ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 12, Issue 06, June 2025

Enhancing Tool Wear Detection in CNC Milling: A Comparative Study of Attention Based Deep Learning Models

[1] Akshit Jain, [2] Rajesh Kaushal, [3] Jatin, [4] M.S. Niranjan

[1] [2] [3] [4] Department of Production and Industrial Engineering, Delhi Technological University, Delhi, India
*Corresponding Author's Email: [1] akshitj285@gmail.com, [2] wreadkaushal@gmail.com, [3] jatinkanojiya641@gmail.com, [4] mahendraiitr2002@gmail.com

Abstract*— Tool wear detection is critical in Computer Numeric Control (CNC) milling to maintain machining precision, reduce material waste, and extend tool life. Traditional methods, including manual inspection and basic sensor-based approaches, are often inefficient and prone to error, underscoring the need for automated, real-time solutions. This study investigates the effectiveness of deep learning (DL) models incorporating attention mechanisms for tool wear classification. Four architectures—CNN with attention, LSTM with attention, Hybrid CNN-LSTM with attention, and Transformer encoder—are evaluated on a benchmark machining dataset. Among the models tested, the Hybrid CNN-LSTM with attention achieves the highest accuracy of 99.85%, effectively capturing both spatial and temporal dependencies in the data. The LSTM with attention model follows with 99.69% accuracy, excelling in modeling sequential patterns. CNN with attention also performs well with 99.15% accuracy. In contrast, the Transformer encoder achieves comparatively lower accuracy at 89.3%, suggesting a need for further optimization for this specific application. The results highlight the strong potential of attention-based DL models in enhancing tool wear detection in smart manufacturing environments. Specifically, the Hybrid CNN-LSTM with attention model demonstrates suitability for real-time monitoring and predictive maintenance. This research contributes to the development of robust, automated systems for tool condition monitoring and lays the groundwork for future improvements in intelligent manufacturing technologies.

Keywords: CNC Milling, Deep Learning, Attention Mechanism, Tool Wear Detection.

I. INTRODUCTION

In modern manufacturing, CNC (Computer Numerical Control) milling is a crucial machining process in modern manufacturing, producing high-precision components across industries such as aerospace, automotive, and medical devices [1,2]. Fig. 1 shows a CNC milling machine in operation, showcasing the control panel, cutting area, and tooling setup used for precision machining in manufacturing industries. However, the efficiency of CNC milling is highly dependent on the condition of the cutting tools, which experience gradual wear due to mechanical stress, high temperatures, and friction with workpiece materials [3, 4]. Tool wear leads to reduced machining accuracy, increased cutting forces, poor surface finish, and even catastrophic tool failure if not monitored and managed properly [5]. This has driven the need for effective tool wear detection systems that enable real-time monitoring and predictive maintenance, reducing downtime and improving production efficiency [6–8].

Fig. 1: CNC milling machine at DTU

Traditionally, tool wear detection has relied on direct and indirect monitoring techniques [9]. Direct methods, such as optical microscopy and scanning electron microscopy (SEM), provide highly accurate measurements but require machine stop pages, making them impractical for real-time applications [10]. Indirect methods utilize sensor data, including vibrations, acoustic emissions, cutting forces, and spindle power, to infer tool wear conditions [11]. While these methods enable real time monitoring, they often require complex feature extraction and signal processing techniques to ensure accuracy [12]. In recent years, Machine learning (ML) and Deep Learning (DL) approaches have revolutionized tool wear detection by automating feature extraction and improving predictive capabilities [13,14].

IFERP
ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 12, Issue 06, June 2025

Reliable tool wear detection is crucial for improving manufacturing efficiency and minimizing downtime. Traditional machine learning approaches rely heavily on manual feature extraction, which is time-consuming and lacks the adaptability to different machining conditions. Deep learning (DL) eliminates this limitation by automatically learning patterns from raw sensor data, making it more effective in handling complex, high-dimensional information [15].

Unlike ML models that require predefined indicators, DL architectures such as CNNs and LSTMs can extract hidden patterns from machining data without human intervention. Attention mechanisms further enhance DL models by dynamically focusing on the most relevant features, improving accuracy and robustness [16]. This adaptability makes DL a superior choice for real-time tool wear detection in CNC milling, where machining conditions frequently vary [17].

Furthermore, integrating attention-based models aligns with the goals of Industry 4.0, enabling smart manufacturing with predictive maintenance and real-time decision making. By leveraging advanced DL techniques, manufacturers can enhance productivity, reduce tool replacement costs, and optimize machining performance, making tool wear detection more efficient and scalable [18].

This paper is structured into six sections. Section 1 provides an overview of tool wear detection in CNC milling and the motivation for using attention-based DL models. Section 2 shows the literature review that reviews existing approaches, of DL techniques and presents the research gaps. Section 3 shows the problem statement and research objectives. Section 4 illustrates the technical background that covers DL architectures, attention mechanisms, dataset preparation, model implementation and evaluation metrices. Section 5 presents the experimental setup and comparative analysis of the models. Finally, Chapter 6 summarizes key findings, contributions to smart manufacturing, and directions for future research.

II. LITERATURE REVIEW

DL-based approaches for tool wear detection have gained significant attention, with various models being explored to improve prediction accuracy and monitoring efficiency [19].

Table 1: Related work on DL-based tool wear detection

Reference	Model	Results (%)	Research Focusses	Comments
Kumar et al. [20] (2025)	EfficientNet-B0 (Pre-trained CNN)	94.11	Vision-based tool wear monitoring with CNN.	Self-attention or hybrid architectures not explored.
Xu et al. [21] (2025)	Multi-scale CNN BiLSTM-GCN	98.50	Hybrid CNN, BiLSTM, GCN for spatial-temporal features.	No explicit attention mechanisms.
He et al. [22] (2025)	Semi-Supervised LSTM	95.78	Semi-supervised LSTM for tool wear detection.	No attention for global dependency learning.
Karabacak et al. [23] (2023)	CNN(GoogleNet, AlexNet, ResNet-50, EfficientNet-B0)	99.00	CNN with spectrograms for tool wear detection.	No sequential feature learning.
Kamat et al. [24] (2022)	CNN, AE-LSTM, k-NN	93.00	Compared DL (CNN, AE-LSTM) and traditional k-NN on sensor data.	CNN best overall; AE-LSTM reduced false positives; k-NN struggled with noise.
Wu et al. [25] (2019)	CNN(ToolWearNet) with CAE and BP-SGD	96.20	Image-based tool wear detection using CNN and CAE.	Lacks transformer-based learning.
Tobon-Mejia et al. [26] (2012)	MoG-HMM with Dynamic Bayesian Networks (DBN)	89.5	DBN-based tool wear and RUL prediction using sensor features.	Strong RUL prediction; future work to address variable conditions and maintenance integration.

Table 1 summarizes key research contributions in this domain. Kumar et al. [20] implemented a vision-based tool wear monitoring system using EfficientNet-B0 with transfer learning, achieving 94.11% accuracy, but their study did not explore self-attention or hybrid architectures. Xu et al. [21] introduced a multi-scale CNN-BiLSTM-GCN model, which enhanced spatial-temporal feature extraction and achieved 98.50% accuracy; however, their approach lacked explicit attention mechanisms, such as Transformers, for feature refinement. He et al. [22] developed a semi-supervised LSTM-based tool wear identification model, leveraging both labeled and unlabeled data to improve accuracy (95.78%), but it did not incorporate attention-based enhancements for global dependency learning. Karabacak et al. [23] utilized CNN models (GoogleNet, AlexNet, ResNet-50, EfficientNet-B0) for tool wear detection with spectrogram

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 12, Issue 06, June 2025

analysis, achieving 99% accuracy; however, the study did not focus on sequential feature learning. Kamat et al. [24] compared deep learning models (CNN, AE-LSTM) and a traditional k-NN classifier on sensor data for tool wear classification. The CNN model achieved the highest accuracy of 93.0%, with AE-LSTM showing better false positive reduction, while k-NN struggled with noisy and high-dimensional data. Wu et al. [25] proposed CNN (ToolWearNet) with a convolutional autoencoder (CAE) and backpropagation-based fine-tuning, achieving 96.20% accuracy, but their approach did not explore Transformer-based architectures for enhanced learning. Similarly, Tobon-Mejia et al. [26] developed a diagnostic and prognostic model using MoG-HMM combined with Dynamic Bayesian Networks (DBN) to estimate tool wear and predict Remaining Useful Life (RUL) based on sensor features, achieving an accuracy of 89.5%. Their approach showed strong RUL prediction performance but lacked adaptability to varying operating conditions and maintenance integration.

A. Research Gaps

1. Most existing studies rely on CNN, LSTM, or hybrid models, but they do not incorporate attention-based architectures like Transformers or self-attention mechanisms, which can refine feature extraction and improve predictive performance.
2. While CNN-based models effectively capture spatial features, they struggle with learning sequential dependencies in tool wear progression. Some hybrid models, such as those combining CNN and BiLSTM, enhance temporal feature extraction, but they lack explicit attention-based components to improve global context understanding.
3. Additionally, there has been no comprehensive comparison of different attention-based models to evaluate their effectiveness in tool wear detection.

Hence, this study aims to address these limitations by conducting a comparative analysis of multiple attention-based DL models for tool wear detection. By evaluating the performance of Transformer architectures and self-attention mechanisms against traditional CNN, LSTM, and hybrid approaches, this study seeks to determine the optimal model for spatial-temporal feature extraction and real-time tool wear monitoring.

III. RESEARCH METHODOLOGY

This section outlines the systematic approach followed to conduct the research. The study uses an open-source dataset to perform tool wear detection using four attention based LD models. The methodology comprises several stages, starting with data preprocessing, followed by model selection and training. The performance of each model is evaluated using various key metrics. Finally, a comparative analysis is conducted to identify the most effective model. Fig. 2 illustrates the entire workflow.

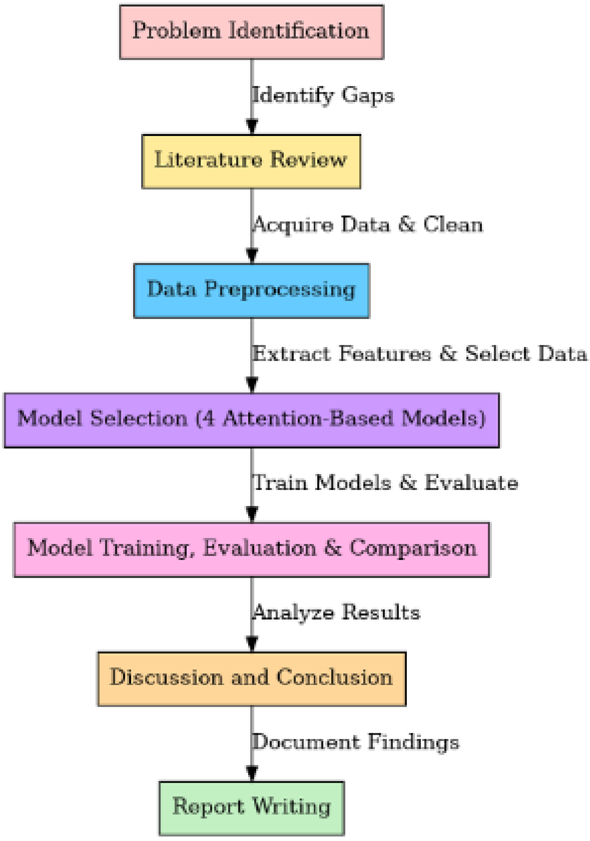

Fig 2: Research Methodology Flowchart

A. Dataset Description

The dataset originates from the University of Michigan SMART Lab and is designed for tool wear detection in CNC milling machining experiments. This dataset can be accessed at https://www.kaggle.com/datasets/shasun/tool-wear-detection-in-cnc-mill. It was obtained from the Kaggle repository. The dataset includes multiple experiments, each focusing on different machining conditions such as tool condition, feed rate, and clamping pressure.

1) Tool and Material Specification

Tool Specification

Tool type: CNC Milling Cutter.

Condition: Worn and Unworn Tools (Eight experiments with unworn tools and ten with worn tools).

Purpose: The tool was used to machine an ”S” shape into the workpiece, representing smart manufacturing processes.

Material: The tool material, carbide or high-speed steel (HSS) is typically used for CNC milling due to their durability, wear resistance, and ability to maintain sharp cutting edges under varying conditions.

Material Specification

Workpiece material: Wax

Dimensions: 2” x 2” x 1.5” (approximately 50.8 mm x 50.8

IFERP Explore Your Research Journey...

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 12, Issue 06, June 2025

mm x 38.1 mm).

Purpose: Wax blocks were selected as the workpiece material to simulate realistic machining operations. They are often used in controlled experiments because they are easy to machine, minimize tool damage, and clearly show wear effects on the cutting tool.

Inspection method: After the machining process, visual inspections were performed to determine whether the workpiece passed quality standards, aiding in tool wear detection analysis.

This tool wear dataset serves as a valuable resource for predictive modeling in manufacturing diagnostics. The dataset comprises multiple CSV files, including experiment-wise data (experiment 01.csv to experiment 18.csv) and a consolidated dataset (main.csv). These files contain sensor readings and machining parameters, which are crucial for identifying worn and unworn tools. Tables 2 summarize the key features of the main.csv dataset and the machining dataset, respectively, providing an overview of the relevant attributes used in the analysis.

Table 2: Summary of Features in main.csv

Feature	Description
Experiment Number	Identifier for the specific machining experiment.
Material	Type of material used in the milling process (e.g., wax).
Feedrate	Speed at which the tool moves during the milling operation.
Clamp Pressure	Pressure applied to hold the workpiece securely.
Tool Condition	Binary label indicating whether the tool is worn or unworn.
Machining Finalized	Indicates if the milling process was completed.
Passed Visual Inspection	Denotes whether the final machined product met quality standards.

B. Data Preprocessing

Data preprocessing is a crucial step in ML that ensures the raw data is cleaned, structured, and transformed for optimal model performance [27]. This process involves handling missing values, normalizing features, and encoding categorical variables to improve data quality and consistency [28].

Handling Missing Values: The dataset contains missing values in the passed visual inspection column. To ensure consistency in the dataset, all missing values were filled with the categorical value ’no’:

$$X_{\text{new}} = \begin{cases} X, & \text{if } X \neq \text{NaN} \\ \text{'no'}, & \text{if } X = \text{NaN} \end{cases} \quad (1)$$

Feature Engineering:

1. Experiment Tracking: A new column exp_num was created to keep track of different experimental conditions.
2. Adding Experiment-Specific Settings: Additional attributes, including material, feedrate, and clamp_ pressure, were incorporated to provide contextual information for each machining process.
3. Merging Experiment Results: Labels such as tool_condition, machining_finalized, and passed_visual_inspection were integrated into the dataset to associate conditions with outcomes.

DataSplitting: The dataset was split into training and testing sets using an 80:20 ratio:

$$D_{\text{train}}, D_{\text{test}} = \text{train\_test\_split}(D, \text{test\_size} = 0.2) \quad (2)$$

where D represents the original dataset, and D_{train} and D_{test} are the resulting training and testing sets, respectively.

Data Normalization: To ensure numerical features have comparable scales, StandardScaler() was applied. Each feature X was standardized as follows:

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma} \quad (3)$$

where μ is the mean and σ is the standard deviation of the feature.

Categorical Encoding: Categorical variables were converted into numerical representations using LabelEncoder(), which assigns unique integers to each category:

$$X_{\text{encoded}} = f(X) \quad (4)$$

where f is the mapping function that assigns each category a distinct numerical value.

C. Attention Mechanism Based DL Models for Tool Wear Detection

To improve the accuracy of tool wear detection, different attention-based DL models are applied and compared. The key models used in the study are described below.

1. *CNN with Attention:*

This model uses self-attention to capture long-range dependencies in sensor readings. Unlike CNN, which captures local spatial dependencies, self-attention assigns dynamic importance scores to all input features.

2. *LSTM with Attention:*

This model relies solely on LSTM for temporal modeling and applies an attention mechanism to focus on the most significant time steps. The attention mechanism computes an alignment score for each hidden state in the LSTM output, allowing the model to learn which time steps contribute most to predicting tool wear.

3. *Hybrid CNN-LSTM with Attention:*

This model integrates CNNs for spatial feature extraction with LSTM networks for temporal sequence modeling. CNN

IFERP Explore Your Research Journey...

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 12, Issue 06, June 2025

captures spatial de pendencies in the sensor data, while LSTM captures the time-series relationships. An attention mechanism is introduced to dynamically assign different importance weights to different time steps in the sequence.

4. *Transformer Encoder:*

This model uses a Transformer encoder, which consists of multi-head self attention followed by a position-wise feedforward network.

IV. RESULTS AND DISCUSSIONS

This section presents a detailed analysis of the performance and configuration of attention-based deep learning models for tool wear detection. The evaluation focuses on four models: CNN with attention, LSTM with attention, Hybrid CNN-LSTM with attention, and Transformer encoder, comparing their performance metrics and hyperparameter configurations.

A. Attention-Based DL Models Comparison

Table 3 and Fig. 3 shows the comparison of various attention-based DL models applied for the tool wear detection. Among the models, the Hybrid CNN-LSTM with attention exhibits the best performance, achieving an accuracy of 99.85%, precision of 99.87%, recall of 99.83%, and an F1-score of 99.85%. These results highlight the benefits of combining CNN's spatial feature extraction capabilities with LSTM's ability to capture temporal dependencies, further enhanced by attention mechanisms.

Table 3: Comparison of attention-based DL models for tool wear detection

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN with attention	99.15	99.58	98.80	99.19
LSTM with attention	99.69	99.72	99.65	99.68
Hybrid CNN-LSTM with attention	**99.85**	**99.87**	**99.83**	**99.85**
Transformer encoder	89.3	89.82	89.86	89.84

The LSTM with attention model also performs well, with an accuracy of 99.69%, precision of 99.72%, recall of 99.65%, and an F1-score of 99.68%. This model leverages LSTM's strengths in handling sequential data, complemented by attention mechanisms to focus on the most relevant time steps. However, it slightly underperforms compared to the hybrid model.

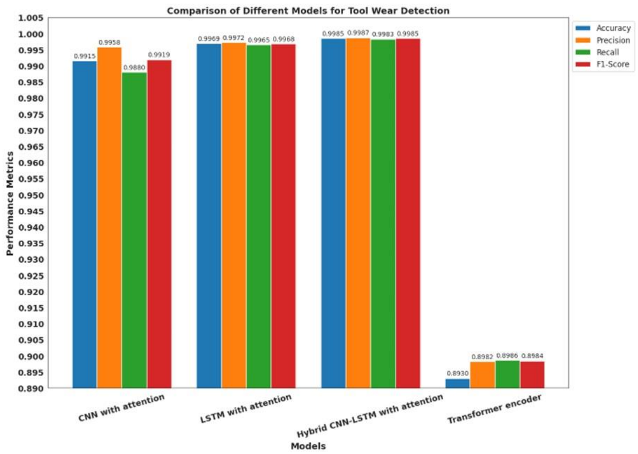

Fig. 3: Comparison of different models for tool wear detection

The CNN with attention model achieves an accuracy of 99.15%, precision of 99.58%, recall of 98.80%, and an F1-score of 99.19%. While the model effectively extracts features from the input data, its lower recall suggests occasional difficulty in identifying all relevant instances, possibly due to limitations in capturing temporal relationships.

In contrast, the Transformer encoder exhibits a significantly lower performance, with an accuracy of 89.3%, precision of 89.82%, recall of 89.86%, and an F1-score of 89.84%. This underperformance may stem from the model's dependency on large datasets for effective training and a higher sensitivity to hyperparameter tuning.

B. Hyperparameter Configurations

Table 4 details the hyperparameters used for each model, which significantly influence their performance. The CNN with attention employs Conv1D layers with 64 and 128 filters, a kernel size of 3, ReLU activation, and a dropout rate of 0.2. These settings enable efficient feature extraction and regularization. The LSTM with attention is configured with

Table 4: Hyperparameters for various attention-based DL models

Model	Hyperparameters
CNN with attention	Conv1D filters: 64, 128 Kernel size: 3 Activation: ReLU Dropout: 0.2
LSTM with attention	LSTM units: 64 Return sequences: True Dropout: 0.2
Hybrid CNN-LSTM with attention	Conv1D filters: 64, 128 LSTM units: 64 Multi-Head Attention heads: 4 Key dimension: 32 Dropout: 0.2
Transformer encoder	Dense layer units: 128, 64 Multi-Head Attention heads: 4 Key dimension: 32 Dropout: 0.2

64 LSTM units, return sequences enabled, and a dropout rate of 0.2. This setup ensures the model retains sequence information while mitigating overfitting.

The Hybrid CNN-LSTM with attention combines the strengths of both architectures. It uses Conv1D layers (64 and 128 filters), 64 LSTM units, and multi-head attention with 4 heads, a key dimension of 32, and a dropout rate of 0.2. This integration optimizes both spatial and temporal feature learning while focusing on the most critical features through attention mechanisms.

The Transformer encoder is designed with dense layers of 128 and 64 units, multi head attention with 4 heads, a key dimension of 32, and a dropout rate of 0.2. Despite its sophisticated architecture, the model's dependency on a large dataset and intricate hyperparameter optimization affects its effectiveness in this task.

C. Insights and Observations

The results indicate that the Hybrid CNN-LSTM with attention is the most effective model for tool wear detection, consistently outperforming other models across all metrics. While the Transformer encoder shows potential, its performance suggests that it may require further optimization or larger datasets to fully leverage its capabilities. It is notable that both the CNN with attention and LSTM with attention models achieve slightly lower accuracy than the hybrid CNN-LSTM model. Although these differences may appear numerically small, in tasks where high accuracy is crucial (such as tool wear detection), even a fraction of a percent can signify a meaningful improvement. However, the results indicate that the individual models are still highly accurate and perform well, but not at the level of the hybrid model.

D. Performance of Proposed Model- Hybrid CNN LSTM with Attention

The proposed hybrid CNN-LSTM with attention model emerged as the most effective for detecting tool wear in CNC milling operations. By combining the strengths of both CNNs and LSTM networks, it demonstrated exceptional performance across key evaluation metrics, including accuracy, precision, recall and F1-score.

1) Model Architecture Effectiveness

CNNs efficiently captured spatial patterns from sensor data, particularly from the X, Y, Z axis, and spindle motor measurements. The spatial feature extraction enabled the model to detect localized patterns associated with tool wear.

LSTM networks are instrumental in capturing temporal dependencies in the time-series data. Since CNC milling data exhibits strong temporal correlations, the LSTM layer effectively analyzed these sequences to predict tool wear conditions.

Attention mechanism further enhanced the model's capability by selectively focusing on the most relevant features during specific machining operations. This ensured that the model prioritized critical time steps and sensor readings that exhibited significant tool wear characteristics.

The proposed hybrid CNN-LSTM with attention model consistently outperformed other models, achieving the accuracy 99.85%, precision 99.87%, recall 99.83% and F1-score 99.85%. These results highlight the robustness and reliability of the model in accurately classifying worn and unworn tools, minimizing false positives and false negatives.

2) Practical Implications

The outstanding performance of the proposed hybrid CNN-LSTM with attention model makes it a strong candidate for real-time tool wear detection systems in manufacturing industries. Implementing this model can lead to:

- Improved predictive maintenance.
- Reduction in machine downtime.
- Enhanced product quality.
- Lower operational costs.

V. CONCLUSION

Tool wear detection plays a critical role in manufacturing, directly impacting productivity, quality, and operational efficiency. Accurate detection of tool wear helps prevent equipment failures, reduces downtime, and enhances overall process reliability. Traditional approaches often rely on statistical methods or ML algorithms, which may struggle with the complexity and non-linear patterns in tool wear data. In this study, attention-based DL models have been explored for tool wear detection, leveraging their ability to process intricate patterns and focus on the most relevant features in the data.

The evaluation of four models—CNN with attention, LSTM with attention, Hybrid CNN-LSTM with attention, and Transformer encoder—revealed that attention mechanisms significantly enhance the performance of deep learning models by directing focus to critical features. Among these, the Hybrid CNN-LSTM with attention achieved the best results, demonstrating the effectiveness of combining CNN's spatial feature extraction and LSTM's temporal learning capabilities. The superior performance of this model highlights the importance of synergistic architectures in tackling complex predictive tasks like tool wear detection.

While individual CNN and LSTM models performed well, their slightly lower accuracy emphasized the limitations of relying solely on spatial or temporal feature learning. The Transformer encoder, while conceptually powerful, underperformed due to its sensitivity to hyperparameters and reliance on large datasets, suggesting it requires further optimization to match the other models' performance.

A. Future Work

While this study demonstrates the potential of attention-based DL models for tool wear detection, several

IFERP Explore Your Research Journey...

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 12, Issue 06, June 2025

areas of future work could further enhance the research:

1. Real-Time Implementation: Developing real-time tool wear detection systems with optimized hybrid models could improve practical applicability in industrial settings. Efforts can focus on reducing computational overhead to meet the constraints of real-time processing.
2. Transformer Optimization: Further investigation into optimizing Transformer based models, including pretraining on large-scale datasets or fine-tuning architectures, may unlock their full potential for tool wear detection.
3. Explainability and Interpretability: Enhancing the interpretability of attention based models can provide deeper insights into the factors influencing tool wear, enabling better decision-making in industrial applications.

REFERENCES

[1] I. E. Ekengwu and K. I. Emeruwa, "Comprehensive review of computer numerical control (cnc) systems," International Research Journal of Scientific Studies, vol. 1, no. 1, pp. 50–55, 2024.

[2] T. Mohanraj, E. Kirubakaran, D. K. Madheswaran, M. Naren, M. Ibrahim et al., "Review of advances in tool condition monitoring techniques in the milling process," Measurement Science and Technology, vol. 35, no. 9, p. 092002, 2024.

[3] A. Mohamed, M. Hassan, R.M'Saoubi, and H. Attia, "Tool condition monitoring for high-performance machining systems—a review," Sensors, vol. 22, no. 6, p. 2206, 2022.

[4] K. Aslantas, A. Hasc¸elik, A. Erc¸etin, M. Danish, L. K. Alatrushi, S. Rubaiee, and A. B. Mahfouz, "Effect of cutting conditions on tool wear and wear mechanism in micro-milling of additively manufactured titanium alloy," Tribology International, vol. 193, p. 109340, 2024.

[5] M. Bhuiyan and I. Choudhury, "13.22—review of sensor applications in tool condition monitoring in machining," Comprehensive Materials Processing, vol. 13, pp. 539–569, 2014.

[6] J. Lee, J. Ni, J. Singh, B. Jiang, M. Azamfar, and J. Feng, "Intelligent maintenance systems and predictive manufacturing," Journal of Manufacturing Science and Engineering, vol. 142, no. 11, p. 110805, 2020.

[7] S. Kasiviswanathan, S. Gnanasekaran, M. Thangamuthu, and J. Rakkiyannan, "Machine-learning-and internet-of-things-driven techniques for monitoring tool. wear in machining process: a comprehensive review," Journal of Sensor and Actuator Networks, vol. 13, no. 5, p. 53, 2024.

[8] S. Ayvaz and K. Alpay, "Predictive maintenance system for production lines in manufacturing: A machine learning approach using iot data in real-time," Expert Systems with Applications, vol. 173, p. 114598, 2021.

[9] C. Scheffer and P. Heyns, "An industrial tool wear monitoring system for interrupted turning," Mechanical Systems and Signal Processing, vol. 18, no. 5, pp. 1219–1242, 2004.

[10] I. M. Fielden, Investigation of microstructural evolution by real-time SEM of high-temperature specimens. Sheffield Hallam University (United Kingdom), 2005.

[11] D. Y. Pimenov, M. K. Gupta, L. R. da Silva, M. Kiran, N. Khanna, and G. M. Krolczyk, "Application of measurement systems in tool condition monitoring of milling: A review of measurement science approach," Measurement, vol. 199, p. 111503, 2022.

[12] T. Schneider, N. Helwig, and A. Sch¨ utze, "Industrial condition monitoring with smart sensors using automated feature extraction and selection," Measurement Science and Technology, vol. 29, no. 9, p. 094002, 2018.

[13] Z. He, T. Shi, J. Xuan, and T. Li, "Research on tool wear prediction based on temperature signals and deep learning," Wear, vol. 478, p. 203902, 2021.

[14] D. Y. Pimenov, A. Bustillo, S. Wojciechowski, V. S. Sharma, M. K. Gupta, and M. Kunto˘ glu, "Artificial intelligence systems for tool condition monitoring in machining: Analysis and critical review," Journal of Intelligent Manufacturing, vol. 34, no. 5, pp. 2079–2121, 2023.

[15] Y. Zhu, M. Wang, X. Yin, J. Zhang, E. Meijering, and J. Hu, "Deep learning in diverse intelligent sensor based systems," Sensors, vol. 23, no. 1, p. 62, 2022.

[16] G. Brauwers and F. Frasincar, "A general survey on attention mechanisms in deep learning," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 4, pp. 3279–3298, 2021.

[17] S. Y. Wong, J. H. Chuah, and H. J. Yap, "Technical data-driven tool condition monitoring challenges for cnc milling: a review," The International Journal of Advanced Manufacturing Technology, vol. 107, pp. 4837–4857, 2020.

[18] A. Singh, A. Jadhav, and P. Singh, "Ai applications in production," Industry 4.0, Smart Manufacturing, and Industrial Engineering, pp. 139–161.

[19] V. Nasir and F. Sassani, "A review on deep learning in machining and tool monitoring: Methods, opportunities, and challenges," The International Journal of Advanced Manufacturing Technology, vol. 115, no. 9, pp. 2683–2709, 2021.

[20] A. S. Kumar, A. Agarwal, V. G. Jansari, K. Desai, C. Chattopadhyay, and L. Mears, "Realizing on-machine tool wear monitoring through integration of vision-based system with cnc milling machine," Journal of Manufacturing Systems, vol. 78, pp. 283–293, 2025.

[21] Z. Xu, B. Zhang, L. L. Fan, E. H. Yan, D. Li, Z. Zhao, W. S. Yip, and S. To, "Deep-learning-driven intelligent tool wear identification of high-precision machining with multi-scale cnn-bilstm-gcn," Advanced Engineering Informatics, vol. 65, p. 103234, 2025.

[22] X. He, M. Zhong, C. He, J. Wu, H. Yang, Z. Zhao, W. Yang, C. Jing, Y. Li, and G. Chen, "A novel tool wear identification method based on a semi-supervised lstm," Lubricants, vol. 13, no. 2, p. 72, 2025.

[23] Y. KARABACAK, "Deep learning-based cnc milling tool wear stage estimation with multi-signal analysis," Eksploatacja i Niezawodnosc-Maintenance and Reliability, vol. 25, no. 3, 2023.

[24] P. V. Kamat, A. Nargund, S. Kumar, S. Patil, and R. Sugandhi, "Tool wear prediction in milling: A comparative analysis based on machine learning and deep learning approaches," International Journal of Computing and Digital System, vol. 3, pp. 206–211, 2021.

[25] X. Wu, Y. Liu, X. Zhou, and A. Mou, "Automatic identification of tool wear based on convolutional neural network in face milling process," Sensors, vol. 19, no. 18, p.

3817, 2019.

[26] D. A. Tobon-Mejia, K. Medjaher, and N. Zerhouni, “Cnc machine tool’s wear diagnostic and prognostic by using dynamic bayesian networks,” Mechanical Systems and Signal Processing, vol. 28, pp. 167–182, 2012.

[27] M. Kang and J. Tian, “Machine learning: Data pre-processing,” Prognostics and health management of electronics: fundamentals, machine learning, and the internet of things, pp. 111–130, 2018.

[28] M. Saar-Tsechansky and F. Provost, “Handling missing values when applying classification models,” 2007.

[29] V. Labatut and H. Cherifi, “Accuracy measures for the comparison of classifiers,” arXiv preprint arXiv:1207.3790, 2012.

[30] D. M. Powers, “Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation,” arXiv preprint arXiv:2010.16061, 2020.

[31] A. H. Fielding and J. F. Bell, “A review of methods for the assessment of prediction errors in conservation presence/absence models,” Environmental conservation, vol. 24, no. 1, pp. 38–49, 1997.

[32] R. Diallo, C. Edalo, and O. O. Awe, “Machine learning evaluation of imbalanced health data: A comparative analysis of balanced accuracy, mcc, and f1 score,” in Practical Statistical Learning and Data Science Methods: Case Studies from LISA 2020 Global Network, USA. Springer, 2024, pp. 283–312.