Robotic Machining Errors Prediction via In-Process Force and Vibration Sensors and Mixture of Experts

Chaoyue Niu[1,2*] and Visakan Kadirkamanathan[1]

***Abstract*—High-precision robotic machining remains challenging, as achieving accurate dimensional and geometric features requires feedback control based on reliable error prediction. Effective solutions must estimate geometric errors accurately while minimizing inspection costs and ideally enabling real-time corrective control. An established approach relies on costly laser tracker position sensing with data-driven modeling, limiting large-scale industrial deployment. This paper presents a machine learning framework for error prediction using in-process low-cost force and vibration sensor data, and robot controller signals. A mixture-of-experts (MoE) architecture is adopted, where the robot controller dynamically weights predictions from multiple experts, enabling complementary information fusion. Both handcrafted features and deep-learned features are developed and evaluated. Experiments on square and circular trajectories demonstrate that the proposed method is comparable to laser-tracker-learning-based models and surpasses direct laser tracker measurements, reducing the mean absolute error sevenfold from 71.4 microns to 9.2 microns relative to a coordinate measuring machine (CMM). The proposed framework enables sub-5 millisecond latency and an energy-delay product of 760 microjoule-second on a CPU board. These results indicate that low-cost sensing combined with machine learning can match expensive metrology and position sensing systems, providing a practical and scalable approach for high-precision manufacturing.**

***Index Terms*—Machining Error Prediction; Robotic Milling; Mixture of Experts; Low-cost In-process Sensing; Machine Learning.**

I. Introduction

Intelligent manufacturing is undergoing rapid transformation driven by technologies such as autonomous production and digitalised value chains. These trends require factories to be faster, more agile, and customer-oriented. They must deliver mass-customised products that are affordable and available on demand. Within this context, metrology [1] plays a key role in ensuring quality while improving productivity. Emerging metrology data analytics aim to unify information across design, production, and verification. They create standardised knowledge systems for intelligent manufacturing. Robotic machining [2], especially for high-force milling, has attracted attention as a flexible and cost-efficient alternative to conventional computer numerical control (CNC) machining. This is particularly valuable for large or complex workpieces. However, robotic machining often suffers from errors due to limited structural rigidity, complex tool–workpiece interactions, and environmental disturbances. These factors restrict robotic machining in sectors that require extremely high accuracy, such as aerospace and automotive. Integrating robotic systems with advanced metrology analytics offers a promising approach. It enables real-time error prediction and adaptive process optimisation to meet strict quality requirements in precision manufacturing.

To address accuracy challenges in robotic machining, diverse strategies such as error compensation, toolpath planning, and stiffness optimization have been explored (see Table I). Recent work has introduced hybrid temporal networks with iterative path compensation [3], movement primitive frameworks for low-distortion trajectory reproduction [4], and redundancy-based posture optimisation for smooth joint motions [5]. Systematic error modelling has further linked machining deviations to kinematic and stiffness factors [6], while vision-based mobile grinding [7] and joint optimisation of robot posture and workpiece setup [8] demonstrate effective error reduction in large-scale machining. Multi-robot collaboration has also been proposed, enabling synchronised large-scale machining with improved stiffness and flexibility [9]. Despite these advances, most approaches remain constrained by first-principle modelling (e.g., toolpath, workpiece geometry and system kinematics) and costly sensing. Additionally, high-precision manufacturing still relies on offline CMM inspection [10], while optical systems [11] and touch probes [12] face limitations with speed, geometry, or process interruption. State of the art solutions for on-line measurements involve accurate position sensing such as with laser trackers [13]. However, these remain costly and impractical for large scale industrial deployment. In this context, low-cost sensing and learning based models seem to be promising and have not been fully explored for machining error prediction.

Force and vibration signals provide direct, real-time insight into the dynamic interaction between a robotic tool and workpiece. They capture tool engagement, material removal, robot compliance, and cutting instabilities that directly affect dimensional accuracy and surface quality [14]. Force sensing can improve machining accuracy by over 70% [15] and suppress vibrations in low-stiffness workpieces [16], while vibration signals reveal transient disturbances and process conditions [17]. These signals are high-frequency, low-cost, and non-intrusive, making them ideal for predictive modelling of machining errors. However, force and vibration signals capture complementary yet heterogeneous aspects of the cutting process. Integrating them within a single model often struggles to perform robustly if they rely on fixed feature

This work was supported by UK Engineering and Physical Sciences Research Council EP/P006930/1. We acknowledge the data provision and discussions with Bin Chen, Simon Fletcher, Peace Afolabi and Erdem Ozturk.

[1] School of Electrical and Electronic Engineering, University of Sheffield, Sheffield S1 3JD, U.K. [2] School of Computer Science, University of Sheffield, Sheffield S1 4DP, U.K. Email: {c.niu, visakan}@sheffield.ac.uk.

*The corresponding author

TABLE I: Comparison of robotic machining approaches in terms of task and error modelling categories. CDMP: Combination of Dynamical Movement Primitives. MoE: Mixture of Experts. DoF: Degree of Freedom.

Literature	Approach	Task	Strengths	Limitations
Peng et al. (2024) [3]	Hybrid temporal network + iterative path compensation.	Path error compensation.	Improves contour accuracy; iterative refinement.	Data-intensive; limited to path errors.
Zhou et al. (2022) [4]	CDMP with trajectory segmentation for machining motion imitation.	Motion imitation / trajectory shaping.	Low-distortion, efficient trajectories.	Geometry-driven; insensitive to cutting dynamics.
Lu et al. (2021) [5]	Redundancy-based posture optimization for smooth joint motion.	Kinematic trajectory optimization.	Exploits DoFs; smooth, efficient motions.	Kinematic-only; ignores process dynamics.
Xie et al. (2020) [6]	Error modeling with kinematics / stiffness; compensation via setup / tool pose.	Model-driven error compensation.	Explicit error–factor modeling; reduced error.	Relies on accurate models; poor adaptability.
Zhao et al. (2022) [7]	Vision-based grinding with surface reconstruction.	Vision-guided adaptive compensation.	Effective for freeform surfaces.	Sensor cost; issues with reflectivity.
Liao et al. (2021) [8]	Joint optimization of robot posture and workpiece setup.	Stiffness enhancement.	Improved structural rigidity; fewer posture changes.	Focused on stiffness; neglects dynamic errors.
Zhao et al. (2021) [9]	Multi-robot cluster machining via distributed coordination.	Collaborative error mitigation.	Scalable, synchronized for large parts.	Complex coordination; not for high precision.
Proposed method	MoE for in-process predictive machining error modelling.	Data-driven error prediction.	Low-cost, real-time, scalable, in-process sensing.	Dependent on the part quality measured.

sets. The mixture-of-experts (MoE) architecture [18] shows potential to address this limitation by assigning specialised experts to different signal modalities and adaptively combining their outputs based on the most informative features, enabling robust and feature-aware decision making.

Machine learning has shown great potential for process monitoring and quality prediction in manufacturing [19]. Wavelet-enhanced CNNs capture multiscale features from vibration data for accurate fault diagnosis [20]. Deep learning models, including stacked autoencoders and feedforward neural networks, predict manufacturing quality effectively [21], while CNN–LSTM architectures handle temporal dependencies to forecast surface roughness [22]. In robotic machining, Gaussian process regression and in-process compensation frameworks have shown promise for linking sensory signals with post-process errors [23], [24], though mostly in drilling applications. In conventional CNC machining, force, vibration, and controller signals are widely used for process monitoring [25], [26]. This motivates investigating their feasibility in robotic machining. High-precision robotic milling is more challenging due to the nonlinear relationship between process signals and geometric errors. Offline inspection with costly coordinate measuring machines ensures accuracy, but cannot provide real-time feedback. Leveraging force and vibration signals with data-driven modelling is promising to predict machining errors online accurately and cost-effectively.

This paper proposes a framework for in-process machining error prediction. It employs a mixture-of-experts (MoE) architecture that integrates external force and vibration signals with internal robot controller data, where the controller dynamically weights expert predictions. We investigate both hand-crafted features, including temporal and spectral descriptors, and self-learned representations using variational autoencoders (VAEs), multimodal VAEs, and CNN bottleneck architectures with squeeze-and-excitation modules. Additionally, we systematically benchmark the proposed framework against laser tracker based method. The main contributions of this work are:

- A MoE framework that fuses force, vibration, and controller signals for machining error prediction, validated on diverse workpieces with square and circular features.
- A comparative study of feature representations, demonstrating that designed handcrafted features achieve superior prediction accuracy to self-learned features from the in-process signals.
- An evaluation of sensor requirements, showing that the proposed framework can be comparable to laser tracker learning based method, offering a low cost alternative to expensive metrology and position sensing systems.
- A computational efficiency analysis, highlighting that the proposed framework enables real-time, energy-efficient deployment suitable for on-device monitoring.

The remainder of this paper is organised as follows: Section II presents problem formulation, data collection and preprocessing, and the proposed framework. Section III describes implementations of experiments. Models are evaluated in terms of prediction accuracy and computational efficiency, as described in Section IV. Section V analyses the findings, limitations, and future work. Section VI highlights the key insights and contributions.

II. Materials and Methods

A. Problem formulation

In high-precision robotic milling, the objective is to predict the instantaneous machining error $e(t) \in \mathbb{R}$ as a function of available in-process signals. Let the general set of signals be

$$\mathcal{X}(t) = \{\mathbf{F}(t), \mathbf{V}(t), \mathbf{C}(t), \mathbf{L}(t)\}, \tag{1}$$

where $\mathbf{F}(t)$ denotes force signals, $\mathbf{V}(t)$ denotes vibration signals, $\mathbf{C}(t)$ denotes internal controller commands, and $\mathbf{L}(t)$ denotes position measurements from a laser tracker. A predictive model $\hat{e}(t)$ can be expressed as:

$$\hat{e}(t) = f_\theta(\mathbf{X}_S(t)), \quad \mathbf{X}_S(t) \subseteq \mathcal{X}(t), \tag{2}$$

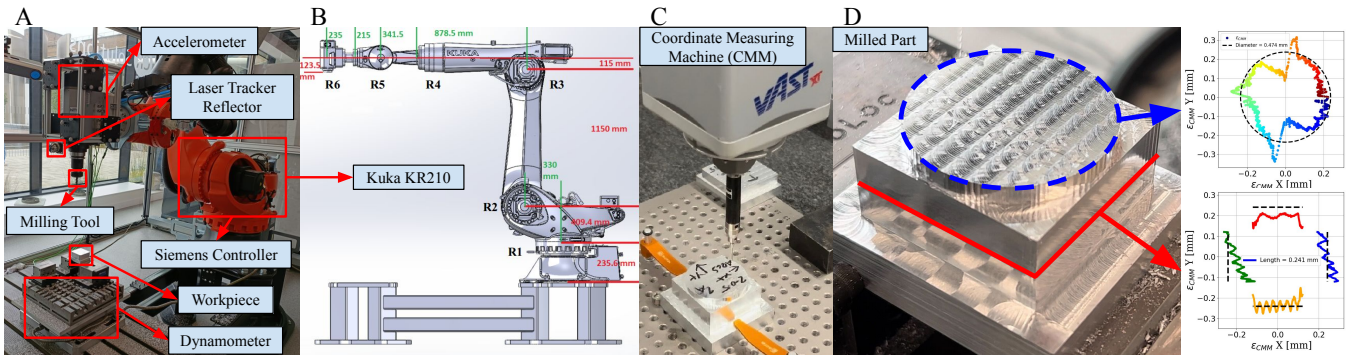


Fig. 1: Robot platform and quality measurement setup. **A** Sensor and workpiece configuration. **B** KUKA KR210 robot with six revolute joints. **C** CMM (Zeiss Prismo Access) with 1 mm probe for part quality measurement. **D** Milled part and zoomed-in view of geometries of square (scaling factor of 200) and circle (scaling factor of 100) measured by CMM, black dashed lines indicate nominal size. Temporal milling sequence for square: right (blue), front (yellow), left (green), rear (red) and for circle: clockwise, right (blue) to rear (red).

where $\mathbf{X}_S(t)$ is a selected subset of available signals, f_θ belongs to a class of candidate machine learning methods, and θ are the learnable parameters. The model may incorporate explicit *feature mappings* $\phi : \mathbf{X}_S \mapsto \mathbb{R}^d$ to transform raw signals into a feature space, such that:

$$\hat{e}(t) = f_\theta \circ \phi(\mathbf{X}_S(t)), \quad \phi \in \{\phi_{\text{features}}\}. \tag{3}$$

The objective is to find an optimal combination of signal subset, feature representation, and model parameters that minimizes loss $\mathcal{L}$ measuring the deviation between predictions and ground truth over the entire machining process $t \in [0, T]$:

$$(\theta^*, \mathbf{X}_S^*, \phi^*) = \arg \min_{\theta \in \Theta,\, \mathbf{X}_S \subseteq \mathcal{X},\, \phi \in \{\phi_{\text{features}}\}} \frac{1}{T} \int_0^T \mathcal{L}\Big(e(t), f_\theta \circ \phi(\mathbf{X}_S(t))\Big)\, dt. \tag{4}$$

The problem is whether deviations in CMM-measured workpiece geometries can be accurately predicted using in-process sensing, leading to three main questions: (i) viability of using existing in-process sensing without requiring external, (ii) the necessity and effectiveness of employing MoE and (iii) selection of predictive features and position sensing.

Q1: Necessity of position sensing: How effective is low-cost in-process sensor (e.g., force, vibration, controller) based error predictions in comparison to direct measurement of position sensing (e.g., laser tracker) $\mathbf{L}(t)$?

$$|f_\theta^*(\mathbf{F}(t), \mathbf{V}(t), \mathbf{C}(t))| \gtrsim |f_\theta^*(\mathbf{L}(t), \mathbf{C}(t))| \tag{5}$$

Q2: Necessity and effectiveness of MoE: The MoE can be expressed as

$$f_{\text{MoE}}(t) = \sum_{k=1}^{K} w_k(t) f_{\theta_k}(\mathbf{X}_S(t)) \tag{6}$$

where $w_k(t) = w_k(\mathbf{C}(t))$ are adaptive gating weights for each expert, $\sum_{k=1}^{K} w_k(t) = 1, w_k(t) \geq 0$ and $\mathbf{C}(t)$ are the input to the weighting function. This question investigates whether the predictive performance of the MoE consistently exceeds that of any individual expert, i.e.,

$$P_{\text{MoE}} \geq \max_k P_{\text{expert},k}. \tag{7}$$

where P denotes a performance metric. It also examines whether the gating mechanism produces balanced weights $w_k(t)$ rather than dominance by a single expert (i.e., $w_k(t) \approx 1$ for one k and ≈ 0 for others).

Q3: Optimal feature representation: Given a candidate signal subset $\mathbf{X}_S$, which feature mapping ϕ^* best captures the relevant information for predicting machining error?

$$\phi^* = \arg \min_{\phi \in \{\phi_{\text{hand}}, \phi_{\text{learned}}\}} \frac{1}{T} \int_0^T \mathcal{L}\Big(e(t), f_\theta \circ \phi(\mathbf{X}_S(t))\Big)\, dt. \tag{8}$$

where ϕ_{hand} corresponds to hand-engineered features in terms of temporal and spectral perspective, and ϕ_{learned} corresponds to self-learned features obtained from supervised or self-supervised learning.

B. Robot platform setup and data collection

The experiments were conducted using a KUKA KR210 6-axis industrial robot (see Fig. 1A & B) equipped with a 10 mm diameter milling tool, controlled via a Siemens 840D solution line controller, with position data recorded through OPC UA at 10 Hz (see more details in [27]). Force signals were collected at 10 kHz using a Kistler dynamometer positioned underneath the vice, while vibration signals were collected at 1 kHz using an Analog Devices ADXL355 accelerometer mounted on the spindle. A Leica Absolute Tracker ATS600 laser tracker (100 Hz) provided high-frequency monitoring of the tool's real-time motion. The machining was performed with a spindle speed of 8500 rpm, a feed rate of 1275 mm/min, and a 3-flute milling tool with a diameter of 10 mm. The CMMs setup and quality measurement are illustrated in Fig. 1C & D. The machining sequence began with a square section milled from 0 mm to -25 mm along the Z-axis in 25 layers of 1 mm each, followed by circular features machined in two rounds with depths of -2 mm, -4 mm, and -4.8 mm, progressively reducing diameters. Two workpieces $\mathcal{W}_1, \mathcal{W}_2$, were machined. Dimensional measurements were obtained via CMM at the 7th and 14th layers ($l = 7, 14$) of the square section and the 1st and 2nd layers ($l = 1, 2$) of the circular section, with volumetric accuracy below 3 µm. Spatial distributions of machining errors for square and circular geometries are shown in Fig. 1D. In square milling, the four sides exhibit different

geometries, while circular milling shows location-dependent variations. Backlash, caused by reversal motion of the robot's revolute joints, occurs at three locations in circular milling: twice at $X = 0$ and once at $Y = 0, X < 0$.

C. Data configuration and preprocessing

Training and testing setup: The dataset, collected from two workpieces with both square and circular geometries, comprises 4,897 CMM/laser tracker samples, 490 controller samples, 48,970 vibration samples, and 489,700 force samples. We train a machine learning model to predict CMM deviations $e(t)$ as, $e_{\mathcal{W},l}(t) = f^*\big(\mathbf{X}_{\mathcal{W},l}(t)\big) + \epsilon_{\mathcal{W},l}(t)$, where $e_{\mathcal{W},l}(t)$ denotes the deviation of CMM measurements from nominal geometry for workpiece $\mathcal{W}$ at layer l, $\mathbf{X}_{\mathcal{W},l}(t)$ is the set of in-process signals at the same index, $\hat{e}_{\mathcal{W},l}(t)$ is the model prediction, and $\epsilon_{\mathcal{W},l}(t)$ is the modeling error. We then consider three generalization testing setups: (i) *cross-workpiece, same layer*—train on $(\mathcal{W}_1, l_1)$ and predict on $(\mathcal{W}_2, l_1)$: $\hat{e}_{\mathcal{W}_2,l_1}(t) = f_{\mathcal{W}_1,l_1}(\mathbf{X}_{\mathcal{W}_2,l_1}(t))$; (ii) *cross-workpiece, cross-layer*—train on $(\mathcal{W}_2, l_1)$ and predict $\hat{e}_{\mathcal{W}_1,l_2}(t) = f_{\mathcal{W}_2,l_1}(\mathbf{X}_{\mathcal{W}_1,l_2}(t))$ with $l_1 \neq l_2$; and (iii) *reversed modeling*—swap training/testing to obtain $\hat{e}_{\mathcal{W}_1,l_1}(t) = f_{\mathcal{W}_2,l_1}(\mathbf{X}_{\mathcal{W}_1,l_1}(t))$.

Preprocessing: The CMM error $e(t) \in \mathbb{R}$, sampled at 100 Hz, serves as the ground truth. *Pre-defined* Siemens controller trajectories at 10 Hz are upsampled to 100 Hz via linear interpolation [28], giving $\mathbf{C}(t) \in \mathbb{R}^2$ for X–Y axes. Force signals $\mathbf{F}(t) \in \mathbb{R}^3$ at 10 kHz are segmented into $\mathbf{F}_w(t) \in \mathbb{R}^{128\times3}$ using 128-sample windows (12.8 ms) with 100-sample hops (10 ms), while vibration signals $\mathbf{V}(t) \in \mathbb{R}^3$ at 1 kHz are segmented into $\mathbf{V}_w(t) \in \mathbb{R}^{13\times3}$ with 13-sample windows (13 ms) and 10-sample hops (10 ms). The aligned input at each time step is $\mathbf{X}(t) = \{\mathbf{F}_w(t), \mathbf{V}_w(t), \mathbf{C}(t)\}$, with prediction target $e(t)$.

D. Mixture-of-Experts framework

We model the instantaneous machining error $e_{\mathcal{W},l}(t)$ of workpiece $\mathcal{W}$ at layer l using a Mixture-of-Experts (MoE) architecture, where multiple expert models are dynamically fused through a router network conditioned on internal controller signals (see Fig. 2A). The MoE prediction is formulated as a convex combination of expert outputs: $\hat{e}_{\mathcal{W},l}(t) = \sum_{k=1}^{3} w_k^{\mathcal{W},l}(t)\, f_{\mathcal{W},l}^{(k)}\big(\phi_k(\mathbf{X}_{\mathcal{W},l}^{(k)}(t)); \theta_k\big)$, where: $f_{\mathcal{W},l}^{(k)}(\cdot;\theta_k)$ is the k-th expert mapping parameterized by θ_k, $\phi_k : \mathbf{X}_{\mathcal{W},l}^{(k)} \mapsto \mathbb{R}^{d_k}$ is a feature mapping (handcrafted or self-learned) for expert k, $\mathbf{X}_{\mathcal{W},l}^{(k)}(t)$ denotes the input subset for expert k: $\mathbf{X}_{\mathcal{W},l}^{(1)} = \{\mathbf{F}_w, \mathbf{C}\}$, $\mathbf{X}_{\mathcal{W},l}^{(2)} = \{\mathbf{V}_w, \mathbf{C}\}$, $\mathbf{X}_{\mathcal{W},l}^{(3)} = \{\mathbf{F}_w, \mathbf{V}_w, \mathbf{C}\}$, $w_k^{\mathcal{W},l}(t) = g_\psi^{(k)}(\mathbf{C}(t))$ is the router-assigned weight for expert k, with weights $\mathbf{w}^{\mathcal{W},l}(t) = [w_1^{\mathcal{W},l}(t), w_2^{\mathcal{W},l}(t), w_3^{\mathcal{W},l}(t)]^\top \in \Delta^3$, the 3-dimensional probability simplex. The router mapping $g_\psi^{(k)} : \mathbb{R}^2 \mapsto \Delta^3$ is a learnable function parameterized by ψ that dynamically assigns contributions of each expert based on the robot controller state. Feature mappings ϕ_k transform raw segments into latent representations, unifying handcrafted and self-learned features in the MoE framework (see Fig. 2A). The joint training objective over the workpiece $\mathcal{W} \in \mathcal{S}$ for the specific layers $l \in \mathcal{L}_\mathcal{W}$ is defined as:

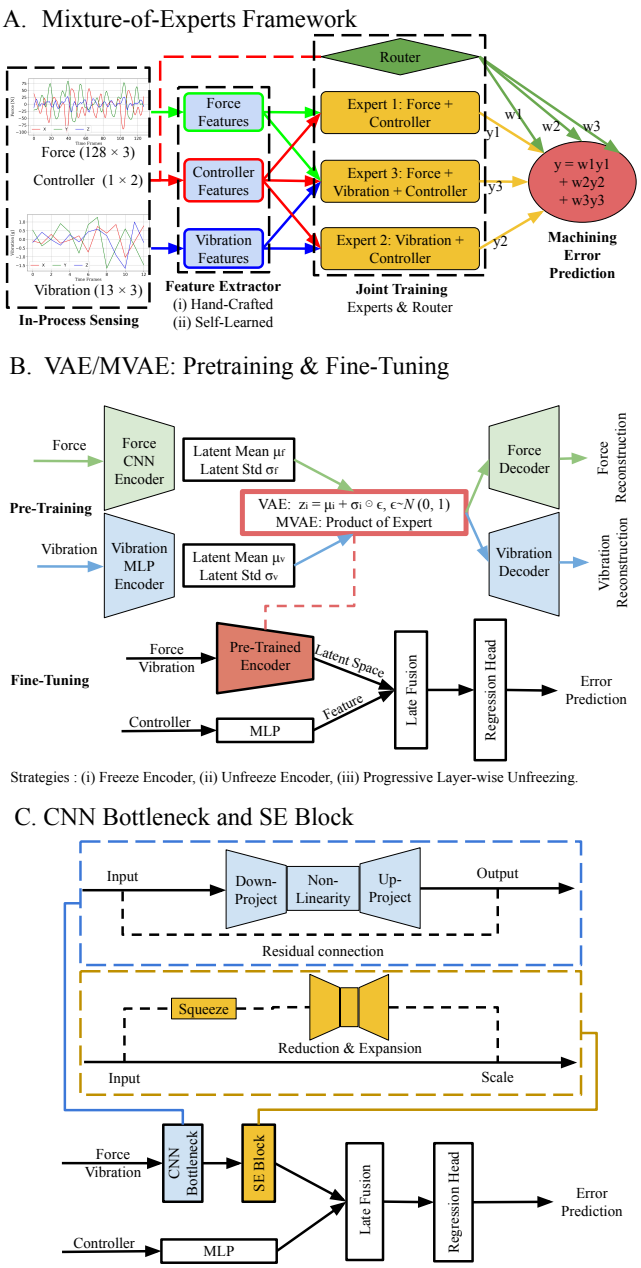


Fig. 2: Overview of MoE Framework and self learned feature representation. MLP: Multi-Layer Perceptron. CNN: Convolutional Neural Network.

$$(\theta_1^*, \theta_2^*, \theta_3^*, \psi^*) = \arg\min_{\theta_k,\psi} \left(\frac{1}{T_l}\right) \int_0^{T_l} \mathcal{L}\Big(e_{\mathcal{W},l}(t), \hat{e}_{\mathcal{W},l}(t)\Big)\, dt \tag{9}$$

where T_l is the duration of layer l and $\mathcal{L}(\cdot,\cdot)$ is a loss function. The experts $f^{(k)}(\cdot)$ and router $g_\psi^k(\cdot)$ are jointly optimized, with the router dynamically weighting contributions. The MoE router uses only the controller signal to capture configuration-dependent stiffness and kinematic variations, enabling expert specialization across robot operation, while the experts rely on external measurements (i.e., force and vibration) to model dynamic disturbances and predict machining errors.

Handcrafted Features: For each multichannel time-series segment, we extracted a set of temporal and spectral features to

provide a compact representation of the signals. The temporal features included the mean, standard deviation, root mean square per channel, and overall root mean square across all channels, capturing the central tendency, variability, and overall signal magnitude. In the spectral domain, we computed the dominant frequency, spectral centroid, spectral bandwidth, and spectral energy per channel, which characterize the frequency content, spread, and overall power of the signals. Additionally, we construct two feature combinations: the temporal combination, which includes mean, standard deviation, and root mean square per channel, and the frequency combination, which consists of dominant frequency, spectral centroid, spectral bandwidth, and spectral energy per channel.

Self-learned features: We employed two strategies to capture rich representations of signals. The first uses Variational Autoencoders (VAEs) [29] and a Multimodal VAEs (MVAEs) to learn compact latent embeddings via unsupervised signal reconstruction, with expert-specific VAEs capturing modality-specific patterns and the MVAE jointly encoding both modalities to capture cross-modal correlations (see Fig. 2B). The second strategy uses a CNN-based architecture with bottleneck [30] and Squeeze-and-Excitation (SE) modules [31]. The bottleneck layers reduce the temporal sequences into compact representations, whereas the SE modules adaptively recalibrate the three sensing channels to model inter-channel dependencies and highlight the most informative components (see Fig. 2C). The VAEs captures global latent representations, whereas the CNN-SE extracts fine-grained features.

III. Experiments

A. Handcrafted features

Preliminary experiments comparing random forest, decision tree, gradient boosting, and K-nearest neighbours demonstrated that random forest consistently performed best, and was chosen as the expert model for subsequent experiments. Three random forest experts were chosen to be architecturally different in terms of the input they receive, namely, force, vibration and their combination, in addition to the controller features. A neural router receiving controller data as input with two hidden layers (32 and 16 ReLU neurons) and a 3-neuron softmax output was trained to produce weights for dynamically combining the three expert predictions. These low cost sensor based mixture of experts predictions are compared with high cost direct laser measurements and with a machine learning model utilising the laser measurements as developed in [27].

B. Self learned features

1) Variational Autoencoders

Signal reconstruction: We implemented single-modal VAEs and a multimodal VAE (MVAE) to learn compact latent embeddings from force and vibration signals via unsupervised reconstruction (see Fig. 2B). Force signals were encoded with a 1D convolutional network and vibration signals with a fully connected MLP, each mapping to latent spaces of 1, 2, 4, 8, 16, 32, 64 and 128 dimensions, respectively. The architectural difference is due to the force signal being sampled at a much higher rate than the vibration signal. The decoders used upsampling and one-dimensional convolutions for reconstruction. The MVAE combined modality-specific encodings via a product-of-experts (PoE) [32], and reconstruction losses using mean squared error (MSE) for each modality were combined with Kullback–Leibler (KL) divergence [33] to form the total loss. The reconstruction loss is defined as MSE between the vectorized input signal x_i and its reconstruction $\hat{x}_i$, i.e., $\mathcal{L}_{\text{recon}} = \frac{1}{N}\sum_{i=1}^{N} \|x_i - \hat{x}_i\|_2^2$. The KL divergence regularizes the latent distribution parameterized by mean μ and standard deviation σ toward a standard Gaussian prior, given by $\mathcal{L}_{\text{KL}} = -\frac{1}{2}\sum_{j=1}^{d}\left(1 + \log\sigma_j^2 - \mu_j^2 - \sigma_j^2\right)$, where d is the latent dimensionality. In the multimodal VAE, the latent distribution is obtained via a PoE. The mean μ and variance σ^2 of the combined latent variable are computed as follows: $\mu = (\sum_{m=1}^{M} \mu_m/\sigma_m^2)/(\sum_{m=1}^{M} 1/\sigma_m^2)$ and $\sigma^2 = 1/(\sum_{m=1}^{M} 1/\sigma_m^2)$, with M denoting the number of modalities, here M=2, indicating force and vibration respectively. The overall loss function is expressed as $\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{recon}} + \beta\mathcal{L}_{\text{KL}}$, the MVAE extends $\mathcal{L}_{\text{recon}}$ to include modality-specific force and vibration reconstruction losses, i.e., $\mathcal{L}_{\text{MVAE}} = \mathcal{L}_{\text{recon}}^{F} + \mathcal{L}_{\text{recon}}^{V} + \beta\mathcal{L}_{\text{KL}}$, where $\beta = 1$ is used for both VAE and MVAE to ensure a consistent weighting between reconstruction fidelity and latent regularization, enabling a fair comparison across architectures without additional hyperparameter tuning bias. MSE ensures accurate reconstruction, while KL divergence regularizes the latent space for smooth and generalizable representations. All models were trained on an NVIDIA GeForce RTX 3060 GPU (12 GB) with Adam optimizer (1×10^{-3}), batch size of 64, for up to 500 epochs with early stopping (20 epochs) and checkpointing, and reconstruction quality was evaluated via MAE and RMSE.

Downstream task: The pretrained VAE/MVAE encoders were incorporated into a downstream regression framework for CMM error prediction and fine-tuned using three strategies: (i) freezing all encoder layers, (ii) unfreezing all layers, and (iii) progressive layer-wise unfreezing across three stages with decreasing learning rates, starting with all layers frozen, then unfreezing the last three layers (i.e., the latent mean μ, latent standard deviation σ, and latent space $z = \mu + \sigma \odot \epsilon$, $\epsilon \sim \mathcal{N}(0, 1)$ [34]) and finally unfreezing all layers. The progressive layer-wise unfreezing strategy was introduced for two reasons. First, it enables a systematic comparison of different transfer learning strategies for pretrained VAE encoders under the same experimental setting. Second, compared with directly unfreezing all layers, it allows gradual adaptation of pretrained representations to the downstream task, thereby reducing the risk of catastrophic forgetting and training instability. This strategy has been widely adopted in studies on edge devices and mobile computing [35], [36] due to its ability to improve optimization stability while preserving previously learned feature representations. The pretrained encoders are fine-tuned for machining error prediction using a router-based framework. The router generates adaptive weights w_1, w_2, and w_3 from input controller signals $\mathbf{C}_t$, satisfying $w_1 + w_2 + w_3 = 1$, and the final prediction is computed as $\hat{y}_i = \sum_{k=1}^{3} w_k(\mathbf{C}_t) y_{k,i}$, where $y_{k,i}$ denotes the output of the k-th expert. The downstream regression is optimized using

TABLE II: Performance metrics (MAE and RMSE in mm) for MoE framework, computed between the ground truth and model predictions. L_{direct}: Direct Laser Tracker Measurement; ML: Machine Learning Prediction using L_{direct}; TCF: Temporal Combined Features; TM: Temporal Mean; TSD: Temporal Standard Deviation; TRMS: Temporal Root Mean Square; TTRMS: Total Temporal Root Mean Square; FCF: Frequency Combined Features; SB: Spectral Bandwidth; SE: Spectral Energy; SC: Spectral Centroid; DF: Dominant Frequency; CNNB: CNN Bottleneck & SE; VAEF: VAE (Frozen); VAELU: VAE (Layer-wise Unfrozen); VAEUF: VAE (Unfrozen). L_{direct} as baseline for comparison with learning based model. Bold numbers indicate the relatively best models discussed in the text.

	Scenario 1		Scenario 2		Scenario 3	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Position Sensing using Laser Tracker						
L_{direct}	0.0756	0.0943	0.0715	0.0930	0.0714	0.0896
ML	0.0097	0.0146	0.0139	0.0202	0.0089	0.0141
In-process Low-cost Sensing of Force and Vibration						
Temporal Features						
TCF	0.0106	0.0171	**0.0145**	**0.0199**	0.0106	0.0162
TM	**0.0098**	**0.0163**	0.0146	0.0207	**0.0093**	**0.0151**
TSD	0.0102	0.0167	0.0148	0.0204	0.0102	0.0159
TRMS	0.0103	0.0169	0.0150	0.0206	0.0101	0.0156
TTRMS	0.0099	0.0167	0.0154	0.0213	0.0098	0.0159
Frequency Features						
FCF	0.0125	0.0195	**0.0146**	**0.0200**	0.0104	0.0159
SB	0.0116	0.0185	0.0148	0.0208	**0.0092**	**0.0151**
SE	0.0104	0.0168	0.0152	0.0209	0.0102	0.0160
SC	0.0122	0.0185	0.0149	0.0209	0.0095	0.0152
DF	**0.0100**	**0.0172**	0.0154	0.0217	0.0095	0.0161
Deep-learned Features						
CNNB	0.0149	0.0203	**0.0167**	**0.0222**	0.0145	0.0205
Variational Autoencoder - 64 dimensional latent space						
VAEF	0.0217	0.0290	0.0217	0.0289	0.0180	0.0240
VAELU	0.0157	0.0210	0.0187	0.0247	0.0160	0.0213
VAEUF	0.0145	0.0211	**0.0171**	**0.0227**	0.0148	**0.0203**
Variational Autoencoder - 16 dimensional latent space						
VAEF	0.0216	0.0292	0.0198	0.0266	0.0180	0.0246
VAELU	0.0183	0.0250	0.0217	0.0292	0.0179	0.0245
VAEUF	**0.0143**	**0.0199**	0.0172	0.0237	**0.0147**	0.0215

the MSE loss $\mathcal{L}_{\text{reg}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$, where y_i and $\hat{y}_i$ are the ground-truth and predicted machining errors respectively. Models were trained with the Adam optimizer, using learning rates of 1×10^{-3} for full training and 1×10^{-3}, 1×10^{-4}, and 1×10^{-5} for the three progressive stages, with batch size 32, up to 50 epochs, early stopping at 20 epochs.

2) CNN with bottleneck and SE

Each expert network processes force and/or vibration sequences with bottlenecks and squeeze and excitation (SE) blocks, temporal pooling, and global average pooling, while a parallel fully connected pathway handles 2D controller input (see Fig. 2C). Expert outputs are fused by a softmax router resulting in the mixture of experts architecture. Training uses Adam (1×10^{-3}), batch size 32, up to 50 epochs with early stopping at 20, and MSE loss.

IV. Results

A. Dimensional error prediction performance

Model performance was evaluated using mean absolute error (MAE) and root mean squared error (RMSE) computed between the predicted workpiece size deviations and the ground-truth deviations measured by the CMM. Preliminary investigation to determine the latent dimension within a set of $1, 2, 4, \cdots, 128$ (see Section III-B1) revealed that 16 and 64-dimensional VAE latent vectors with encoder freezing achieved the lowest error. These were selected for subsequent experiments with full and progressive unfreezing strategies.

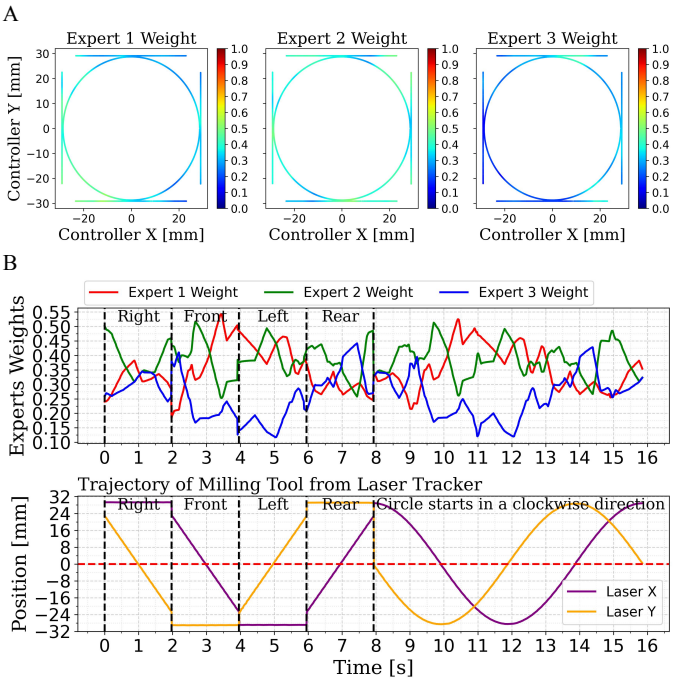


Fig. 3: Visualisation of individual experts' weights via space (A) and temporal (B) domain using temporal combined features (TCF), alongside the real trajectory of the milling tool recorded by laser tracker. See more details for the temporal order in Fig. 1D.

TABLE III: Performance metrics (MAE and RMSE in mm) for Experts 1, 2, 3, and MoE using temporal combined features (TCF).

	Scenario 1		Scenario 2		Scenario 3	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Expert 1	0.0108	0.0175	**0.0143**	**0.0197**	**0.0105**	0.0167
Expert 2	0.0116	0.0185	0.0157	0.0216	0.0111	0.0168
Expert 3	0.0112	0.0178	0.0145	0.0197	0.0113	0.0174
MoE	**0.0106**	**0.0171**	0.0145	0.0199	0.0106	**0.0162**

We provide a baseline for comparison between direct laser tracker measurement (L_{direct}) and the data-driven predictions. L_{direct} represents the tool trajectory deviation between the measured laser tracker positions and the predefined controller trajectory. These deviations are projected onto the workpiece surface, aligning the number of uniformly laser tracker samples (constant cutting speed) for cutting one cycle with the corresponding CMM measurement points. The MAE and RMSE between $L_{\text{direct}}(t)$ and $e(t)$ are reported in Table II.

All the mixture of experts models consistently achieved comparably good performance, matching costly laser tracker learning-based method (see Table II). They significantly surpass direct laser tracker measurements, achieving around five-fold to sevenfold improvement in MAE and RMSE. These findings address the first research question (see Section II-A) and demonstrate that accurate predictions can be achieved without relying on laser tracker position sensing.

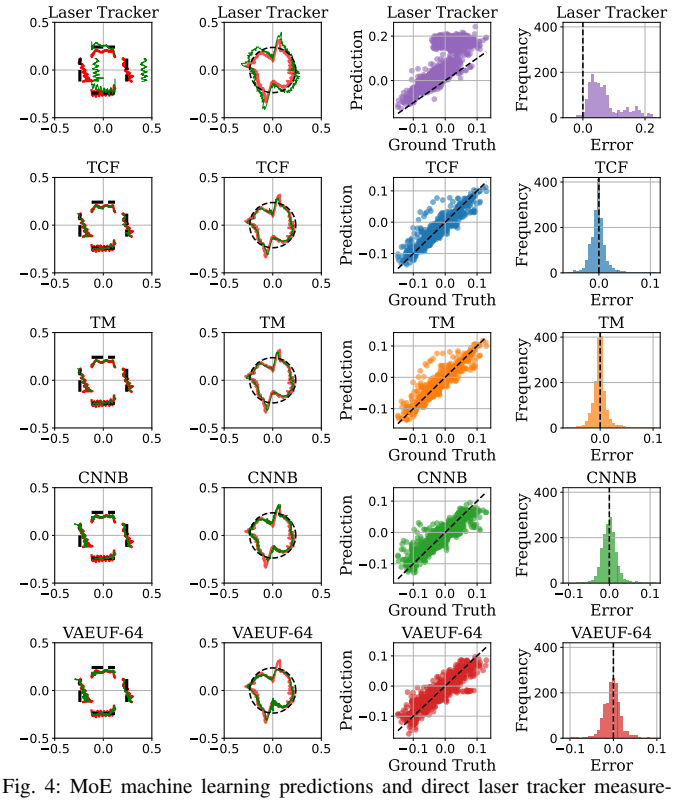


Fig. 4: MoE machine learning predictions and direct laser tracker measurements (green) are compared to ground truth (bold red) for square and circle geometries in the first and second columns, with nominal sizes as black dashed lines. Scatter plots show their correlations, and error histograms indicate discrepancies in the third and fourth columns.

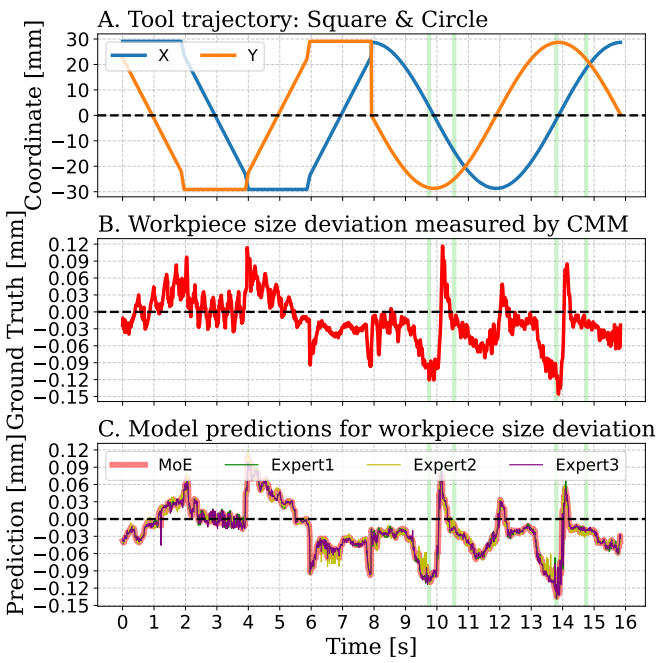


Fig. 5: Visualisation of tool trajectory, workpiece size deviation, and model predictions. A. Controller-based tool trajectories (square and circular). B. Workpiece size deviation (difference between CMM measurements and nominal), with spatial distribution shown in Fig. 1D. C. Predicted size deviation from data-driven models (Experts and MoE), with backlash and non-backlash regions green-highlighted. Time intervals and metrics are given in Table IV.

We then examine the second research question using temporal combined features (Mean, Std, RMS). Expert weights for three experts across spatial and temporal domains (Fig. 3) show adaptive contributions from all three experts. A similar trend is observed for other feature sets, confirming that no single expert dominates the MoE. Additionally, individual experts may achieve strong performance in one scenario but fails in another, whereas the MoE dynamically assigns weights to experts, ensuring more consistent and near optimal performance across different scenarios (see results in Table III and scenario setting in Section II-C). Results from employing a mixture of experts also demonstrate cross-workpiece generalisation under three scenarios in Table II. Scenarios 1 and 3 exhibit similar MAE and RMSE, whereas Scenario 2 shows higher errors but still acceptably small, reflecting that training with data from a layer and predicting across workpieces at the same layer yields lower errors than across different layers.

Single temporal or frequency features achieved better performance than their combination in the same-layer setting of two workpieces (Scenarios 1 & 3), whereas feature combination outperforms single features in the cross-layer setting (Scenario 2). A similar trend is observed with the pre-trained encoder using VAE, where fully unfreezing the encoder is the best fine-tuning strategy for downstream error prediction. In cross-layer scenarios (Scenario 2), a 64-dimensional latent space outperforms a 16-dimensional one, while lower-dimensional feature works better in same-layer scenarios (Scenarios 1 & 3). These findings suggest that when data distributions are similar and stable (e.g., same-layer for training and prediction), adding more features may introduce noise, whereas in cross-layer settings with distinct data distributions, richer features enhance model generalisation. Besides, the CNN bottleneck with squeeze and excitation in the cross-layer setup (Scenario 2) outperforms the VAE, demonstrating the effectiveness of the bottleneck and SE architecture in enhancing generalisation. Finally, we observe that handcrafted features outperform self-learned features, thereby addressing the third research question.

We selected several MoE models for visualisation in the spatial domain to compare their predictions with the ground truth (see Fig. 4) in Scenario 3. The temporal mean (TM) feature is selected for its high accuracy, while the temporal combined features (TCF), which are three times higher in dimensionality than TM, are used to assess the effect of feature dimensionality in handcrafted features. For self-learned features, we include CNN bottlenecks (CNNB) and VAE embeddings, with the VAEUF model employing a 64-dimensional latent space. These are visually compared with L_{direct}. The maximum error is reduced from 0.25 mm with L_{direct} to 0.05 mm using the MoE framework.

B. Robustness analysis under transient machining conditions

In Scenario 2 (see Table III), although Expert 1 achieves the best overall MAE and RMSE, the performance difference compared to the MoE is marginal, motivating further analysis under critical transient conditions. On two backlashes at X=0 in circular milling, which correspond to the largest workpiece

TABLE IV: Time intervals and performance metrics. MAE and RMSE are computed between ground truth (Fig. 5B) and model predictions (Fig. 5C) during each time interval. Bold text shows the best performance.

Backlash Period			Non-Backlash Period		
Model	MAE	RMSE	Model	MAE	RMSE
Duration: 9.70-9.80s			Duration: 14.70-14.80s		
MoE	**0.0070**	**0.0085**	MoE	**0.0030**	**0.0038**
Expert 1	0.0100	0.0110	Expert 1	0.0033	0.0041
Expert 2	0.0095	0.0110	Expert 2	0.0041	0.0061
Expert 3	0.0098	0.0108	Expert 3	0.0033	0.0042
Duration: 13.75-13.85s			Duration: 10.50-10.60s		
MoE	**0.0164**	**0.0182**	MoE	0.0079	0.0085
Expert 1	0.0208	0.0221	Expert 1	0.0088	0.0094
Expert 2	0.0185	0.0210	Expert 2	0.0087	0.0102
Expert 3	0.0189	0.0205	Expert 3	**0.0057**	**0.0063**

size deviations (up to -0.1457 mm at 13.75–13.85 s and -0.1205 mm at 9.70–9.80 s, see Fig. 5B), the MoE consistently achieves the lowest MAE and RMSE within fixed 0.1 s windows, as reported in Table IV, demonstrating superior performance in worst-case scenarios. In contrast, during non-backlash periods with small deviations (around -0.03 mm, see Fig. 5B), MoE outperforms all individual experts in a certain period, while in others, a single expert achieves the best performance. This indicates that single models may suffice under less challenging conditions. Furthermore, individual experts exhibit noticeable noise and outliers, whereas the MoE produces smoother and more stable predictions (Fig. 5C). These results highlight that, while individual experts may be competitive in some cases, the MoE provides clear advantages in reducing maximum errors and improving robustness under backlash transient machining conditions.

TABLE V: Routing-weight transitions near the workpiece centre lines. E1, E2 and E3 indicate three experts. See visualisation in Fig. 3.

Milling stage	Time [s]	Centre line	Comparable experts	Dominance transition
Square Right	(0.5, 1.5)	$t \approx 1$, $Y \approx 0$	E1, E2, E3	E1 $\rightarrow$ E2
Square Front	(2.5, 3.5)	$t \approx 3$, $X \approx 0$	E2, E1	E2 $\rightarrow$ E1
Square Left	(4.5, 5.5)	$t \approx 5$, $Y \approx 0$	E1, E2	E1 $\rightarrow$ E2
Square Rear	(6.5, 7.5)	$t \approx 7$, $X \approx 0$	E2, E3	E2 $\rightarrow$ E3
Circle	(9.5, 10.5)	$t \approx 10$, $X \approx 0$	E2, E1	E2 $\rightarrow$ E1
Circle	(11.5, 12.5)	$t \approx 12$, $Y \approx 0$	E2, E1	E2 $\rightarrow$ E1
Circle	(13.5, 14.5)	$t \approx 14$, $X \approx 0$	E3, E2	E3 $\rightarrow$ E2

C. Analysis of routing weights dynamics

The routing weights are adaptive and reflect each expert's suitability to the current machining state rather than fixed or uniform assignments. Across the entire machining process, the average weights for Experts 1, 2, and 3 are 0.35, 0.39, and 0.26, indicating that the mixture of experts does not collapse to a single expert, while local dominance varies with the machining phase. As shown in Fig. 3, the spatial and temporal evolution of routing weights along with the laser-tracker tool trajectory indicates that expert allocation changes consistently with process conditions. Table V further shows that routing transitions frequently occur near workpiece center lines ($X \approx 0$ or $Y \approx 0$), e.g., at $t \approx 1$ (E1$\rightarrow$E2) and $t \approx 3$ (E2$\rightarrow$E1), with similar patterns observed in both square and circular milling. This suggests that the routing network learns phase-dependent expert specialization, where a single expert dominates under stable local conditions, whereas transition regions lead to comparable expert contributions and dynamic switching. Overall, the routing mechanism achieves globally balanced yet locally specialized expert allocation that is physically consistent with tool position, robot configuration, and machining phase.

TABLE VI: Industrial deployment was evaluated for MoE with Temporal Combined Features (TCF), Temporal Mean (TM), CNN Bottleneck (CNNB), and VAE unfreezing encoder with 64-dimensional latent space (VAEUF-64). Metrics include inference time (median, 95th, 99th percentiles), throughput, jitter (standard deviation and P95–P50), energy per inference, energy efficiency, and energy delay product (EDP), aggregated from 1000 samples.

Metrics	TCF	TM	CNNB	VAEUF-64
Inference time (Computational latency)				
P50 (ms)	4.33	4.30	47.58	15.01
P95 (ms)	4.42	4.46	49.66	15.25
P99 (ms)	4.75	4.60	52.77	15.41
Thru (samples/s)	228	232	20	65
Jitter SD (ms)	0.12	0.11	1.27	0.17
Jitter P95-P50 (ms)	0.09	0.16	2.09	0.24
Energy consumption				
Energy/Inf (mJ)	179	176	2048	616
Eff (samples/W)	5.58	5.66	0.49	1.59
EDP (mJ·ms)	773	758	97432	9243

D. Computational efficiency analysis

In addition to prediction performance, computational efficiency determines how well the model integrates into robotic machining systems. Feature choice does not only depend on model accuracy, but also on whether they fit the real-time and energy demands of the system. Real time performance was evaluated using inference time (median/P50, P95, P99), throughput (samples per second), and jitter (inference time standard deviation and P95–P50 difference). Energy metrics included energy per inference (mJ), energy efficiency (throughput per watt), and the energy–delay product (energy $\times$ P50 inference). Tests were conducted on an AMD Ryzen 7 7700X CPU to reflect deployment feasibility without relying on GPUs. Handcrafted features (TCF, TM) achieve sub-5 ms latency per 13 ms segment, $>$220 samples/s throughput, and $<$180 mJ/inference energy (see Table VI), providing sufficient margin for real-time monitoring. The system uses a 13 ms sliding window with a 10 ms hop, with predictions updated every 10 ms (see Fig. 6). In contrast, CNN and VAE features have higher latency and energy costs, making them better suited for offline analysis or less time-critical tasks.

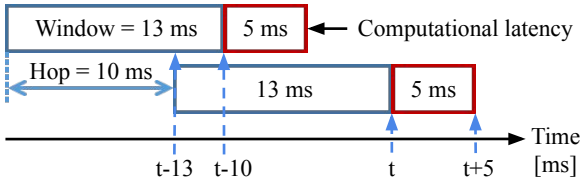


Fig. 6: System timing diagram. A 13 ms sliding window with a 10 ms hop is used, with < 5 ms inference latency. t denotes the current time corresponding to the end of each window, at which the machining error is predicted.

V. Discussion

This work proposes a mixture of experts framework that fuses low-cost in-process signals of force, vibration, and controller data to predict high-precision machining errors, achieving up to sevenfold improvement over direct laser tracker measurements. Handcrafted features within the MoE enable near metrology level accuracy, comparable to a learning-based method with a laser tracker, while supporting real-time and energy-efficient deployment. Since any shape can be decomposed into straight lines and curves, validating on a square and a circle provides insights into the model's ability to capture both linear and curved geometric features, which generalises to more complex shapes. Moreover, due to the limited machining trials that can be run, we evaluated three scenarios to demonstrate that knowledge learned from one workpiece can be transferred to others. The proposed method accurately predicts diverse geometric patterns along all edges of squares and circles (Fig. 1D) within one model, demonstrating its capability to handle varied geometric features without requiring multiple models.

The predicted machining error enables online quality monitoring and timely interventions, such as parameter adjustment or error compensation, in line with the concept of *Inspection by Exception* [23], [37], rather than relying solely on post-process CMM inspection. Analysis shows that cutting force partially correlates with machining error, while vibration signals are less correlated. This may be due to the complex influence of factors such as thermal changes and robot compliance. The proposed framework captures these interactions to provide accurate predictions and practical guidance for robotic machining, with further exploration of signal–error relationships and process parameter integration remaining future directions.

Comparing individual experts with MoE across feature types, MoE consistently outperformed single experts for self-learned representations (VAE, CNN), showing the benefit of modality fusion in latent spaces. With handcrafted features, individual experts matched or surpassed MoE, suggesting that while MoE is robust under feature uncertainty, simpler experts may suffice when effective signal–feature pairs are known.

In this study, although alternative fusion strategies, such as attention-based fusion [38] and stacking-based ensemble methods [39], are possible, we adopt the mixture of experts based fusion scheme because it provides an interpretable and computationally efficient solution for real-time compensation. Adaptivity is achieved through router-based time-varying expert weights and nonlinear expert networks specialised to different trajectory regions.

While the proposed framework shows strong performance, its applicability may be limited by the range of sensing–feature combinations and workpiece configurations. Although MoE with well-designed features performs well, generalization to more diverse machining scenarios has no guarantees. The diversity comes from varied geometries, cutting conditions and sensor modalities and the generalisation achieved by MoE depends on the consistency of the data gathered from the machined parts. Guaranteeing such generalisation may need to include field tests on machining of arbitrary shapes to enable the identification of generalized signal–feature relationships for error prediction across tasks.

VI. Conclusion

This work proposes a mixture of experts framework fusing low-cost in-process signals, including external sensing of force and vibration, and internal Siemens controller data, to predict high-precision machining errors. It achieves a sevenfold improvement over direct laser tracker measurements based error prediction. Additionally, well-designed handcrafted features via a Mixture-of-Experts framework enable near metrology level accuracy (MAE = 0.0093 mm), comparable to a learning-based laser tracker method (MAE = 0.0089 mm). Such features allow real-time (latency: sub-5 ms) and energy-efficient (EDP: < 760 μJ·s) deployment on the CPU board. Future work will validate the framework on field trials with robotic systems that involve a variety of different geometries of parts being machined. Extending the framework to complex freeform surfaces represents a promising avenue for future work. Overall, this study demonstrates how low cost sensor based fusion with machine learning enable accurate, efficient, and scalable predictive manufacturing.

References

[1] V. D. Majstorovic, N. Durakbasa, Y. Takaya, and S. Stojadinovic, "Advanced manufacturing metrology in context of industry 4.0 model," in *Proceedings of the 12th International Conference on Measurement and Quality Control-Cyber Physical Issue: IMEKO TC 14 2019 12*. Springer, 2019, pp. 1–11.

[2] L. Yuan, Z. Pan, D. Ding, S. Sun, and W. Li, "A review on chatter in robotic machining process regarding both regenerative and mode coupling mechanism," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 5, pp. 2240–2251, 2018.

[3] Q. Peng, W. Li, C. Jiang, Y. Cheng, D. Wang, and W. Xu, "A novel path error compensation method for robotic milling by using hybrid temporal network," *IEEE/ASME Transactions on Mechatronics*, 2024.

[4] P. Zhou, X. Zhao, B. Tao, and H. Ding, "Combination of dynamical movement primitives with trajectory segmentation and node mapping for robot machining motion learning," *IEEE/ASME Transactions on Mechatronics*, vol. 28, no. 1, pp. 175–185, 2022.

[5] L. Lu, J. Han, F. Dong, Z. Ding, C. Fan, S. Chen, H. Liu, and H. Wang, "Joint-smooth toolpath planning by optimized differential vector for robot surface machining considering the tool orientation constraints," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 4, pp. 2301–2311, 2021.

[6] H. Xie, W.-l. Li, D.-H. Zhu, Z.-p. Yin, and H. Ding, "A systematic model of machining error reduction in robotic grinding," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 6, pp. 2961–2972, 2020.

[7] X. Zhao, H. Lu, W. Yu, B. Tao, and H. Ding, "Vision-based mobile robotic grinding for large-scale workpiece and its accuracy analysis," *IEEE/ASME Transactions on Mechatronics*, vol. 28, no. 2, pp. 895–906, 2022.

[8] Z.-Y. Liao, Q.-H. Wang, H.-L. Xie, J.-R. Li, X.-F. Zhou, and T.-H. Pan, "Optimization of robot posture and workpiece setup in robotic milling with stiffness threshold," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 1, pp. 582–593, 2021.

[9] X. Zhao, B. Tao, and H. Ding, "Multimobile robot cluster system for robot machining of large-scale workpieces," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 1, pp. 561–571, 2021.

[10] R. J. Hocken, P. H. Pereira, *et al.*, *Coordinate measuring machines and systems*. CRC press Boca Raton, FL, USA:, 2012, vol. 6.

[11] D. Bernd, M. Henriette, H. Gross, *et al.*, *Handbook of optical systems, volume 5: Metrology of optical components and systems*. John Wiley & Sons, 2012.

[12] P. Blecha, M. Holub, T. Marek, R. Jankovych, F. Misun, J. Smolik, and M. Machalka, "Capability of measurement with a touch probe on CNC machine tools," *Measurement*, vol. 195, p. 111153, 2022.

[13] B. Iñigo, A. Ibabe, G. Aguirre, H. Urreta, and L. N. López de Lacalle, "Analysis of laser tracker-based volumetric error mapping strategies for large machine tools," *Metals*, vol. 9, no. 7, p. 757, 2019.

[14] S. Wojciechowski, M. Suszyński, R. Talar, V. Černohlávek, and J. Štěrba, "Cutting force—vibration interactions in precise—and micromilling processes: A critical review on prediction methods," *Materials*, vol. 18, no. 15, p. 3539, 2025.

[15] L. Cen, S. N. Melkote, J. Castle, and H. Appelman, "A wireless force-sensing and model-based approach for enhancement of machining accuracy in robotic milling," *IEEE/ASME Transactions on Mechatronics*, vol. 21, no. 5, pp. 2227–2235, 2016.

[16] Q. Wang, W. Wang, L. Zheng, and C. Yun, "Force control-based vibration suppression in robotic grinding of large thin-wall shells," *Robotics and Computer-Integrated Manufacturing*, vol. 67, p. 102031, 2021.

[17] Y. Dai, Y. Xue, and J. Zhang, "Vibration-based milling condition monitoring in robot-assisted spine surgery," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 6, pp. 3028–3039, 2015.

[18] S. E. Yuksel, J. N. Wilson, and P. D. Gader, "Twenty years of mixture of experts," *IEEE transactions on neural networks and learning systems*, vol. 23, no. 8, pp. 1177–1193, 2012.

[19] H. Tercan and T. Meisen, "Machine learning and deep learning based predictive quality in manufacturing: a systematic review," *Journal of Intelligent Manufacturing*, vol. 33, no. 7, pp. 1879–1905, 2022.

[20] W. Sun, B. Yao, N. Zeng, B. Chen, Y. He, X. Cao, and W. He, "An intelligent gear fault diagnosis methodology using a complex wavelet enhanced convolutional neural network," *Materials*, vol. 10, no. 7, p. 790, 2017.

[21] Y. Bai, Z. Sun, J. Deng, L. Li, J. Long, and C. Li, "Manufacturing quality prediction using intelligent learning approaches: A comparative study," *Sustainability*, vol. 10, no. 1, p. 85, 2017.

[22] B. Bhandari and G. Park, "Non-contact surface roughness evaluation of milling surface using cnn-deep learning models," *International Journal of Computer Integrated Manufacturing*, vol. 37, no. 4, pp. 423–437, 2024.

[23] M. Leco, T. McLeay, and V. Kadirkamanathan, "A two-step machining and active learning approach for right-first-time robotic countersinking through in-process error compensation and prediction of depth of cuts," *Robotics and Computer-Integrated Manufacturing*, vol. 77, p. 102345, 2022.

[24] M. Leco and V. Kadirkamanathan, "A perturbation signal based data-driven gaussian process regression model for in-process part quality prediction in robotic countersinking operations," *Robotics and Computer-Integrated Manufacturing*, vol. 71, p. 102105, 2021.

[25] Y. Cheng, Y. Li, X. Liu, and Y. Cai, "Mechanism-based structured deep neural network for cutting force forecasting using CNC inherent monitoring signals," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 4, pp. 2235–2245, 2021.

[26] K. Zhu and Y. Zhang, "A cyber-physical production system framework of smart CNC machining monitoring system," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 6, pp. 2579–2586, 2018.

[27] C. Niu, B. Chen, S. Fletcher, P. Onawumi, E. Ozturk, M. Mahfouf, and V. Kadirkamanathan, "Learning-based robotic machining error prediction for high precision manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 100, p. 103217, 2026.

[28] T. Blu, P. Thévenaz, and M. Unser, "Linear interpolation revitalized," *IEEE Transactions on Image Processing*, vol. 13, no. 5, pp. 710–719, 2004.

[29] L. Pinheiro Cinelli, M. Araújo Marins, E. A. Barros da Silva, and S. Lima Netto, "Variational autoencoder," in *Variational methods for machine learning with applications to deep networks*. Springer, 2021, pp. 111–149.

[30] S. Hu, Z. Lou, X. Yan, and Y. Ye, "A survey on information bottleneck," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 8, pp. 5325–5344, 2024.

[31] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[32] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[33] A. Asperti and M. Trentin, "Balancing reconstruction error and kullback-leibler divergence in variational autoencoders," *IEEE Access*, vol. 8, pp. 199 440–199 448, 2020.

[34] M. N. Suvon, P. C. Tripathi, W. Fan, S. Zhou, X. Liu, S. Alabed, V. Osmani, A. J. Swift, C. Chen, and H. Lu, "Multimodal variational autoencoder for low-cost cardiac hemodynamics instability detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 296–306.

[35] G. Yuan, Y. Li, S. Li, Z. Kong, S. Tulyakov, X. Tang, Y. Wang, and J. Ren, "Layer freezing & data sieving: Missing pieces of a generic framework for sparse training," *Advances in Neural Information Processing Systems*, vol. 35, pp. 19 061–19 074, 2022.

[36] Y. Zeng, J. Zhang, S. Guo, B. Ye, Z. Qu, Z. Hong, B. Tang, J. Chen, J. Zhou, and J. Yu, "Afedlf: Adaptive layer freezing of foundation models in heterogeneous federated learning," *IEEE Transactions on Mobile Computing*, 2025.

[37] M. Papananias, T. E. McLeay, O. Obajemu, M. Mahfouf, and V. Kadirkamanathan, "Inspection by exception: A new machine learning-based approach for multistage manufacturing," *Applied Soft Computing*, vol. 97, p. 106787, 2020.

[38] A. Lu, W. Wang, C. Li, J. Tang, and B. Luo, "AFter: Attention-based fusion router for RGBT tracking," *IEEE Transactions on Image Processing*, 2025.

[39] D. Wu, C.-T. Lin, J. Huang, and Z. Zeng, "On the functional equivalence of TSK fuzzy systems to neural networks, mixture of experts, CART, and stacking ensemble regression," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 10, pp. 2570–2580, 2019.



Chaoyue Niu received an M.Sc. degree in Robotics from King's College London, U.K., in 2017 and a Ph.D. degree in Computer Science from the University of Southampton, U.K., in 2023. He is currently a research associate at the University of Sheffield. His research interests include autonomous navigation in unstructured environments, multimodal sensing and in-process metrology in robotic and composite manufacturing, and acoustic and multimodal analysis for pervasive healthcare monitoring.



Visakan Kadirkamanathan received the B.A. degree in electrical and information sciences and the Ph.D. degree in information engineering from the Department of Engineering, University of Cambridge, U.K., in 1987 and 1992, respectively. He is currently a Professor in Signal and Information Processing and the Director of the Rolls-Royce University Technology Centre for Control, Monitoring, and Systems Engineering at Sheffield. His research interests are in the areas of system identification, data-driven modeling, signal processing, and control with applications in aerospace, manufacturing, biomedical, and other dynamic systems. Dr. Kadirkamanathan was a recipient of the 2013 Proceedings of the National Academy of Science (PNAS) Cozzarelli Prize, USA. He is a past Associate Editor of the IEEE Transactions on Neural Networks and Editor-in-Chief of the International Journal of Systems Science.