to enhance speed and accuracy of training. The technique has better mean-squared error rates than traditional backpropagation networks, however, its ability to handle extreme wear conditions or invisible machine classes is not so high. A DL-based system was released to improve predicting multi-step as well as tool wear in diverse machining conditions [12]. The framework adopts an integrated architecture with a combination of Convolutional Neural Network (CNNs) and the BiLSTM networks that are based on the spatial-temporal correlation. To predict the short- and long-term tool wear, a dense residual network (ResNetD) was proposed. The efficiency and resiliency of the model were experimentally valid, but the model may be ineffective in dynamic or covert work conditions.

The LSTM-based DL model has been proposed to enhance the online machine tool wear diagnosis and prediction accuracy [13] It has the ability to generalize on tools and contexts, employing long-term relationships, and no prior knowledge of the system. Milling tools were experimentally validated with less MSE. Nevertheless, the operating noise and sensor quality may vary, which might influence the performance of the model, and scalability is not an easy issue to reach in dynamic industrial environments. A cutting tool provided with a thin-film thermocouple in it was used to predict tool wear [14]. A backpropagation neural network and stacked sparse autoencoder were tested and increased prediction accuracy by sparsity and weight penalties. This approach was better than the conventional approaches and was more robust and generalized. Necessity to use temperature could, however, restrict flexibility on varying machining circumstances.

As compared to the traditional machine learning models, the research [15] minimized machining costs and scrap by 36.53 percent through the introduction of a WPT-based tool wear monitoring solution in real time. The technique operates with acoustic emission signals of complex data, but preprocessing and feature selection are very extensive. The high wear of the tools in milling TC18 alloy impacts on the quality of the surface and leads to high costs of production. In the research [16], there is a real-time control method that involves CNN-BILSTM, CNN- (BI Grated Recurrent Unit) BIGRU, and based on cutting force measurements, a prediction of the wear process in the tool is presented. High accuracy was established with the DL models having prediction errors of less than 8%. Performance may however be different in some different cutting conditions or tool geometries. More confirmation in various machining scenarios should be done to allow larger applicability.

DeepTool predicts milling tool wear and life by measuring sensor data on real cutting experiments in the real world [17]. It integrates encoder-decoder LSTM and hybrid autoencoder-LSTM networks in order to model the patterns of tool deterioration. The system has prediction reliability by rich feature extraction process. Findings indicate a R2 of more than 95 per cent in wear onset detection and life estimation. It requires additional confirmation in a variety of industrial configurations. A real-time system can improve the monitoring of the condition of the tools by providing an AE root mean square signal[18] to identify sudden tool failure. LSTM autoencoder is used to extract and compress time-

frequency features in the system to detect anomalies. Deviation during reconstruction is used as an indicator of pre-failure. Its strength was also experimentally proven, but its performance at extreme conditions of chip adhesion-separation under machining of ultra-hard material requires further research.

A DL model named Sequence-to-Sequence Monotonicity-Aware Meta-Learning (SMAML) [19] is a DL model used to enhance the monitoring process of tool wear, using the degradation pattern of the cutting process. It employs a sequence-to-sequence design, monotonicity loss that is unique and the attention methods to simulate the wear progression well. SMAML is more accurate and predictive on real world CNC machine datasets compared to other models. Nonetheless, its applicability to other machining conditions is still low, and further research should be done to ensure it is applicable to other types of tools and materials. To predict tool wear in CFRP milling, a hybrid DL model is applied, which is a multi-channel 1D CNN and LSTM [20]. The model then takes dynamic force signal inputs of milling experiments at different angles between the fibers, feed rates and cutting speeds. It has MAE of 2.94 μm and R2 of 95.04 percent, which gives an improvement of more than 25 percent in its accuracy over the traditional versions. But performance can be less in changing signal of unstable forces or multi-directional laminates.

The research [21] aimed to optimize Computer Numerical Control (CNC) machining center processing parameters to achieve energy saving and cost reduction in manufacturing operations. It presented a multi-objective optimization approach integrating process parameter optimization with energy consumption and production cost evaluation. To develop a data-driven method for automated machine condition monitoring and cutting tool wear prediction using vibration sensor signals [22]. It presented a Long Short-Term Memory based Recurrent Neural Network (LSTM-RNN) model to learn system transition and observation functions for tool wear estimation and remaining useful life prediction to precisely determine the state of cutting tool wear for bettering mechanical processing yield and machining precision [23]. It presented a Convolutional Neural Network (CNN) model that analyzed two-dimensional images generated from force signals using Continuous Wavelet Transform (CWT), Short-Time Fourier Transform (STFT), and Gramian Angular Summation Field (GASF) methods.

2.1. Research Gap

The previous studies have developed the DL-based tool wear prediction, though a significant number of them have problems with unstable accuracy throughout the wear stages, overdependence on some types of sensors, and lack of adaptability to altering cut states. More often than not, these models require a great deal of human tuning or do not work well in the dynamic world. The suggested DG-GA-LSTM methodology will overcome these constraints and combine a dynamic gravitational search scheme with a gate-modulated LSTM framework. Such integration increases flexibility, learning effectiveness, and predictability consistency when conditions of real-time CNC machining are met.

3. TOOL WEAR IN CNC MACHINE

Tool wear is a process of development that takes place in CNC machining when cutting tool becomes deformed due to constant heat and mechanical contact with component material. This wear ultimately causes a decrease in the quality of the machined part in terms of its overall quality, surface integrity and dimensional accuracy. The machine's stability and performance may be endangered by the ongoing rise in cutting forces, temperature, and vibration brought on by gradual tool deterioration. As seen in Figure 1, CNC maintenance operations monitor tool degradation.

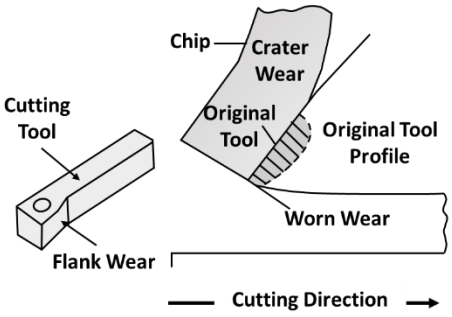

Fig. 1. Tool Wear Mechanisms in CNC Machining

The four main types of tool wear are built-up edge (BUE), notch, crater and flank wear. Flank wear is also associated with sticky and abrasive processes and it commonly occurs in the clearance face of the tool. Crater wear occurs on the rake face due to high temperatures and diffusion of the material whereas notch wear is observed near the depth-of-cut line due to repeated mechanical and thermal stress. In the presence of particular cutting conditions that may cause the workpiece material to adhere to the tool edge, accumulated edge may occur and result in unequal wear behavior and poor quality of the surface.

The extent of the reduced rate of feeding is one of the variables that affects the rate and pattern of tool corrosion, the rate of the blade and the state of lubrication as well as the hardness and abrasiveness of the workpiece material. Uncontrolled tool wear not only reduces the life cycle of the tool but also affects the final product with defects, makes it expensive but also leads to unexpected failure of the tools that causes more unplanned downtimes. Conventionally, monitoring tool wear has been

dependent on either the manual examination approach or the indirect approach which encompasses the examination of surface roughness or dimensional variations after machining. Nevertheless, they tend to be responsive, do not allow real time operation, and may fail to counteract the interruptions of the processes or quality decline.

These restrictions are raising the demand of smart systems that are capable of tracking and predicting tools wear in real time. In a modern production environment, it assists in increasing productivity and cost-efficiency as well as aiding predictive maintenance methods.

4. METHODOLOGY

This part explains the entire design constructed to predict tool wear in CNC machining environment. The proposed approach entailed signal preprocessing, feature extraction, model construction and optimization. Raw sensor data such as spindle power, vibration and motor current is initially preprocessed in the removal of noise and normalized to ensure consistency. In order to extract time as well as frequency domain information, feature extraction is carried out using features like WPT. The DG-GA-LSTM prediction model is designed in a manner to address time dependencies and enhance accuracy of learning. The process diagram of the applied method is shown in Figure 2.

4.1. Data Acquisition

The multi-sensor tool wear data was collected from an open-source dataset called Kaggle: https://www.kaggle.com/datasets/ziya07/multi-sensor-cnc-tool-wear-dataset/data. This dataset was created for CNC mechanical machining to monitor and anticipate tool wear in real time. It includes tool wear labels and multi-sensor assessments of tool operation, allowing for both classification and regression tasks for industrial research and predictive maintenance.

To guarantee both methodological validation and real-world applicability, this research adopts a hybrid experimental framework. The publicly available Kaggle multi-sensor CNC tool wear dataset is used as the primary offline dataset for model training, validation, and benchmarking of the proposed DG-GA-LSTM model Moreover, the research describes an OPC UA-based CNC monitoring architecture, which is the real-life industrial context where machining sensor signals, including vibration, spindle power, and cutting current, can be the gathered ones in real-time. This is an interim method that will close the gap between offline data-driven model development and real-time smart manufacturing deployment cases.

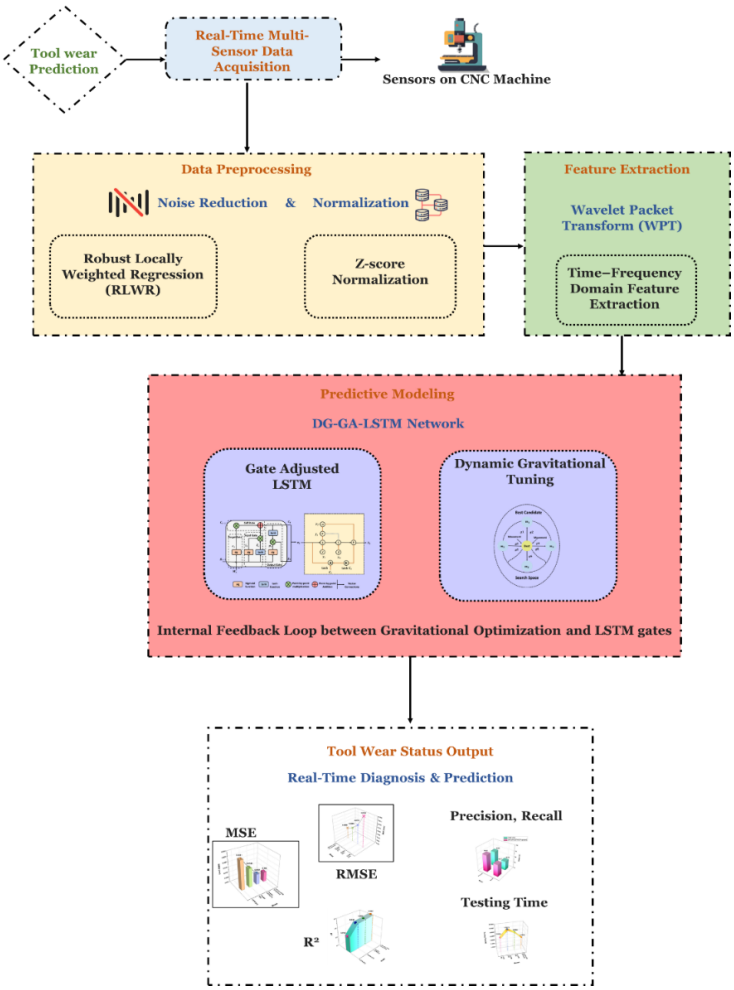

Fig. 2. Flow of the proposed approach

4.2. Data Preprocessing

Good and steady quality of the incoming signals should be ensured through efficient data preparation and then fed into the prediction model. In this research, RLWR is applied to eliminate noise in raw multi-sensor data due to outliers and variations. Standardization of data through Z-score normalization after the denoising step makes each of the features contribute equally and increases the convergence of the model. These preprocessing methods directly enhance the accuracy of tool wear diagnosis and forecast and also enhance the reliability of the data derived.

4.2.1. Z-score Normalization

A critical preprocessing step is signal normalization, which makes that input characteristics from various sensors are on the same scale. The preprocessed data is standardized using Z-score normalization prior to feature extraction and model training. DL models can converge more effectively during training when the data distribution is transformed to have a 0 mean and 1 variance using Z-score normalization, sometimes referred to as standard score normalization. This approach is particularly important in time-series machining data, where raw sensor values can vary significantly in scale and range depending on tool-material interaction, spindle speed, and depth of cut. The transformation is mathematically defined in eq. 1.

$$y_j = \frac{w_j - \mu}{\sigma} \quad (1)$$

Where $w_j$is the raw data value, $\mu$is the mean of the dataset (feature-wise), the standard deviation is represented by σ, and y_j is the normalized value. Features with wide numerical ranges cannot influence the learning process due to this standardization. Gradient-based optimizers perform better due to consistent scale. The model becomes less sensitive to numerical instabilities during training.

In this research, Z-score normalization is applied independently to each sensor channel following RLWR-based denoising. This two-stage preprocessing pipeline robust outlier suppression via RLWR, followed by distribution standardization via Z-score normalization, provides a stable, clean, and well-scaled input for the downstream DG-GA-LSTM model. In real-time machining situations, this combination greatly improves the accuracy and dependability of predictive modeling and wear on tools state forecasts.

Clarification on Sensor Signal Representation and Normalization

In this research, raw sensor signals obtained from spindle power, vibration, and motor current are not directly converted into a single physical quantity. Instead, each signal is processed to extract wear-relevant representations based on its physical meaning and sensing mechanism. Cutting force (CF) is indirectly estimated from spindle power and motor current measurements, as spindle power is proportional to cutting force under constant spindle speed and tool geometry conditions. The relationship has been used extensively in indirect force monitoring when direct force sensors are necessary.

Vibration (Vib) and acoustic emission (AE) sensors are considered to be independent measurements in the sensor domain. The results of the accelerators are vibration data and acoustic emission characteristics are obtained as high-frequency signal components at the moment of machining. The signals are not converted to force but rather processed using timefrequency feature extraction to be able to capture wear-induced dynamic behavior.

Signal normalization does not mean the standardization of various physical units. The denoised sensor channels are normalized using Z-score, which ensures that the features have zero-mean and unit variance. This step brings the numbers into parity without affecting the underlying physical properties of every signal, which enables combination of multi-sensors in the fusion of DL models without dimensional bias in the training of DL models.

4.2.2. Robust Locally Weighted Regression (RLWR)

With signal processing in time-series machining, the data collected on a sensor is prone to be contaminated by noise and outliers because of sudden cut condition or hardware interference. These deviations may have a serious impact on the performance of downstream learning algorithms. Hence, an efficient way of smoothing and normalizing is needed in order to make sure that it has good features extraction and model training.

An informal technique for regression is locally weighted regression (LWR) that is used to approximate local linear regression models using the data within a particular area. Nonetheless, standard LWR is also very sensitive to the outliers because it makes use of the least squares criterion. This is even more exaggerated in local models, where an abnormal point even just one point may have a disproportionately large effect on the regression output, because the subset size of each local fit is lower.

In an attempt to overcome this weakness, RLWR posits a robustness mechanism, which mitigates the effect of outliers during the model fitting process. This is done by adding a bisquare weighting component on the basis of a residual and this increases the rate of penalization on gross deviations. The bisquare weight (z_j) for each observation is computed in equation 2.

$$z_j = \begin{cases} \left(1 - \left(\frac{q_j}{d.\mathrm{MAD}}\right)^2\right)^2 & \text{if } |q_j| < d.MAD \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Here, q_j is the residual of the j^{th} point from the initial fit,and MAD i is the median absolute deviation of the residuals that was represented in equation 3. d

is a tuning constant, determining the sensitivity to outliers.

$$\mathrm{MAD} = \mathrm{median}\left(\left|q_j - \mathrm{median}(q_j)\right|\right) \quad (3)$$

The overall RLWR approach minimizes the weighted least squares' objective in equation 4.

$$\sum_{j=1}^{m} x_j \cdot \left(z_j - \hat{z}_j\right)^2 \quad (4)$$

Here, indexes the elements (from 1 to m), where m is the entire set of elements. The variable x_j is a weight assigned to the j^{th}element, indicating its significance or reliability. z_j denotes the actual or observed value for the j^{th}instance, and $\hat{z}_j$ is the corresponding predicted or estimated value. The term $\left(z_j - \hat{z}_j\right)^2$ calculates the squared error for each instance, and multiplies it by x_j scales the error based on the importance of that instance.

This estimation is performed iteratively: Initial coefficients are estimated using LWR. Residuals and MAD are computed. Bisquare weights are updated. Coefficients are re-estimated using the updated weights. This iterative process continues until convergence, i.e., when changes in coefficients become negligible.

In this research, RLWR is applied during the preprocessing stage to denoise and normalize the sensor signals before feature extraction. This makes the input to the DG-GA-LSTM model resistant to noise fluctuations and sudden anomalies that are temporary, which increases the robustness of the tool wear diagnostic and prediction in real time setting.

4.3. Feature Extraction

In continuous tool wear modeling applications, sensor data related to machining, such as motion, chopping power, and sound radiation, are inherently non-linear and non-stationary. Such signals usually include temporary properties, abrupt spikes, and concentrated time-frequency data on the condition of degradation of the tool. Conventional methods of signal analysis, e.g. Fourier Transform, cannot describe these localized variations in time. The WPT is used in order to counter this in form of a powerful time-frequency domain feature extraction technique. WPT also builds on the above mentioned DWT by allowing the approximation and detail coefficients to be fully decomposed at every level, giving even frequency bandwidths across sub-bands. The uniformity can also be further analyzed to concentrate on high-frequency elements which is especially helpful in identifying subtle trends in the process of tool wear.

The signal is progressively processed through two low-pass and high-pass filters during WPT decomposition, producing a tree structure of sub-signals that correspond to various frequency bands. Mathematically, the WPT basis functions are defined in equation 5.

$$X_m^{i,l}(s) = 2^{i/2} X_m(2^i s - l) \quad (5)$$

Where:

i and $l$are scale and translation parameters,

$m$is the modulation or oscillation index,

s represents a continuous or discrete signal variable

$X_m(s)$represents the wavelet packet basis function.

This iterative decomposition continues until a predefined decomposition level L is reached, resulting in 2^Lnarrowband sub-signals that together preserve the complete time-frequency energy content of the original signal. The DG-GA-LSTM model learns complex temporal patterns linked to various tool wear phases according to these extracted characteristics, which operate as extremely informative inputs.

4.4. Dynamic Gravitational-tuned Gate Adjusted Long Short-Term Memory (DG-GA-LSTM)

The accurate modeling of tool wear progression over time requires deep temporal representation learning with adaptability to dynamic operating conditions. To meet this requirement, propose a novel sequence modeling architecture named DG-GA-LSTM. This model improves the classic LSTM network by embedding two biologically and physically inspired mechanisms:

- A Dynamic Gravitational Tuning unit that modulates learning rates and error sensitivity based on prediction deviation.
- A Gate Adjustment Module that controls the LSTM gates dynamically, allowing more refined memory flow across time steps.

This hybrid design greatly enhances learning efficiency, noise tolerance, predictive accuracy especially in the case of real-time, multi- sensor, non- stationary signals such as those obtained on a machining system.

4.4.1. Gate Adjusted Long Short-Term Memory

When relating to mechanical machining, tool wear changes over time and its precise prediction demands a model, which is able to capture temporal dependencies and dynamic trends across many sensor streams of data. Conventional feedforward neural networks are unable to store historical tendencies, which is essential in time-series situations such as the present. To address this weakness, LSTM networks are taken as the temporal modeling block forming the basis in the present research.

Recurrent neural networks Recurrent neural networks (RNNs) of the LSTM type were specifically designed to solve the vanishing gradient problem through the training of long-range connections in sequential inputs. In contrast to the simple RNNs, LSTMs include memory cells and gating mechanisms that enable the selective storage, updating and retrieval of the temporal information. The architecture of the adjusted LSTM of the gate is presented in Figure 3.

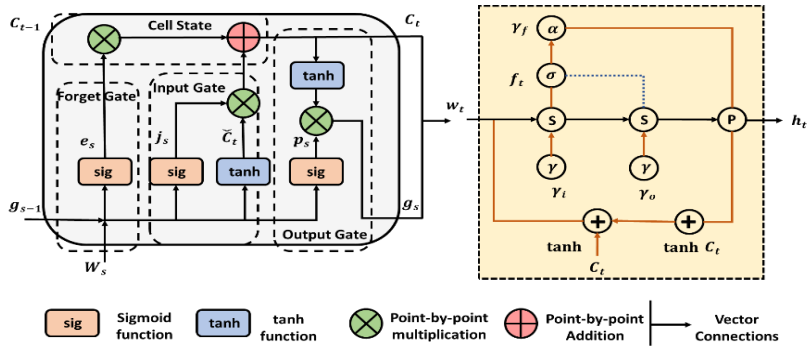

Fig. 3. Architecture of the proposed GA-LSTM

In this research, LSTM will be used to simulate the changing behaviour of tool wear by taking turns of the extracted features of time-frequency transformed machining signals. It records not only short-term changing (e.g. instant vibration changes) but also long-term degradation patterns in machining cycles.

a) Forget Gate

The forget gate is used to choose the data in the previous one memory to be retained or forgotten. This is particularly important in the machining settings, where previous sensor information may be out of date as a result of a shift in the cutting factors or material characteristics. The formulation of the forget gate is shown in equation 6.

$$e_s = \sigma(X_e w_s + V_e g_{s-1} + a_e) \quad (6)$$

In this equation, e_s represents the forget gate output at time step s, where X_e is the input, w_s and V_e were the weight matrices associated with the input vector and the previous hidden state g_{s-1}, and a_e denotes the corresponding bias parameter. The sigmoid activation function σ regulates how much past information is retained. In this research, this mechanism enables the network to ignore obsolete tool wear patterns that are no longer relevant to the current machining state.

b) Input Gate

The input gate regulates how much of the newly computed candidate memory should be incorporated into the current cell state. In tool wear context, it filters the incoming information from the current vibration, AE, or force signals to allow only meaningful updates that were represented in equation 7.

$$j_s = \sigma(X_j w_s + V_j g_{s-1} + a_j) \quad (7)$$

In this formulation, X_j is the input vector at the current time step, and g_{s-1} denotes the output from the preceding time step. The matrices w_s and V_j are learned weight parameters that transform the input and prior hidden state, respectively, while a_j is a bias vector. The sigmoid activation function σ crowds the resulting sum into a range between 0 and 1, producing the gate activation j_s. This number establishes how much the new input should affect the existing state, allowing the network to selectively update its memory and maintain long-term dependencies in sequence data. This gate ensures that only wear-relevant signal features (e.g., transient peaks or wear-induced distortions) are added to memory, avoiding noise or redundant data accumulation.

c) Candidate Cell Gate

Before the cell state is updated, a vector of candidate values is created according to the prior hidden state and the present input. These values are potential contributors to updated memory representation. Equation 8 depicts the computation of the candidate cell state.

$$\tilde{C}_s = \tanh(X_j w_s + V_d g_{s-1} + a_j) \quad (8)$$

The candidate cell state $\tilde{C}_s$ introduces a new memory content proposal depending the prior condition and the present input. It is generated using the hyperbolic tangent function $\tanh(.)$, which produces values between -1 and 1, suitable for memory content representation. In this case, current input X_j, past hidden state g_{s-1}, and bias a_j combined, transformed by weights w_s and V_d, are added to this intermediate update term, then before being clustered by the gates. The candidate values are a shortened version of the most recent sensor values, which have already been preconditioned to be added to the memory cell.

d) Cell State Update

The incoming candidate values are added to the Retained memory and combined with them to update the cell state. Equation 9 is the main memory update of the LSTM.

$$C_s = e_s \odot C_{s-1} + j_s \odot \tilde{C}_s \quad (9)$$

This equation is what controls the dynamic refinement of memory through time i.e. the forgetting of irrelevant past signals, and incorporation of new wear relevant information. The cell state C_s is optimised by combining the contents of retained memory and new content calculated. The forget gate is applied to the preceding memory and the input gate to the candidate state via the element-wise product $e_s of \odot C_{s-1}$ and $j_s \odot \tilde{C}_s$ respectively. The combined formulation is used to make sure that the LSTM cell is retaining valuable information as time goes by and forgetting less valuable information. It enables the model to monitor the slow wear development and abrupt changes in the state of tools.

e) Output Gate

Depending on what should affect the current output, it is the output gate that determines which portions of the internal cell memory are to be considered. It regulates the amount of internal state that is downstream processed or recurred to the next step. The equation 10 illustrates how the output gate is formulated.

$$p_s = \sigma\left(X_p w_s + V_p g_{s-1} + a_p\right) \quad (10)$$

The output gate p_s is what controls the amount of the cell state C_s that is to be disclosed to the following hidden state g_s. It is computed based on the current input X_p, the past hidden state g_{s-1}, bias a_p and weights w_s, V_p The sigmoid activation is that only the most pertinent memory information is uncovered in every time step. This gate enables the model to highlight or minimize some wear features in the generation of the output, enhancing predictive readability as well as interpretability.

f) Hidden State

Applying the output gate to the active cell state at time s is the hidden state or the LSTM unit final output which was represented in equation 11.

$$g_s = p_s \tanh \odot (C_s) \quad (11)$$

The output gate p_s is applied to the transformed (tanh-activated) cell state C_s to determine the ultimate concealed state g_s, which is sent to the following time step. This combination allows for flexible and context-aware sequence modeling by guaranteeing that the data carried forward to subsequent time steps is both temporally gated and filtered using a nonlinear transformation.

Each gate in the LSTM architecture plays a specific and critical role in managing the flow of machine signal information. In the proposed DG-GA-LSTM model, these gates are further enhanced with gate adjustment mechanisms and gravitational tuning, allowing the network to adapt memory behavior based on signal context, accelerate correction on high-error predictions, and achieve precise modeling of wear dynamics across time. This makes the gate-based LSTM design highly suitable for real-time, high-precision predictive maintenance in advanced manufacturing.

g) Gate Adjustment Module

LSTM gate activations can be too rigid or weak depending on the input signal's condition. Hence, adaptive gate modulation is introduced using learnable control coefficients in equation 12.

$$e_s^{'} = \gamma e.\, e_s,\quad j_s^{'} = \gamma j.\, j_s,\quad p_s^{'} = \gamma p.\, p_s \quad (12)$$

In these equations, $e_s^{'}, j_s^{'}, p_s^{'}$ are the modulated forms of the original vectors e_s, j_s, p_s, respectively. The scalars γe , γj , γp were learnable gating parameters that regulate every element's involvement while the model processes information. These coefficients tune the responsiveness of the input, forget, and output gates, respectively. By multiplying the original gate outputs with these coefficients, the model dynamically adjusts its sensitivity to changing conditions or optimization needs during training. This gate-adjusted mechanism improves temporal flexibility and contributes to more stable long-sequence learning. By enabling the model to tune the influence of each gate, the forget gate can suppress noise accumulation, the input gate can enhance critical transitions (e.g., tool degradation onset), and the output gate can focus on contextually important signal segments.

4.4.3. Dynamic Gravitational Search Algorithm

This research combines the DGSA, population-based metaheuristic, which is inspired by Newtonian gravitation law to improve the adaptive learning and optimization capacity of the implied predictive method in the process of precise tool wear tracking. DGSA is used to model the movement of particles (candidate solutions) in a multi-dimensional search space due to the effect of gravity and directs them to optimal solutions through attraction-repulsion dynamics. The general process contained in the DGSO is as shown in figure 4.

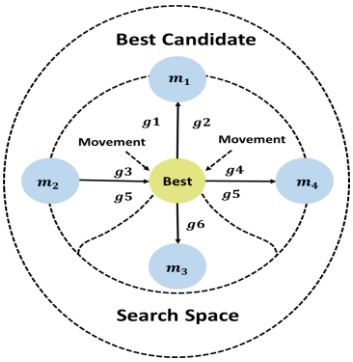

Fig. 4. The DGS algorithm structural procedure

All the particles are given mass that varies in accordance with its fitness, the highly-fitted solution has a stronger gravitational pull-on other particles. Due to the dynamic nature of particle positions in relation to each other as the algorithm repeats, DGSA is especially useful when the global search in an optimization landscape should be balanced with local search, resulting in a reduced impact of localization inherent in most optimization problems. The given dynamic search mechanism is an important part of the proposed DG-GA-LSTM framework that allows refining the LSTM gate parameters, making the prediction of tool wear more accurate and adaptable.

a) Gravitational Mass and Fitness Encoding

The fitness of each particle is equal to its mass. Particles with a higher gravitational pull are better solutions (particles) and more interesting to others. The fitness of each particle is normalized by an equation 13 to calculate the corresponding mass of the particle, which means that improved solutions have a greater mass, and this allows the proposed DG-GA-LSTM framework to be effective in the optimization of LSTM parameters to achieve the desired tool wear prediction.

$$n_j = \frac{fit_j - worst}{best - worst} \quad (13)$$

n_j Raw (unnormalized) gravitational mass of particle j

fit_j: Fitness of particle j

best : A best fitness value within the present population.

worst: Worst Fitness in the current population.

Equation 14 transforms the raw masses to a normalized form which makes the overall impact of the gravitational effect constant across the population.

$$N_j = \frac{n_j}{\sum_{i=1}^{M} n_i} \quad (14)$$

Here, the normalized gravitational mass of particle j was represented by N_j. n_j is the raw mass. M was the total number of particles. This mass determines the extent of gravitational influence that each particle exerts, with fitter particles guiding the search process more significantly.

b) Particle Interaction and Gravitational Force

Each particle experiences a gravitational force due to the presence of all other particles in the population. This force is based on the gravitational law, where attraction is stronger between more massive and nearby particles. Equation 15 calculates the directional gravitational pull on the particle i exerts on the particle j, based on their masses and distance, enabling the DG-GA-LSTM framework to dynamically fine-tune the LSTM network for improved prediction of tool wear progression.

$$E_{ji}^{c} = G.\frac{N_j.N_i}{Q_{ji}.\varepsilon}.\left(w_i^c - w_j^c\right) \quad (15)$$

Here, E_{ji}^{c} represents the gravitational force acting on the particle j from particle i along the c^{th} dimension. The terms N_i and N_j denote the normalized masses of particles i and j, respectively, These masses are proportional to the fitness scores of the particles and this means that more fit particles have a greater attraction. Q_{ji} gives the distance between the two particles calculated in Euclidean distance of the multidimensional search space. The term $\left(w_i^c - w_j^c\right)$ describes the directional aspect of the force, where w_i^c and w_j^c are the positions of particles i and j in dimension c. Lastly, G is the gravitational constant that scales the force and typically decays over iterations to allow for more refined searches in the later stages of optimization.

Equation 16 aggregates the net force acting on the particle ij from the top-performing L_{best} particles only, adding stochasticity for exploration.

$$E_j^c = \sum_{i \in L_{best}} rand_i.E_{ji}^c \quad (16)$$

Here, E_j^c represents the gravitational force acting on the particle j along the c^{th} dimension. This gravitational force governs the movement direction and speed of particles, driving them toward optimal regions.

c) Gravitational Constant Update

To control the strength of attraction over iterations, the gravitational constant is reduced exponentially to predict the tool wear in mechanical machining. Equation 17 controls how gravitational influence diminishes over time to shift from exploration to exploitation gradually.

$$G(s) = G_0.f^{-\alpha s/S} \quad (17)$$

Where $G(s)$ Gravitational constant at iteration s, G_0 is the initial gravitational constant, s is the current iteration, S is the total number of iterations, and α controls the decay rate. This guarantees more exploration at the initial stages and concentration in the exploitation at the later stages.

d) Acceleration, Velocity, and Position Update

There are three stages that are involved in Acceleration, Velocity and Position Update:(d) Acceleration, Velocity, and Position Update.The motion of every particle is calculated by applying Newton second law. The net force is calculated to obtain acceleration and determine velocity and position. Equation 18 calculates the rate at which a particle accelerates due to its force and mass, according to the second law of Newton to use internal weights adaptively to better predict tool wear and respond to any variation.

$$b_j^c = \frac{E_j^c}{N_j} \quad (18)$$

Equation 19 corrects velocity using previous momentum and current acceleration, which is calculated, to make movement dynamic.

$$u_j^c(s+1) = rand_j.u_j^c(s) + b_j^c \quad (19)$$

Equation 20 moves the position of the particle according to its present velocity.

$$w_j^c(s+1) = w_j^c(s) + u_j^c(s+1) \quad (20)$$

b_j^c is acceleration, u_j^c is velocity, and w_j^c is position in dimension c. E_j^c represents the total overall amount. The number of corresponding samples connected to the j^{th} element is denoted by N_j. $w_j^c(s+1)$ is the succeeding state of the position. $rand_j$introduces exploration randomness. $u_j^c(s)$ and $w_j^c(s)$ denote the velocity and position of the j^{th} particle in class c at iteration s, respectively. This dynamic update will allow the algorithm to model physical motion and change direction in a way that is adaptive towards likely solutions.

e) Global Memory Incorporation

A global memory structure is added to prevent stagnation and increase the accuracy of convergence that stores the best solution ever found. In equation 21, the velocity is adjusted to draw particles towards the global optimum discovered to date, a trade off between exploration and exploitation a vital improvement in improving continuing learning effectiveness in tool wear prediction.

$$u_j^c = rand_j.u_j^c + d_1.b_j^c + (2 - d_1).\left(w_{gbest}^c - w_j^c\right) \quad (21)$$

In this expression, u_j^c is the velocity of particle j in dimension c, and $rand_j$ is a random value between 0 and 1 that maintains exploration diversity. The term b_j^c denotes the acceleration of particle j, while w_j^c represents the current and global best positions, respectively. Equation 22 decreases d_1 over time to enhance exploitation in later stages of the algorithm.

$$d_1 = 2 - 2.\left(\frac{s}{S}\right)^3 \quad (22)$$

Where w^c_{gbest} is the position of the best particle in dimension c, and d_1 is a time-decaying coefficient controlling the influence of memory.By integrating this global memory term, DGSA enhances its exploitation capability and increases the likelihood of converging to a globally optimal solution. This term guides particles toward historically optimal regions, improving local search capability without losing diversity for the diagnosis of wear in the tool.

f) Exponential K-best Strategy

To reduce computational cost while maintaining search performance, DGSA dynamically reduces the number of influencing particles over time using exponential decay. Equation 23 dynamically decreases the number of particles contributing to the gravitational force, helping refine search precision over time a key strategy within the proposed DG-GA-LSTM context to optimize resource efficiency during predictive modeling.

$$K_{best} = M.\left(\frac{per}{100}\right)^{s/S} \quad (23)$$

Here, M is the population size, per is the final percentage of top particles used, and s/S denotes the current progress in iterations. Only the top K_{best} solutions apply a force as opposed to using all the particles. This plan eliminates the unwarranted calculation of low-impact particles and permits converged emphasis at the final phases. DGSO provides a scalable, adaptive, and robust optimization engine by simulating the gravitational dynamics and incorporating the memory based learning. It can be combined with DL structures (including DG-GA-LSTM) to improve the robustness of the model through the optimal tuning of hyperparameters, feature selection or learning sensitivities in dynamic and noisy machining environments. The tradeoff between exploration and exploitation, that exponential $Kbest$ and global memory offer, guarantees that processes converge to solutions, which are globally optimal, and are not prematurely stagnated. The procedure that is incorporated in the methodological approach is detailed in algorithm 1.

Algorithm 1 – DG-GA-LSTM

Initializeparticleswith randomLSTMparameters
FOReach particlej
TrainDG-GA-LSTMwith particlej'sparameters
Computefitnessbasedonpredictionerror
Calculatemassn_j = (fit_j - worst) / (best - worst)
ENDFOR
NormalizemassesN_j
WHILEnotconverged
FOReach particlej
ComputenetforceE_jfromtopK_bestparticles
IFglobal_memory_enabledTHEN
Updatevelocityusingtheglobalbestposition
ELSE
Updatevelocitywith thestandardrule
ENDIF
Updatepositionw_j = w_j + u_j
TrainDG-GA-LSTMwith updatedw_j
Updategates:
IFgate_adjustment_enabledTHEN
*Scalegates: e = γe * e, j = γj * j, p = γp * p*
ENDIF
Updatecellstateand hiddenstate
ENDFOR
IFcurrent_best<global_bestTHEN
Updateglobal_best
ENDIF
DecaygravitationalconstantandK_best
ENDWHILE
ReturnbestDG-GA-LSTMparameters
END

This algorithm summarize the incorporation of DGSA with GA-LSTM for optimizing model parameters. Each particle in the population represents a candidate set of LSTM parameters and was adjusted randomly. The fitness of each particle was assessed based on prediction error, and masses were allocated equally. Particles interact via gravitational forces, where top K_best particles influence movement. A global memory mechanism optionally guides convergence toward the best solution. During each iteration, particle positions w_j and velocities u_j were updated, LSTM gates (forget e, input j, output p) were scaled using coefficients (γe,γj,γp), and cell states were refined. The algorithm iteratively updates the global best, decays the gravitational constant, and returns the optimal parameters.

5. RESULTS AND DISCUSSION

The recital evaluation of the proposed DG-GA-LSTM model is shown in this section. The key fields of research are prediction accuracy, the ability to work under a variety of machining conditions, and performance in comparison with existing equipment. Evaluation measures that were applied to measure the efficiency of the model include Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). Furthermore, the benefits and drawbacks of the representation in the tool wear prediction are substantiated by the multi-dimensional visualizations, the comparison of the execution time, and the precision-recall evaluation.

5.1. Experimental Setup

The performance analysis of the proposed DG-GA-LSTM model was conducted using the publicly available Kaggle CNC tool wear dataset described in Section 4.1. To illustrate the potential deployment of the model in industrial environments, an OPC UA–based CNC monitoring architecture is presented as a conceptual framework for real-time sensor data acquisition. This framework demonstrates how machining signals such as spindle power, vibration, and cutting force could be transmitted and processed within a smart manufacturing system; however, the experiments reported in this research were

performed using the offline dataset. In this conceptual deployment framework, Open Platform Communications Unified Architecture (OPC UA) protocols enable the monitoring of machining parameters such as vibration, cutting current, and spindle power. As a conceptual strategy to explain how machining conditions may be methodically manipulated to increase data diversity under realistic industrial application, the strategy of an orthogonal experimental design is talked about. Before the DG-GA-LSTM model received training and testing, time-series division, normalization, and filtering of the noise (noise), preprocessing operations were performed on the dataset.

Parameters summarized in ensure reproducibility and the continuous inference capability of the proposed DG-GA-LSTM framework is summarized in Table 1.

Table 1. Experimental and Training Configuration of the Proposed DG-GA-LSTM Model

Parameter	Specification
Data splitting strategy	Tool-wise (run-wise) split to prevent data leakage
Training / validation / testing ratio	70% / 15% / 15%
Windowing method	Sliding window segmentation
Window length	T consecutive samples
Window overlap	Fixed overlap between adjacent windows
Label assignment	Flank wear value (VB) at the final time step
Model architecture	DG-GA-LSTM
Optimization algorithm	Adam
Initial Optimization Rate	0.001
Batch Processing Size	32
Total epochs	100
Early termination criterion	Enabled (patience value of10 epochs)
Random seed	Fixed for reproducibility
Optimization method	Dynamic Gravitational Search Algorithm (DGSA)
Optimized parameters	LSTM gate scaling coefficients
Inference time measurement	Average per-sample testing latency
Average inference time	0.00048 s
CPU	Intel Core i7
GPU	NVIDIA GTX / RTX
RAM	16 GB
Operating system	Windows
Implementation framework	Python 3.8
Library	TensorFlow (Keras API)

Table 1 summarizes the experimental setup that was employed in training and testing the proposed DG-GA-LSTM framework. In the case of time-series modeling, sensor signals were split into time-series windows by a sliding window method that has T = 128 consecutive samples in a time-series window meaning that there was enough temporal information to be learnt about wear progression patterns. Sequential windows were created with a 50 per cent overlap to maintain continuity of sequential machining state and at the same time, a larger number of training samples was created. In order to completely replicate the experiments, shuffling of data was done with a constant seed value of 42, weight initialisation and training. The implementation of the model was performed in Python 3.10 and DL libraries on a workstation with Intel Core i7-12700K processor (3.6 GHz), 32GB RAM, and others. an NVIDIA RTX 3080 GPU with 10 GB VRAM. These computing tools provided effective trainifiguewng and consistent optimization of the DG-GA-LSTM structure.

5.2. Feature Correlation Analysis for Tool Wear Prediction

In predictive maintenance, it is essential to pick the most appropriate features of raw multi-sensor data to improve the accuracy of the model and minimize costs in terms of computation. Since the multi-modal data observed through cutting force, vibration, and AE sensors are taken, the relationship between them and tool wear offers knowledge to the discriminative strength of each measure. This step can be considered as a precursor to the model training and it is dynamic in the optimization of the performance of the proposed DG-GA-LSTM model.

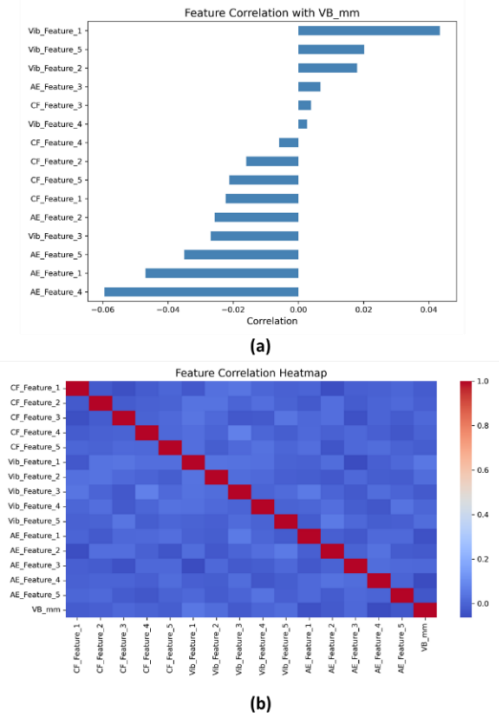

Fig. 5. Graphical Representation of (a) Feature-wise correlation with tool wear, (b) correlation between all extracted features

In Figure 5, a detailed correlation analysis of features was done. Figure 5 (a) shows Pearson correlation of each of the extracted features with the

tool wear measure (VB_mm), and Figure 5 (b) shows the total feature-to-feature correlation heatmap. The correlation coefficients between features are also very low which supports the fact that there is not much redundancy between features thus indicating the uniqueness of every input dimension that has been obtained via the WPT. The model could best represent wear-relevant patterns by considering the features with larger correlation magnitude, and this led to an increase in the quality of real-time detection and prediction.

5.3. Exploratory Data Analysis of Tool Wear Classes

An EDA has been conducted to get initial information on the sensor data and tool wear behavior. The purpose was to be able to visually see the spread of the categories of wear and the relationship of these individuals to attributes gained through multi-sensor signals. The step is necessary in identifying feature correlation, the class imbalance, and potential noise to assist in the selection of models and preprocessing methods..

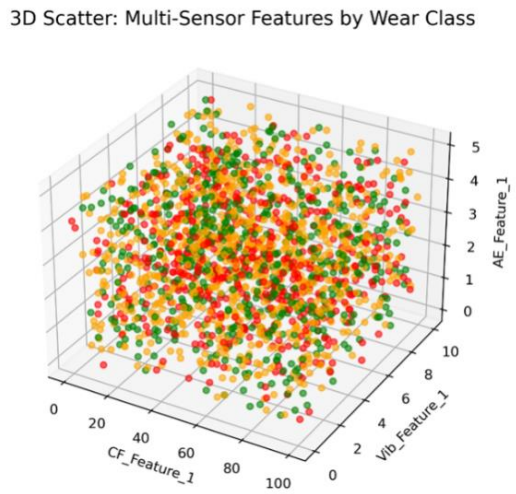

Fig. 6. Representation of sensor-level features distributed by wear class

The difficulty of classifying wear categories using raw characteristics is demonstrated by Figure 6, which shows a notable overlap in feature values. Visible cluster trends, however, imply that latent representations might be successfully identified using DL methods like DG-GA-LSTM. Although moderate wear samples are somewhat more prevalent, the dataset maintains a well-balanced representation of all wear stages, which makes it suitable for supervised classification tasks without undue bias, as indicated by the class distribution histogram.

5.4. Tool Wear Trend Analysis Using Measurements

The wear development over time might be visualized as an efficient way to identify the pattern of cutting tool degradation. , a measure of flank wear, was tracked and examined in this research utilizing time-series visualization methods. For easier reading, the charts overlay wear class transitions and smooth high-frequency variations using the Rolling Mean technique. These visual cues assist in creating ground truth for the DG-GA-LSTM model's classification and regression tasks.

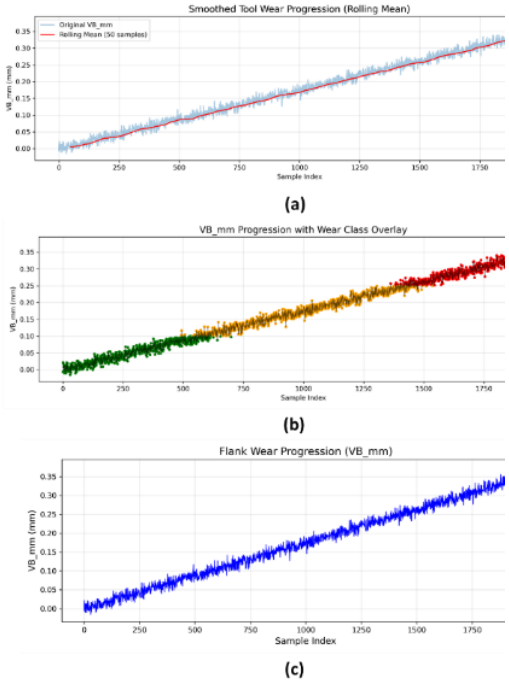

Fig. 7. Graphical Representation of (a) Smoothed tool wear progression using a 50-sample rolling average. (b) values with class-wise segmentation. (c) Raw progression of flank wear in mm across all samples

Figure 7 demonstrates a nearly linear increase in VB_mm with machining time, reinforcing its suitability as a primary indicator of degradation. The smoothed trend aligns closely with the DG-GA-LSTM's learning objectives, offering a stable target for prediction. Additionally, the class overlay visually validates the gradual transition across wear states, confirming the label fidelity used in downstream modeling and performance evaluation.

5.5. Sensor Signal Evaluation and Feature Observation

The temporal behavior of sensor signals is necessary in understanding how the cutting tools wear out during the machining process. In this research, cutting force (CF), vibration (Vib) and acoustic emission (AE) inputs signals were cut and presented in a manner that would evaluate the change of features with time. This research assists in identifying the relative value of each sensor to tool wear prediction, as well as confirming the stability of the signals.

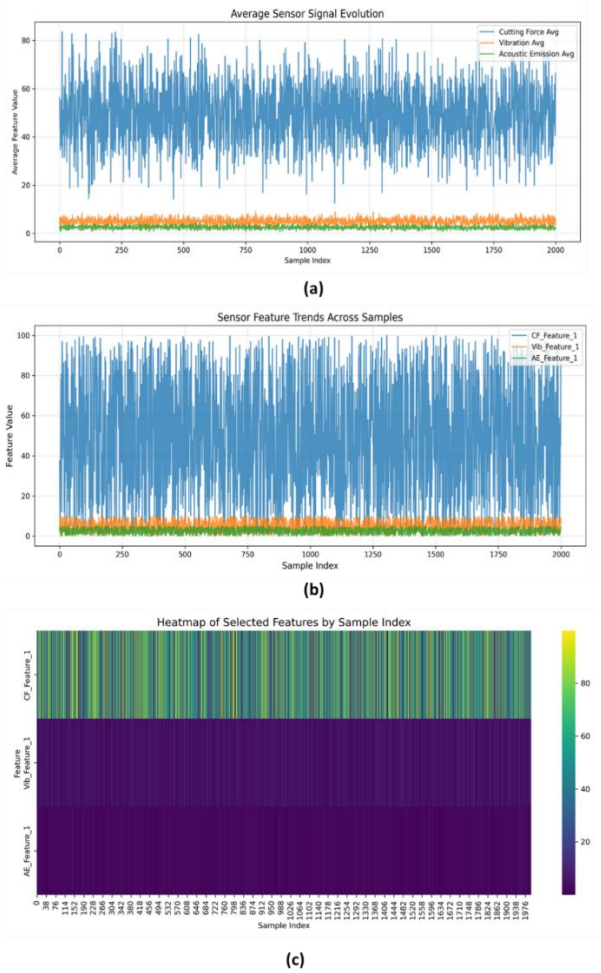

Fig. 8. Graphical representation of (a) the evolution of mean signal of CF, Vib and AE sensors during the machining period. (b) Sample-wise feature trends. (c) Heatmap of the intensity of feature distribution across the sample indices of selected sensor features

Figure 8(a) shows the amplitude evolution of average feature values of various sensor modalities on a comparative basis. The Cutting force exhibits high dynamism, which denotes that it is sensitive to progressive wear. Figure 8(b) further decomposes that at the feature level with CF_Feature_1 having specific patterns of fluctuation with Vib_Feature_1 and AE_Feature_1 having fairly constant values. Finally, Figure 8(c) shows a heatmap that reflects the sample-wise intensity distribution, which validates the fact that CF-based features were characteristics with high likelihood of wear progression. These are the insights that back the multi-sensor fusion approach in the DG-GA-LSTM approach.

5.6. Feature Interaction and Distribution Analysis

Scatter matrix visualization techniques were used to investigate the correlations between major sensor characteristics and how it relates to the course of tool wear. These plots aid in the comprehension of pairwise feature distributions and the identification of overlaps or non-linear correlations in wear progression patterns. The DG-GA-LSTM model's justification for multi-sensor feature fusion is supported by this examination.

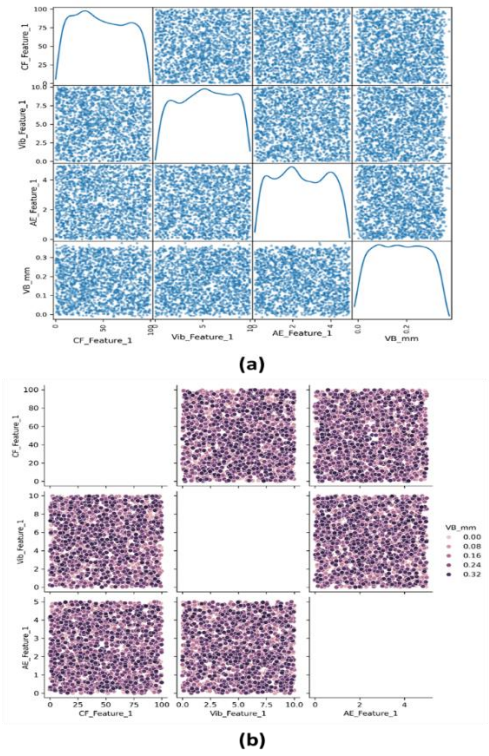

Fig. 9. Graphical Representation (a) pairwise distributions of selected features and (b) the distribution spread across the wear progression levels

Every feature in Figure 9(a) has a distinct distribution, with CF_Feature_1 and AE_Feature_1 displaying the most variability. By representing the wear depth (VB_mm) as color intensities, Figure 9(b) offers a more thorough representation, allowing for a visual examination of the connection among wear progression and specific sensor data. The combination of the two figures shows that there isn't a clear linear association, but it also supports the necessity for deep feature learning techniques to identify latent temporal patterns that are important for predicting tool wear.

5.7. Sensor Feature Visualization and Distribution Trends

Three supplemental visualizations were created to better examine the distribution and behavior of multi-sensor characteristics related to tool wear prediction. These include a kernel density estimate (KDE) plot for distribution analysis, a 3D surface plot to look at temporal variability in cutting force characteristics, and a bivariate scatter plot of wear vs cutting force. Over time, these visualizations aid in revealing the underlying trends and variances in sensor responses.

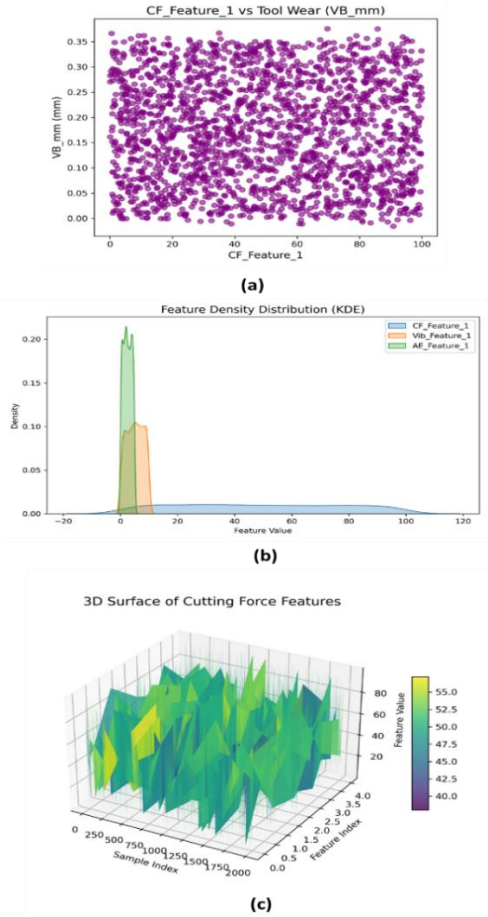

Fig. 10. Graphical representation of tool wear versus c1 feature, (b) feature density distribution, and (c) cutting force features

Figure 10(a) highlights the non-linear character of the relationship between cutting force values and tool wear data by showing the dispersion between them. While cutting force characteristics display a wider range, indicating higher variability, vibration, and acoustic emission features display densities that are tightly packed, as seen in Figure 10(b). A (3-dimensional) 3D surface representation of CF_Feature_1 is shown in Figure 10(c), which amply illustrates the cutting force signal's dynamic, non-stationary character over time. The necessity of sophisticated feature extraction techniques like the DG-GA-LSTM model is confirmed by this multimodal representation.

Confusion matrix

A confusion matrix summarizes the DG-GA-LSTM model's predictions versus actual tool wear classes, showing correct classifications and misclassifications across all wear categories.

Figure 11 The confusion matrix shows strong classification performance, with most samples correctly identified along the diagonal for Healthy, Moderate, and Worn classes. Only a small number of instances are misclassified between neighboring conditions, indicating high model reliability. Overall, the results demonstrate effective discrimination among wear states.

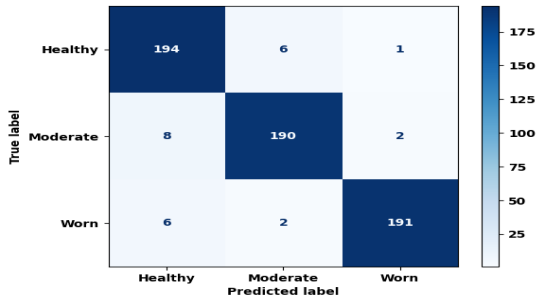

Fig. 11. Confusion matrix showing the classificaross all wear categories simultaneously

Multi class ROC curve

The multiclass ROC curve evaluates the DG-GA-LSTM model's ability to distinguish multiple tool wear levels, quantifying diagnostic performance across all wear categories simultaneously.

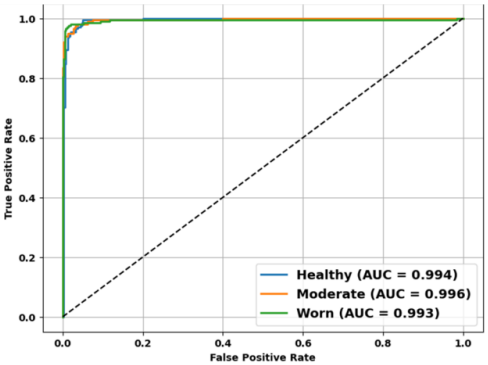

Fig. 12. Multiclass ROC curve showing model performance across wear states

Figure 12 shows multiclass ROC curve illustrates the model's classification performance for tool wear states. AUC values indicate excellent discrimination: Healthy = 0.994, Moderate = 0.996,

and Worn = 0.993, showing the model accurately distinguishes all wear conditions.

Multi class precision recall curve

The multiclass Precision-Recall curve measures the DG-GA-LSTM model's accuracy and completeness in identifying each tool wear category, highlighting performance across all classes simultaneously.

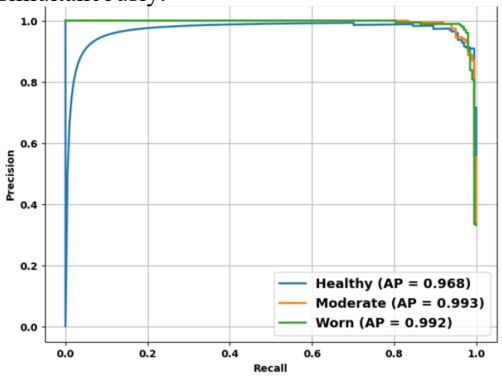

Fig. 13. Multiclass Precision-Recall curve showing classification performance across wear states

Figure 13 multiclass Precision-Recall curve evaluates model accuracy for tool wear classification. Average Precision (AP) scores are high: Healthy = 0.968, Moderate = 0.993, and Worn = 0.992, indicating strong precision and recall, especially for moderate and worn states.

5.8. Performance Evaluation of Tool Wear Prediction Models

The performance of the suggested DG-GA-LSTM model was assessed against that of other

DL architectures, such as the LSTM, ResNet, and ResNet-LSTM models described in [21], to confirm the model's efficacy. Three common regression assessment measures were used in the comparison: R^2, MAE, and RMSE. Table 2 represents the comparative performance of the proposed approach over models.

Table 2. Performance comparison of Tool Wear Prediction Models

Model	MAE (mm)	RMSE (mm)	R^2
LSTM [21]	0.0182	0.0281	0.8744
ResNet [21]	0.0118	0.0182	0.9745
ResNet-LSTM [21]	0.0085	0.0101	0.9825
DG-GA-LSTM [proposed]	0.0068	0.0094	0.9967

As evident from Table 2, the proposed DG-GA-LSTM model significantly outperforms traditional LSTM and hybrid ResNet-LSTM approaches in all three-evaluation metrics. It achieves the lowest MAE of 0.0068 mm and RMSE of 0.0094 mm, indicating highly accurate and consistent tool wear prediction. Furthermore, the model exhibits a near-perfect R^2 score of 0.9967, demonstrating excellent predictive alignment with actual tool wear trends. This improvement can be attributed to the dynamic gravitational tuning and gate-adjustment mechanisms, which enhance temporal sensitivity and memory retention in the proposed model, resulting in superior regression performance under varying machining conditions.

MAE

A popular regression performance indicator called MAE measures the average size of errors between expected and actual values without taking direction into account. By comparing the anticipated wear values to the actual wear measurements under various machining settings, In the context of this research, MAE helps evaluate the correctness of tool wear estimation. Equation 24 depicts the formulation of MAE.

$$\mathrm{MAE} = \frac{1}{m}\sum_{j=1}^{m}\left|z_j - \hat{z}_j\right| \tag{24}$$

The model's estimates are more in line with the actual tool wear states when the MAE value is smaller, which is essential for ensuring reliable decision-making in real-time diagnostics and predictive maintenance. In this equation, m is the total number of data points. The corresponding tool degradation value for the jth sample is represented by z_j. The expected tool wear value for the jth sample is $\hat{z}_j$. The absolute difference (error) between the actual and projected values is represented by $\hat{z}_j$. As shown in Figure 14, MAE is a key metric used in this study to verify the accuracy of the DG-GA-LSTM approach for evaluating tool degradation over time.

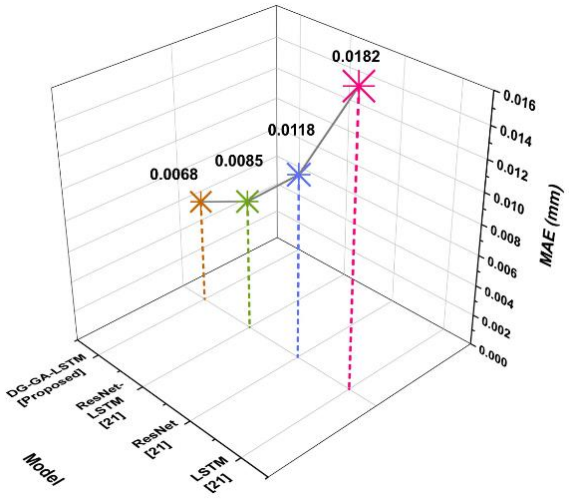

Fig. 14. Graphical representation of MAE

RMSE

A common regression statistic called Root Mean Square Error (RMSE) calculates the square root of the average squared discrepancies between the actual and predicted values. When assessing models where greater deviations are more crucial, such as in tool wear prediction, RMSE is especially helpful since it penalizes larger mistakes more severely than MAE. Equation 25 represents the calculation formula of RMSE.

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(z_j - \hat{z}_j\right)^2} \tag{25}$$

A lower RMSE value signifies that the predictive model produces fewer and less severe errors. In this research, RMSE is used to evaluate the stability and accuracy of DG-GA-LSTM predictions, ensuring that large deviations in tool wear estimation are minimized to avoid operational risks. Figure 15 depicts the illustration of RMSE.

Coefficient of Determination (R^2)

A measurement of statistical significance that illustrates how well the expected and actual data match is called the coefficient of determination. Higher numbers, which range from 0 to 1, suggest a better match. R^2 aids in evaluating how well the model reflects the underlying patterns in sensor-based wear progression in tool wear prediction represented in equation 26.

$$R^2 = 1 - \frac{\sum_{j=1}^{m}\left(z_j - \hat{z}_j\right)^2}{\sum_{j=1}^{m}\left(z_j - \bar{z}_j\right)^2} \tag{26}$$

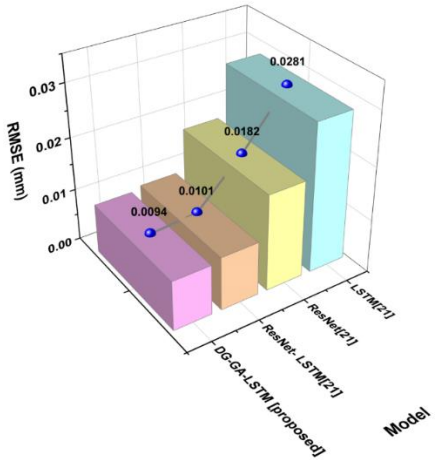

Fig. 15. Graphical representation of RMSE

In Figure 16, the model's predictive power and generalizability are confirmed by an R^2 value near 1, which shows that it accounts for the majority of the variation in tool wear. The high R^2 that DG-GA-LSTM achieved in this research is indicative of its exceptional ability to simulate intricate wear behavior under actual machining circumstances.

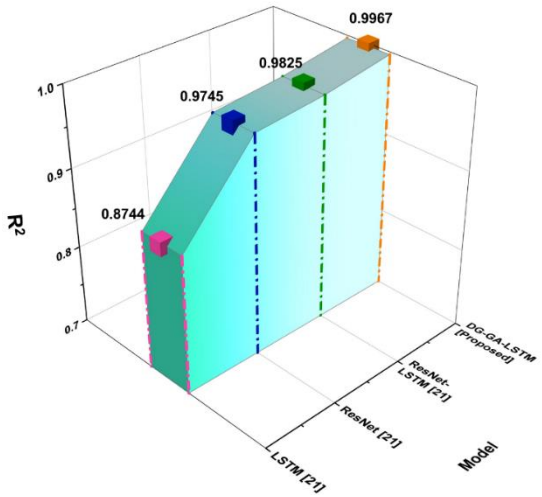

Fig. 16. Graphical representation of R2

5.9. Precision and Recall-Based Performance Comparison

In tool wear classification, precision and recall are important performance metrics, especially when assessing the model's capacity to accurately detect worn tool states and minimize false positives. A comparison between the suggested DG-GA-LSTM framework and the CWT [22] based model is shown in the following table. The accuracy and recall comparison with conventional methods is shown in Table 3 and Figure 17.

Precision

One important indicator is precision, which assesses the percentage of accurately predicted positive cases out of all instances that were anticipated to be positive. By only marking true wear conditions as worn, the model is able to prevent false alarms in the context of tool wear categorization. Equation 27 depicts the formulation of precision.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{27}$$

Here, TP represents the true positive, which correctly identifies worn tool instances, and FP depicts the false positive, where normal tools are incorrectly labeled as worn. A higher precision score indicates that the model is highly reliable when it predicts that a tool is worn, which is essential in reducing unnecessary tool replacements in CNC machining environments.

Recall

A proportion of actual positive cases that the model correctly identified is measured by recall. It is sometimes referred to as true positive rate or sensitivity. It measures the model's capacity to identify every occurrence of tool wear in this research, reducing missed malfunctions. Equation 28 depicts the formulation of recall.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{28}$$

Here, TP represents the true positives that worn tools correctly detected, and FN represents false negatives that worn tools missed by the model. High recall ensures that critical wear-related events are not missed, thereby reducing the risk of unexpected tool failures and production downtime. Table 3 depicts the comparative performance of the implied approach over the method.

Table 3. Precision and Recall Comparison of Competing Models

Model	Precision (%)	Recall (%)
CWT[22]	92.93	87.78
DG-GA-LSTM [proposed]	94.68	90.47

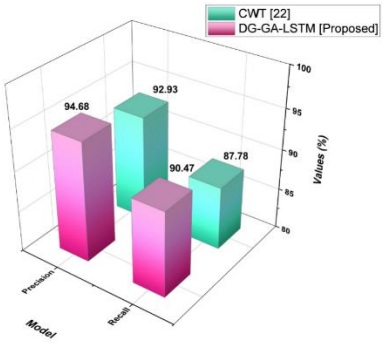

Fig. 17. Graphical representation of precision and recall

5.10. Testing Time

Each model variant's testing time per inference was assessed to assess the computational cost related to various DL architectures. This statistic sheds light on the model's viability for real-time deployment, particularly in situations where low-latency predictions are crucial, such as tool status monitoring. Conventional LSTM, BiLSTM, hierarchical LSTM (HLLSTM), and the suggested

DG-GA-LSTM are among the models taken into consideration that were represented in Table 4 and Figure 18.

Table 4. Average per-sample inference time (seconds) for different deep learning models measured after warm-up under identical hardware and software configurations

Model	Testing Time (s)
LSTM [23]	0.00054
BiLSTM [23]	0.00072
HLLSTM [23]	0.00063
DG-GA-LSTM [proposed]	0.00048

To ensure fair and reproducible comparison of inference latency across models, all experiments were conducted on the same hardware and software platform (Intel Core i7 CPU, NVIDIA GTX/RTX GPU, 16 GB RAM, Python 3.8, TensorFlow/Keras). Inference time was measured as the average per-sample forward-pass latency after warm-up runs, with identical batch size and execution conditions for all models.

Although the absolute latency differences among models are small (on the order of microseconds), DG-GA-LSTM consistently achieved the lowest inference time (0.00048 s). This demonstrates that the proposed dynamic gravitational tuning and gate-adjustment mechanisms do not introduce additional computational overhead compared to existing DL architectures, confirming its suitability for real-time CNC tool wear monitoring applications.

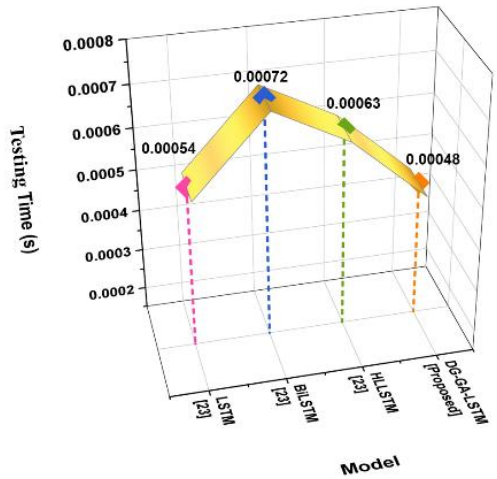

Fig. 18. Comparison of different DL models with respect to their execution time (s)

Figure 18 shows the average prediction time of the various models and it can be noted that the proposed framework can be used in real time. DG-GA-LSTM has the lowest average inference time compared to traditional LSTM, Bi-LSTM, and GRU models, which means that it is effective when it is necessary to make a decision quickly. This is important in real time tool wear diagnostics in industrial CNC settings due to its short latency. The results prove the applicability of the model to be employed in time-sensitive manufacturing systems.

Table 5. The proposed DG-GA-LSTM model results on the classification of various tool wear classes

Wear Class	Precision (%)	Recall (%)	F1-score (%)
Normal	95.4	93.7	94.5
Mild Wear	93.8	92.1	92.9
Moderate	94.6	95.3	94.9
Severe	96.2	96.8	96.5
Macro Avg	95.0	94.5	94.7
Micro Avg	95.9	95.9	95.9
Overall Accuracy	95.7 %		

Table 5 shows precision, recall, and F1-score for each tool wear class predicted by the proposed DG-GA-LSTM model. Large values of high precision indicate that the model has low false-positive prediction rates whereas large values of high recall indicate that the model identifies the majority of true cases of each wear class. The F1-score (a combination of precision and recall) shows that the model has a high level of consistent performance throughout all the wear classes. Both the macro and micro averages are published to provide a summary of the general model performance which both surpass 94 percent to show consistent and balanced classification results of normal, mild, moderate, and severe tool wear.

5.11. Discussion

A effcetive tool wear prediction process is important in making sure that machining is reliable and efficient. The current models such as LSTM [21,23], BiLSTM [23], HLLSTM[23], ResNet[21] and CWT [22] are characterized by low flexibility, reduced convergence and poor performance with non-stationary signals. To overcome these limits, the proposed DG-GA-LSTM model uses a DGSO to incorporate GA-LSTM so that adaptive learning and efficient feature retention are realized. The results of the experiment confirm higher performance based on accuracy, duration of execution, and generalization. This confirms the model as highly efficient in dealing with various machining environments and points to its ability to have real-time intelligent tool wear monitoring.

6. CONCLUSION

Efficient tool wear forecasting is crucial in ensuring machine efficiency, minimizing the downtime, and improving the quality of products. The suggested DG-GA-LSTM model was quite effective in the prediction accuracy in this research with an MAE of 0.0068mm, RMSE of 0.0094 mm, R^2 of 0.9967, Precision of 94.68 and recall of 90.47, which was better than the traditional models, LSTM, BiLSTM and ResNet-LSTM. The interaction between DGSO convergence behavior, with GA in the LSTM making learning of time-dependences improved. Although its performance looks promising, it still has a limitation in that it is not

capable of adapting to hidden machine types or highly volatile wear patterns. The future studies could expand on larger data sets, incorporation of multiple sensors and real time implementation to guarantee improved generalization and scalability to different machining environments.

Source of funding: *This research received no external funding.*

Author contributions: *research concept and design, W.Y., C.B.; Collection and/or assembly of data, W.Y., C.B.; Data analysis and interpretation, W.Y., C.B.; Writing the article, S.R., W.Y., C.B.; Critical revision of the article, W.Y., C.B.; Final approval of the article, W.Y., C.B.*

Declaration of competing interest: *The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.*

REFERENCES

1. Chan YW, Kang TC, Yang CT, Chang CH, Huang SM, Tsai YT. Tool wear prediction using convolutional bidirectional LSTM networks. The Journal of Supercomputing. 2022; 78(1):810-832. https://doi.org/10.1007/s11227-021-03903-4.
2. Cheng M, Jiao L, Yan P, Li S, Dai Z, Qiu T, Wang X. Prediction and evaluation of surface roughness with a hybrid kernel extreme learning machine and monitored tool wear. Journal of Manufacturing Processes. 2022;84:1541-1556. https://doi.org/10.1016/j.jmapro.2022.10.072.
3. Cheng M, Jiao L, Yan P, Jiang H, Wang R, Qiu T, Wang X. Intelligent tool wear monitoring and multi-step prediction based on a deep learning model. Journal of Manufacturing Systems. 2022;62:286-300. https://doi.org/10.1016/j.jmsy.2021.12.002.
4. Cheng Y, Lu M, Gai X, Guan R, Zhou S, Xue J. Research on a multi-signal milling tool wear prediction method based on GAF-ResNext. Robotics and Computer-Integrated Manufacturing. 2024;85: 102634. https://doi.org/10.1016/j.rcim.2023.102634.
5. Danish M, Gupta MK, Irfan SA, Ghazali SM, Rathore MF, Krolczyk GM, Alsaady A. Machine learning models for prediction and classification of tool wear in sustainable milling of additively manufactured 316 stainless steel. Results in Engineering. 2024;22: 102015. https://doi.org/10.1016/j.rineng.2024.102015.
6. Ferrando Chacón JL, Fernández de Barrena T, García A, Sáez de Buruaga M, Badiola X, Vicente J. A novel machine learning-based methodology for tool wear prediction using acoustic emission signals. Sensors. 2021;21(17):5984. https://doi.org/10.3390/s21175984.
7. Guo H, Zhang Y, Zhu K. Interpretable deep learning approach for tool wear monitoring in high-speed milling. Computers in Industry. 2022;138:103638. https://doi.org/10.1016/j.compind.2022.103638.
8. Hassan M, Sadek A, Attia H. A real-time deep machine learning approach for sudden tool failure prediction and prevention in machining processes. Sensors. 2023;23(8):3894. https://doi.org/10.3390/s23083894.
9. He Z, Shi T, Xuan J, Li T. Research on tool wear prediction based on temperature signals and deep learning. Wear. 2021;478:203902. https://doi.org/10.1016/j.wear.2021.203902.
10. Kamat P, Kumar S, Kotecha K. DeepTool: A deep learning framework for tool wear onset detection and remaining useful life prediction. MethodsX. 2024;13:102965. https://doi.org/10.1016/j.mex.2024.102965.
11. Korkmaz ME, Gupta MK, Çelik E, Ross NS, Günay M. Tool wear and its mechanism in turning aluminum alloys with image processing and machine learning methods. Tribology International. 2024;191:109207. https://doi.org/10.1016/j.triboint.2023.109207.
12. Li B, Lu Z, Jin X, Zhao L. Tool wear prediction in milling CFRP with different fiber orientations based on multi-channel 1DCNN-LSTM. Journal of Intelligent Manufacturing. 2024;35(6):2547-2566. https://doi.org/10.1007/s10845-023-02164-7.
13. Liu J, Yu D, Hu Y, Yu H, He W, Zhang L. CNC machine tool fault diagnosis integrated rescheduling approach supported by digital twin-driven interaction and cooperation framework. IEEE Access. 2021;9: 118801-118814. https://doi.org/10.1109/ACCESS.2021.3106797.
14. Ma J, Luo D, Liao X, Zhang Z, Huang Y, Lu J. Tool wear mechanism and prediction in milling TC18 titanium alloy using deep learning. Measurement. 2021;173:108554. https://doi.org/10.1016/j.measurement.2020.108554.
15. Parwal V, Rout BK. Machine learning based approach for process supervision to predict tool wear during machining. Procedia CIRP. 2021;98:133-138. https://doi.org/10.1016/j.procir.2021.01.018.
16. Truong TT, Airao J, Hojati F, Ilvig CF, Azarhoushang B, Karras P, Aghababaei R. Data-driven prediction of tool wear using Bayesian regularized artificial neural networks. Measurement. 2024;238:115303. https://doi.org/10.1016/j.measurement.2024.115303.
17. Twardowski P, Tabaszewski M, Wiciak–Pikuła M, Felusiak-Czyryca A. Identification of tool wear using acoustic emission signal and machine learning methods. Precision Engineering. 2021;72:738-744. https://doi.org/10.1016/j.precisioneng.2021.07.019.
18. Wang D, Hong R, Lin X. A method for predicting hobbing tool wear based on CNC real-time monitoring data and deep learning. Precision Engineering. 2021;72:847-857. https://doi.org/10.1016/j.precisioneng.2021.08.010.
19. Wang G, Zhang F, A sequence-to-sequence model with attention and monotonicity loss for tool wear monitoring and prediction. IEEE Transactions on Instrumentation and Measurement. 2021;70:1-11. https://doi.org/10.1109/TIM.2021.3117082.
20. Wang K, Wang A, Wu L, Xie G. Machine tool wear prediction technology based on multi-sensor information fusion. Sensors. 2024;24(8):2652. https://doi.org/10.3390/s24082652.
21. Xiao Y, Jiang Z, Gu Q, Yan W, Wang R. A novel approach to CNC machining center processing parameters optimization considering energy-saving and low-cost. Journal of Manufacturing Systems. 2021;59:535-548. https://doi.org/10.1016/j.jmsy.2021.03.023.
22. Zhang J, Zeng Y, Starly B. Recurrent neural networks with long-term temporal dependencies in machine tool wear diagnosis and prognosis. SN Applied Sciences.

2021;3(4):442. https://doi.org/10.1007/s42452-021-04427-5.
23. Zhang Y, Qi X, Wang T, He Y. A tool wear condition monitoring method based on deep learning with force signals. Sensors. 2023;23(10):4595. https://doi.org/10.3390/s23104595.

Yuegui WU was born in Yangjiang, Guangdong, China, in 1981. She received her Bachelor of Engineering degree in Mechatronics Technology Education from Zhanjiang Normal University in 2005 and her Master's degree in Software Engineering from Guangdong University of Technology in 2018.

Since 2005, she has been working at the Department of Mechanical and Electrical Engineering, Yangjiang Polytechnic, as a teacher specializing in numerical control. She has published more than 10 academic papers and holds two invention patents, as well as several utility model patents and design patents. Her research interests include mechanical design, mold design, numerical control programming, and numerical control machining.
e-mail: yuegui_wu@hotmail.com

Bailiang CHEN was born in Gaozhou, Guangdong, China, in 1982. He received his Bachelor of Engineering degree in Mechatronics Technology Education from Zhanjiang Normal University in 2005.

From 2005 to 2006, he worked as a teacher at Zhanjiang Technician College. Since 2007, he has been a teacher majoring in mechatronics at the Department of Mechanical and Electrical Engineering, Yangjiang Polytechnic.

He has published several academic papers and holds one invention patent, as well as a number of utility model patents and design patents.

His research interests include mechatronics technology and numerical control machining.
e-mail: chenbailiang86@outlook.com