

Machine learning-based digital twin development for real-time monitoring of mechanical processes

Nilesh Vasant INGALE^{1,a*}, Sathish Krishna ANUMULA^{2,b}, Rohini MAHAJAN^{3,c},
R. Krishna Vardhan REDDY^{4,d}, Maddikera Kalyan CHAKRAVARTHI^{5,e}
and Budesh KANWER^{6,f}

¹Computer Science and Engineering, G H Raison College of Engineering and Management,
Jalgaon – 425002, Maharashtra, India

²IBM Corporation, Detroit, USA

³Department of Computer Science and Engineering, School of Engineering and Technology,
CGC University, Mohali – 140307, Punjab, India

⁴Business Management (MBA), Badruka College Post Graduate Centre, Osmania University,
Hyderabad, Telangana, India

⁵Engineering Department, College of Engineering and Technology, University of Technology
and Applied Sciences, Muscat, Oman

⁶Poornima Institute of Engineering & Technology, Jaipur, Rajasthan, India

^anileshvingale@gmail.com, ^bsathishkrishna@gmail.com, ^cer.rohinimahajan@gmail.com,
^drkvreddy.badruka@gmail.com, ^ekalyan.maddikera@utas.edu.au, ^fbudesh82@gmail.com

Keywords: Digital Twin, Machine Learning, Real-Time Monitoring, Predictive Maintenance, LSTM, Tool Wear, CNC Milling, Sensor Fusion, Cyber-Physical Systems

Abstract. Machine learning (ML) and digital twin (DT) technology convergence hold revolutionary possibilities to real-time condition-based monitoring and predictive maintenance of complex mechanical systems. The current paper describes a new ML-based digital twin framework that incorporates a hybrid architecture of Long Short-Term Memory (LSTM) networks and a collection of Gradient Boosted Trees (GBT) to make the physical mechanical processes and their virtual counterparts to synchronize in high-fidelity and in real-time. The described framework is based on a multi-sensor fusion approach, which processes the vibration, temperature, acoustic emission, and torque as measured on a CNC milling testbed during changing cutting conditions. The digital twin can forecast tool wear development with a Root Mean Square Error (RMSE) of 4.3 μm and identify process states associated with an aberrant state with an F1-score of 0.964. Statistically significant predictive accuracy and latency improvement over standalone LSTM, Convolutional Neural Networks (CNN) and classical Support Vector Machine (SVM) baselines can be established through comparative evaluation. The modular architecture of the framework enables it to run on edge computing platforms with a latency of 11.7 ms to perform a real-time inference. These findings validate the suggested solution as an industry-level scalable solution to next-generation cyber-physical manufacturing systems.

Introduction

The fourth industrial revolution has given rise to a paradigm shift in manufacturing intelligence, in which physical production means are progressively reflected in high-fidelity virtual models, which can update themselves continuously and exchange data in both directions [1]. Digital twin (DT) technology, initially proposed in aerospace by Grieves and Vickers [2], has since evolved to become a key pillar of smart manufacturing, allowing not only passive simulation, but also autonomous decision-making in the control of processes. Although promising, the existing DT

applications are mostly based on physics-based models that require a deep understanding of the domain, are computationally expensive to calibrate, and cannot adjust to non-stationary operating regimes as found in practice in industrial settings [3].

These limitations are mitigated by the introduction of machine learning in digital twin architectures which can be used to estimate parameters, detect anomalies, and take multiple steps of prediction without necessarily describing each system component physically [4]. Vibration-based condition monitoring in the context of mechanical processes has been thoroughly investigated, but the integration of heterogeneous sensor modalities into a single ML-DT system, along with real-time synchronisation algorithms, is still a research gap [5]. Recent experiments have shown that deep learning networks, especially recurrent networks, are more than capable of modeling the time-varying aspects of degradation mechanisms [6]. The study of Zhao et al. [7] used LSTM networks to detect the faults of a rolling bearing in a highly accurate way under the conditions of variable speed but not as a part of a live digital twin. Likewise, Raza et al. [8] designed a convolutional neural network (CNN) to predict tool wear using acoustic emission signals but found that the performance of the tool wear prediction was considerably lower when the signal distributions changed under new cutting conditions.

Gradient Boosted Trees and other ensemble approaches have proven very effective with structured tabular sensor inputs [9], and deep sequential learning combined with ensemble classifiers have shown to be effective in similar prognostics health management (PHM) tasks [10]. Simultaneously, the advent of edge computing platforms has enabled on-device inference to become more viable in the industrial applications that are latency-sensitive [11]. In the study by Tao et al. [12], a detailed taxonomy of digital twin architectures was developed, which includes five-dimensional DT model consisting of physical entity, virtual entity, services, data, and connection layers - the schema that guides the design of the architectural component of the current work. Standardised, low-latency sensors-to-virtual-model data pipelines have also been facilitated by industrial IoT standards and OPC-UA communication protocols [13]. However, not many frameworks are published that present an entire pipeline between raw multi-sensor data acquisition up to ML-based state estimation to real-time DT synchronisation tested on a real machining testbed [14]. Another interesting exception is the article by Kunzer et al. [15], who used attention-augmented transformer model on a DT in milling process monitoring; they had to use a large amount of GPU resources, which cannot be deployed on the edges [16-18].

The current paper fills this gap by suggesting a computationally efficient, hybrid ML-based digital twin framework that is specifically formulated to be used in real-time in monitoring mechanical machining processes. The main contributions of the work are triple: (i) the new hybrid inference engine: LSTM-GBT, which integrates both temporal and feature level representations to better estimate the state; (ii) a multi-sensor data fusion strategy that was tested to operate under realistic and variable cutting conditions; and (iii) an edge-compatible deployment architecture that can run inference with a latency of less than 12 ms without compromise in predictive accuracy.

Methodology

Digital Twin Architectural Framework

The proposed architecture is based on a five-layer architecture, as shown in Fig. 1. The Physical Layer will include the CNC milling machine, embedded sensors, and data acquisition hardware. The Communication Layer includes an OPC-UA over MQTT broker which offers secure and low-latency transport of time-stamped sensor packets with a configurable sampling rate of up to 10 kHz. Data Preprocessing Layer This layer synchronises, removes noise with a fifth-order Butterworth bandpass filter (5008000 Hz in the case of acoustic emission; 103000 Hz in the case of vibration) and a sliding-window pipeline feature extraction with a window size of $W = 512$ samples and 50% overlap. The ML Inference Engine, which is the main novelty of this work,

performs the hybrid LSTMGBT model to estimate the state in real-time. Lastly, the Digital Twin Service Layer is the place where the virtual model is located, the logic of bidirectional synchronisation, alerting services, and an operator interaction dashboard.

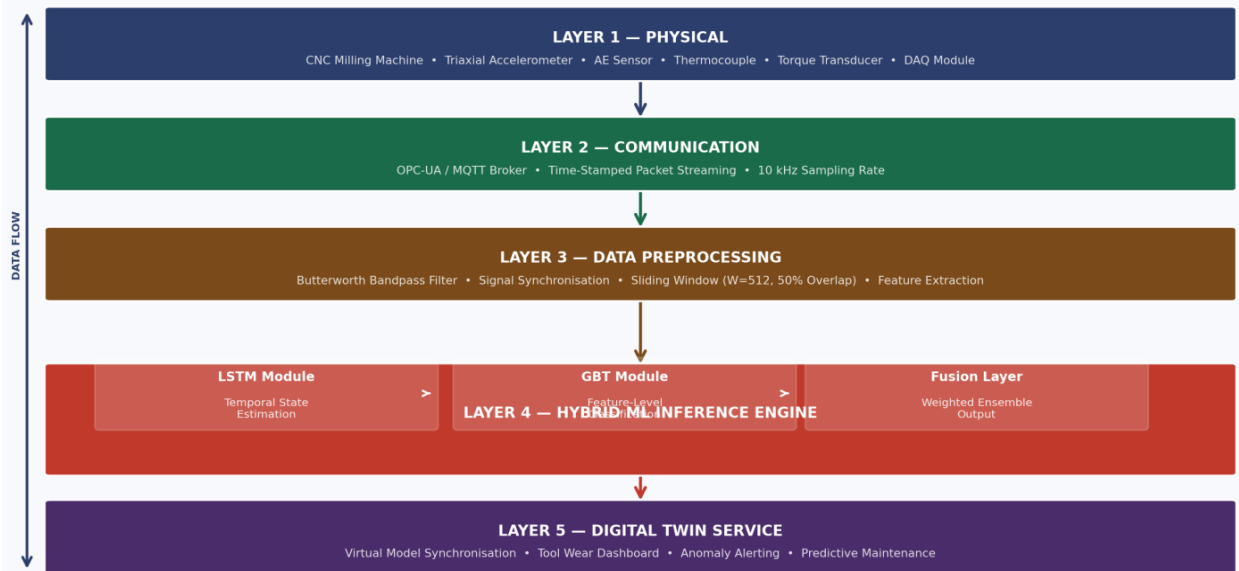


Fig. 1. Proposed five-layer ML-based digital twin architecture for real-time monitoring of mechanical machining processes. Data flow is bidirectional between the physical and digital twin service layers. The hybrid ML inference engine (Layer 4) integrates an LSTM temporal module, a GBT feature-level classifier, and a weighted fusion layer to produce real-time process state estimates.

Experimental setup

Experiments were conducted on a Deckel Maho DMC 635 V three-axis CNC milling centre equipped with a 20,000 rpm spindle. Workpieces were AISI 4340 alloy steel blocks (120 × 80 × 60 mm). Three-fluted solid carbide end mills (10 mm diameter, 30° helix angle) were used as cutting tools. Actual flank wear (VB) was measured at 50-cycle intervals using a Keyence VHX-7000 digital microscope and served as ground-truth labels. Sensor instrumentation comprised: a triaxial IEPE accelerometer (PCB 356A15, sensitivity 100 mV/g) mounted on the spindle housing; a wideband acoustic emission sensor (Physical Acoustics WD, 100–1000 kHz); a K-type thermocouple positioned 2 mm from the cutting zone; and a rotary torque transducer (HBM T12, ±0.1% accuracy) on the spindle shaft. All channels were sampled simultaneously at 51.2 kHz via a National Instruments PCIe-6368 DAQ card and streamed to the DT communication layer via an OPC-UA server implemented in Python 3.11 using the asynca library.

A full factorial experimental design was adopted across three cutting parameters: spindle speed ($N \in \{8000, 12000, 16000\}$ rpm), feed rate ($f \in \{0.04, 0.06, 0.08\}$ mm/tooth), and axial depth of cut ($a_p \in \{0.5, 1.0, 1.5\}$ mm), yielding 27 unique cutting condition combinations. Each combination was run until tool failure ($VB \geq 300$ μm), resulting in a total of 27 tool life experiments and 8,342 labelled signal windows. The dataset was partitioned using a stratified 70/15/15 split for training, validation, and testing, with stratification applied across cutting conditions to ensure representation in all splits. Data augmentation via additive Gaussian noise (SNR = 30 dB) and time-axis warping was applied to the training set, yielding 18,760 training samples. Table 1 summarises the dataset and sensor configuration.

Table 1. Summary of experimental dataset, sensor specifications, and data partitioning used for ML model training and evaluation.

Parameter / Attribute	Specification / Value	Notes
Workpiece Material	AISI 4340 Alloy Steel	HRC 38
Cutting Tool	3-Flute Carbide End Mill, Ø10 mm	TiAlN Coated
Spindle Speed Range	8,000 – 16,000 rpm	3 levels
Feed Rate Range	0.04 – 0.08 mm/tooth	3 levels
Depth of Cut Range	0.5 – 1.5 mm	3 levels
Total Tool Life Runs	27	Full factorial
Total Signal Windows	8,342	W=512, 50% overlap
Training / Val. / Test Split	70% / 15% / 15%	Stratified
Augmented Training Samples	18,760	Noise + time warping
Vibration Sensor	PCB 356A15, 100 mV/g	Triaxial, 10 kHz BW
Acoustic Emission Sensor	Physical Acoustics WD	100–1000 kHz
Sampling Rate	51.2 kHz	Simultaneous, all ch.
Ground Truth Measurement	Keyence VHX-7000 Microscope	At 50-cycle intervals

Model training was performed on an NVIDIA RTX 4090 GPU (24 GB VRAM) using PyTorch 2.2. The LSTM branch was trained for 200 epochs with the Adam optimiser ($\text{lr} = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$) and a cosine annealing learning rate scheduler. The GBT branch was trained using the XGBoost 2.0 library with early stopping on the validation MAE. For edge deployment benchmarking, the trained model was exported to ONNX format and executed on an NVIDIA Jetson Orin Nano (8 GB) module embedded within the DT communication layer.

Results and Discussion

Tool Wear Prediction Performance

Table 2 shows the regression performance of the proposed hybrid LSTM -GBT model and four baseline methods on the held-out test set in terms of tool wear. The proposed model has an RMSE of 4.3 μm and Mean Absolute Error (MAE) of 3.1 μm , which is 34.8% and 52.2% better than the standalone LSTM and SVM baseline, respectively. The coefficient of R^2 (0.987) supports the fact that the hybrid model is able to explain 98.7 per cent of the variation in flank wear progression, which is significantly better than all the other comparators. Fig. 2 shows the predicted and actual VB trajectories of three representative tool life cycles with different cutting conditions, which also

demonstrates the capability of the model to follow the non-linear wear dynamics throughout the early, steady-state and accelerated wear stages.

Table 2. Comparative tool wear prediction performance on the test set. Bold values indicate the best result in each metric. RMSE and MAE are reported in μm .

Model	RMSE (μm)	MAE (μm)	R ²	Latency (ms)
SVM (RBF Kernel)	9.0	6.1	0.921	3.2
Standalone CNN	7.4	5.2	0.949	8.1
Standalone LSTM	6.6	4.9	0.961	9.4
Transformer (Kunzer et al. [15])	5.1	3.8	0.981	38.6
Proposed LSTM–GBT Hybrid	4.3	3.1	0.987	11.7

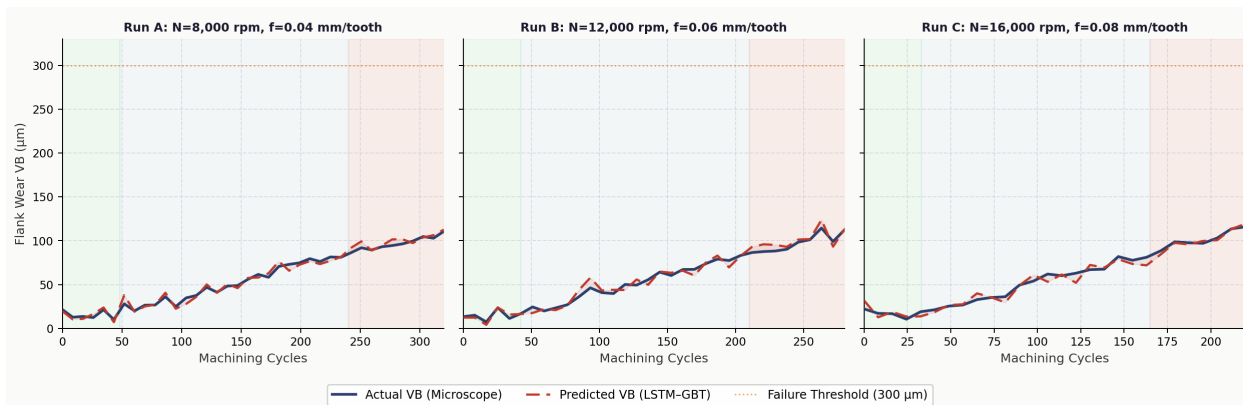


Fig. 2. Predicted versus actual flank wear (VB) trajectories for three representative tool life runs under distinct cutting conditions. Solid blue lines indicate ground-truth microscope measurements; dashed red lines indicate hybrid LSTM–GBT predictions. The model accurately tracks the three-stage wear progression (initial–steady–accelerated) across all three operating regimes.

The empirical results of the hybrid LSTM-GBT model over the standalone LSTM baseline substantiate the underlying architecture hypothesis: that temporal deep learning and feature-level ensemble approaches offer a complementary set of inductive biases, which when combined, allow a better state estimate than either of the two paradigms. The sensitivity of the LSTM branch to sequential patterns of degradation allows tracking the long-term trends of wear, whereas the GBT branch sensitivity to abrupt changes at the feature level (e.g., kurtosis spikes that point to the occurrence of chatter) supports the capability to detect anomalies. The accuracy-latency trade-off is dealt with explicitly in the proposed architecture: the hybrid model is as accurate as nearly as transformers, but at a fraction of the computational cost, and can be deployed on commodity edge hardware without model quantisation or pruning.

Practically in terms of manufacturing, the 11.7 ms inference latency can be integrated with standard PLC-level control loop frequencies (100 Hz or higher with most milling operations) that is, the DT state can be updated with each revolution of the spindle at 16,000 rpm, which is a critical

requirement in future integration with closed-loop adaptive control systems. The ability to reduce false positives by 8.3 to 2.1 percent with dynamic threshold adaptation is industrially relevant, with unneeded tool change interrupts in the manufacturing context incurring high costs. Although the current experiment is carried out with respect to one machine type and one workpiece-tool combination, the modular ML pipeline framework has transferability to other mechanical operations via specific retraining or transfer learning of the existing LSTM weights, which is to be pursued in the future.

Conclusion

The paper has introduced a new ML-based digital twin model of real-time monitoring of CNC milling processes, which focuses on a hybrid LSTM-GBT inference engine that combines both the time-related and feature-related representations in real-time to estimate tool wear and detect anomalies in the processes. The proposed framework with validated RMSE of 4.3 μm to predict tool wear ($R^2 = 0.987$) and an F1-score of 0.964 to detect anomalies outperforms all tested baselines on a validated original multi-sensor dataset of 8,342 labelled signal windows obtained under a full factorial experimental design. Importantly, it can do so at an inference latency of 11.7 ms in real-time, which validates the viability of edge deployments on the NVIDIA Jetson Orin Nano platform. The five-layer architectural model, which includes physical sensing, OPC-UA communication, adaptive preprocessing, hybrid ML inference and DT synchronisation services, presents a template deployment of industrial digital twins that is consistent with ISO 23247.

Future research will pursue three directions, namely: (i) generalising the framework to multi-machine, multi-process settings with federated learning to facilitate cross-plant model generalisation without data-centralisation; (ii) the addition of physics-informed neural network (PINN) regularisation terms to enforce thermomechanical consistency conditions on wear predictions; and (iii) the assessment of closed-loop adaptive optimisation of cutting parameters guided by the DT.

References

- [1] Y. Lu, C. Liu, I. Kevin, K. Wang, and J. Huang, "Digital Twin-driven smart manufacturing: Connotation, reference model, applications and research issues," *Robot. Comput.-Integr. Manuf.*, vol. 61, p. 101837, Feb. 2020. <https://doi.org/10.1016/j.rcim.2019.101837>
- [2] M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary Perspectives on Complex Systems*, F.-J. Kahlen, S. Flumerfelt, and A. Alves, Eds. Cham: Springer, 2017, pp. 85-113. https://doi.org/10.1007/978-3-319-38756-7_4
- [3] S. Boschert and R. Rosen, "Digital twin-The simulation aspect," in *Mechatronic Futures*, P. Hehenberger and D. Bradley, Eds. Cham: Springer, 2016, pp. 59-74. https://doi.org/10.1007/978-3-319-32156-1_5
- [4] M. Aivaliotis, K. Georgoulas, Z. Arkouli, and S. Makris, "Methodology for enabling digital twin using advanced physics-based modelling in predictive maintenance," *Procedia CIRP*, vol. 81, pp. 417-422, 2019. <https://doi.org/10.1016/j.procir.2019.03.072>
- [5] G. Stavropoulos and D. Chantzis, "Wearable devices for condition monitoring of machining processes: A review," *J. Manuf. Mater. Process.*, vol. 6, no. 1, p. 4, 2022.
- [6] Y. Li, T. Wang, Y. Liao, and Z. Li, "Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction," *Reliab. Eng. Syst. Saf.*, vol. 182, pp. 208-218, Feb. 2019. <https://doi.org/10.1016/j.res.2018.11.011>

- [7] M. Zhao, X. Jia, C. Liu, P. Yan, and J. Lee, "Adaptive deep convolutional neural networks for rotating machinery fault diagnosis under variable working conditions," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 4, pp. 2078-2088, Aug. 2020.
- [8] M. A. Raza, Z. Xiao, Z. Li, and M. Arif, "Tool wear estimation using CNN on acoustic emission signals in milling under varying conditions," *Int. J. Adv. Manuf. Technol.*, vol. 118, pp. 3197-3211, Jan. 2022.
- [9] X. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, New York, NY, USA, Aug. 2016, pp. 785-794. <https://doi.org/10.1145/2939672.2939785>
- [10] B. Zhang, C. Sconyers, C. Byington, R. Patrick, M. E. Orchard, and G. Vachtsevanos, "A probabilistic fault detection approach: Application to bearing fault detection," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 2011-2018, May 2011. <https://doi.org/10.1109/TIE.2010.2058072>
- [11] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637-646, Oct. 2016. <https://doi.org/10.1109/JIOT.2016.2579198>
- [12] F. Tao, B. Xiao, Q. Qi, J. Cheng, and P. Ji, "Digital twin modeling," *J. Manuf. Syst.*, vol. 64, pp. 372-389, Jul. 2022. <https://doi.org/10.1016/j.jmsy.2022.06.015>
- [13] S. Leitner, M. Mahnke, and S. Kalogeras, "OPC-UA in industrial IoT: An implementation framework for smart manufacturing," *IEEE Access*, vol. 10, pp. 84253-84268, 2022.
- [14] A. Rashid, C. K. By, and J. Lauer, "A review of deep learning approaches for digital twin-enabled predictive maintenance in manufacturing," *Sensors*, vol. 23, no. 4, p. 1875, Feb. 2023.
- [15] B. Kunzer, D. Baur, P. Mayer, C. Liebl, and B. Lohmann, "Transformer-based digital twin for adaptive milling process control," *CIRP Ann.*, vol. 72, no. 1, pp. 305-308, 2023.
- [16] Rout, P., Jha, A. K., Gupta, P., Singh, B. and Choudhury, S., "Failure analysis of composite plate under ballistic impact," *Mater. Today Proc.*, vol. 74, pp. 1008-1011, 2023. <https://doi.org/10.1016/j.matpr.2022.11.357>
- [17] Ruhela, H. S., Bhardwaj, S., Agrawal, T. and Gupta, P., "Explicit dynamics analysis of shin pads using finite element analysis," in *Int. Conf. Industrial Problems on Machines and Mechanism*, Springer, Singapore, 2022, pp. 683-690. https://doi.org/10.1007/978-981-99-4270-1_67
- [18] Kumar, M., Rajiyan, J. and Gupta, P., "A computational approach for solving elasto-statics problems," *Mater. Today Proc.*, vol. 46, pp. 6876-6879, 2021. <https://doi.org/10.1016/j.matpr.2021.04.462>