Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Smartphone-Based Vibration Monitoring for Drill Bit Wear Classification Using Lightweight CNN Fusion Strategies

Bahram Rafizadeh Ghahderijan
Iran University of Science and Technology
Mohammad Shahbazi
Iran University of Science and Technology
Seyed Ali Niknam
seyed-ali.niknam@polymtl.ca

Polytechnique Montreal

Research Article

Keywords: Tool condition monitoring; Smartphone vibration sensing; Drill wear classification; Lightweight convolutional neural networks; Multi-axis sensor fusion

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11353116/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Smartphone-Based Vibration Monitoring for Drill Bit Wear Classification Using Lightweight CNN Fusion Strategies

Bahram Rafizadeh Ghahderijani[1], Mohammad Shahbazi[1], and Seyed Ali Niknam*[1,2]

[1]School of Mechanical Engineering, Iran University of Science and Technology, Narmak, Tehran 13114-16846, Iran
[2]Department of Mechanical Engineering, Polytechnique Montreal, Canada

Abstract

This study investigates the feasibility of using a consumer-grade smartphone as a low-cost, fixtureless sensing node to monitor and classify drilling bit wear. Tri-axial vibration signals were acquired using the embedded MEMS accelerometer of a smartphone placed freely on the machine table, representing a practical worst-case deployment scenario without rigid sensor mounting. This work evaluates whether severely bandwidth-limited, sub-Nyquist vibration measurements can still retain diagnostic information for drilling-state identification and progressive drill wear. Raw vibration recordings were segmented, transformed into compact two-dimensional pseudo-RGB representations, and classified using lightweight convolutional neural network architectures. We systematically compared three multi-axis fusion strategies—Data-Level, Feature-Level, and Decision-Level fusion— using LeNet-5 and Mini-MobileNetV2 models, with Support Vector Machine and Random Forest classifiers as traditional baselines. The results show that Mini-MobileNetV2 with Data-Level fusion achieved the best drilling-phase detection performance, whereas Feature-Level fusion provided the strongest tool-wear classification performance. Robustness analyses further demonstrated that Feature-Level and Decision-Level fusion were more resilient to simulated sensor saturation and angular misalignment than Data-Level fusion. Overall, the findings support the potential of smartphone-based vibration sensing as an accessible proof-of-concept approach for rapid, non-invasive tool condition monitoring, while highlighting the need for broader validation using additional machine-tool platforms, smartphone devices, and reference-grade sensing systems.

Keywords: Tool condition monitoring; Smartphone vibration sensing; Drill wear classification; Lightweight convolutional neural networks; Multi-axis sensor fusion

1 Introduction

Machining processes are fundamental to modern manufacturing and are widely used in industrial sectors such as automotive, aerospace, medical, oil and gas, and precision engineering [1]. These processes offer high accuracy, repeatability, and production flexibility; however, increasing demands

*Corresponding author. Email: seyed-ali.niknam@polymtl.ca

for productivity, cost reduction, and sustainability have intensified the need for intelligent monitoring and optimization strategies [2].

Cutting tool wear is one of the most important factors affecting machining quality, productivity, and operating cost. Progressive wear increases cutting forces and friction at the tool–workpiece interface, leading to dimensional inaccuracies and degraded surface finish. Tool condition monitoring (TCM), therefore, plays a vital role in predictive maintenance and quality assurance. Conventional TCM systems typically rely on high-frequency industrial sensors, such as piezoelectric accelerometers, with sampling rates in the kilohertz range to capture fast transient phenomena related to tool wear. Although these systems provide high-resolution information, they are often expensive, invasive, and dependent on rigid mounting fixtures that can disrupt production setups. This limits their practicality for small workshops and decentralized manufacturing environments where rapid deployment is essential [3].

Machine learning (ML) and artificial intelligence (AI) have become central tools in smart manufacturing. ML algorithms can learn patterns from historical and real-time data, enabling predictive analysis and decision-making with limited manual intervention [4]. In machining, AI-driven methods have been used to predict tool degradation, estimate remaining useful life, and optimize maintenance schedules. These approaches can reduce production downtime, extend tool lifespan, and improve machining efficiency, particularly in controlled or well-instrumented environments [5].

In parallel, smartphones have become ubiquitous and are equipped with embedded micro-electromechanical system (MEMS) sensors, including triaxial accelerometers. These devices offer an attractive low-cost alternative for decentralized data acquisition. However, commercial smartphones operate at severely constrained sampling rates and generally lack the dedicated analog anti-aliasing and data acquisition hardware used in industrial monitoring systems. Rather than treating this limitation solely as a barrier, this study examines whether sub-Nyquist smartphone vibration measurements can still retain useful diagnostic information for drilling-state identification and tool wear classification.

According to sampling theory, vibration components above the Nyquist limit may be aliased into the observable frequency range when they are sampled without sufficient anti-aliasing. One hypothesis examined in this work is that undersampled high-frequency signatures associated with flank wear, micro-fracture, and friction-induced vibration may appear as structured low-frequency patterns in the recorded smartphone signals. This study does not claim that constructive aliasing is the only mechanism behind the observed signal characteristics. Instead, it evaluates whether smartphone-based vibration measurements, despite limited sampling bandwidth and complex structural resonances, can provide discriminative features consistent with tool wear progression under realistic workshop conditions [6].

By transforming these 1D vibration signals into 2D pseudo-RGB images, convolutional neural networks (CNNs) can learn spatial texture patterns associated with drilling state and tool condition. Based on this concept, the present study investigates the feasibility of using a smartphone as a low-cost sensing node for tool condition monitoring in drilling operations. Unlike conventional studies that rely on rigidly mounted sensors, this work adopts a practical “worst-case” deployment scenario in which the smartphone is placed freely on the machine table. This fixtureless setup serves as a proof of concept for accessible shop-floor monitoring where operators need rapid, non-invasive solutions without complex fixturing.

The primary objective is to develop an intelligent, data-driven framework to identify active cutting phases and classify discrete drill wear conditions, including healthy, moderately worn, and severely worn tools. The main contributions of this paper are summarized as follows:

- A low-cost, fixtureless monitoring framework using multi-axis vibration data acquired from a commercial smartphone accelerometer.
- Validation of the diagnostic potential of sub-Nyquist smartphone vibration measurements through signal-to-image transformation and lightweight CNN-based classification.
- A systematic comparison of Data-Level, Feature-Level, and Decision-Level multi-axis fusion strategies using lightweight CNN architectures, including Mini-MobileNetV2.
- A robustness analysis of the proposed fusion frameworks under simulated sensor saturation and angular misalignment.

By adopting a "worst-case" unconstrained deployment, this research emphasizes practicality and rapid scalability over idealized laboratory conditions, establishing a proof of concept for accessible intelligent manufacturing tools.

The remainder of this paper is organized as follows. Section 2 reviews related work on vibration-based tool condition monitoring, smartphone sensing, and data-driven machining diagnostics. Section 3 describes the experimental setup, signal preprocessing, learning architectures, fusion strategies, and robustness simulations. Section 4 presents the classification and robustness results. Section 5 discusses the main findings, limitations, and future directions. Finally, Section 6 concludes the paper.

2 Literature Review

The manufacturing industry has undergone a major transformation through the integration of the Internet of Things (IoT) [7], cloud computing, mobile connectivity [8], and AI [9]. These technologies have accelerated the transition from conventional automation toward smart manufacturing, where data-driven monitoring and decision-making are essential for improving productivity, reliability, and process quality.

Online monitoring of machining processes is a central component of smart production [10]. Existing studies generally address three related monitoring targets: workpiece condition, including surface integrity [11], roughness [12], and geometric accuracy [13, 14]; tool condition, including wear and breakage detection [15, 16]; and process–machine interaction, including stability analysis [17], chatter detection [18], and energy efficiency [19]. Among these topics, TCM remains particularly important because tool degradation directly affects dimensional accuracy, surface finish, production cost, and machine reliability.

2.1 Sensing Modalities in Tool Condition Monitoring

A wide range of sensors has been used to capture physical phenomena associated with machining. Acoustic emission (AE) sensors detect transient elastic waves generated by rapid energy release

and effectively identify crack initiation, tool fracture, and other high-frequency events [20]. Force sensors provide indirect but informative measurements of the cutting process by capturing changes in cutting load and tool–workpiece interaction [21]. Vibration sensors are also widely used because they can capture dynamic changes in displacement, velocity, and acceleration across time, frequency, and modal domains [22]. In TCM applications, vibration signals are especially attractive because they are sensitive to tool wear, chatter, imbalance, and other dynamic instabilities.

Despite their effectiveness, many conventional sensing configurations rely on industrial-grade sensors, dedicated data acquisition systems, high sampling rates, and rigid mounting conditions. These requirements improve measurement fidelity but can increase cost, installation complexity, and invasiveness. As a result, low-cost, rapidly deployable sensing approaches are still needed to support practical monitoring in small workshops, educational laboratories, and decentralized manufacturing environments.

2.2 Data-Driven and Machine Learning Approaches

ML has become a key tool for extracting diagnostic information from machining data. Schueller and Saldana [23] investigated the generalizability of TCM models and showed that changes in cutting parameters, such as speed, can strongly influence wear prediction performance. Demir and Yesilyurt [24] compared several ML techniques, including Random Forest and neural-network-based models, with Continuous Wavelet Transform (CWT) approaches for tool breakage detection. Their findings indicate that ML methods can provide useful failure detection performance while reducing reliance on computationally intensive signal-processing pipelines.

The performance of ML-based TCM systems depends strongly on data quality, preprocessing, feature extraction, and model tuning. Dey and Yodo [25] showed that data cleaning, feature selection, and hyperparameter optimization can substantially improve neural-network-based wear prediction. These findings highlight that model performance depends not only on classifier architecture but also on the quality and structure of the sensing and preprocessing pipeline.

2.3 Deep Learning-Based Tool Condition Monitoring

Recent TCM research has increasingly shifted toward Deep Learning (DL), which can automatically learn features from raw or transformed signals. Recurrent architectures are commonly used for temporal modeling. Hassan et al. [26] combined Discrete Wavelet Transform (DWT) with a Long Short-Term Memory (LSTM) autoencoder and used reconstruction error to predict sudden tool failure from acoustic emission signals. Gudelek et al. [27] proposed a hybrid LSTM–Perceptron model with CWT preprocessing and demonstrated adaptability to variations in cutting speed and tool diameter for wear regression.

Another important trend is converting 1D signals into 2D representations so CNNs can exploit spatial feature learning. Li et al. [28] converted multi-source signals into 2D time–frequency images using CWT and applied transfer learning with Bayesian optimization, achieving high prediction accuracy. For chatter detection, Kounta et al. [29] used Fast Fourier Transform (FFT) images as inputs to pre-trained networks such as VGG16 and ResNet50, while Zhang et al. [30] proposed a hybrid CNN architecture using cutting-force signals. These studies show that image-based

representations can make vibration, force, or acoustic data more compatible with modern CNN architectures.

Transfer learning has also been used to reduce training time and improve performance when labeled machining data are limited. Ross et al. [31] evaluated pre-trained networks, including AlexNet, VGG-16, and ResNet, for tool wear detection under different cutting environments. Venkatesh et al. [32] converted vibration signals into image-like plots and compared architectures such as VGG-16, GoogLeNet, and ResNet-50 for single-point cutting tool monitoring. Ou et al. [33] extended this direction by applying residual transfer learning to multi-sensor data, including audio, vibration, and current signals, converted into image representations for impeller machining monitoring.

2.4 Low-Cost and Smartphone-Based Sensing

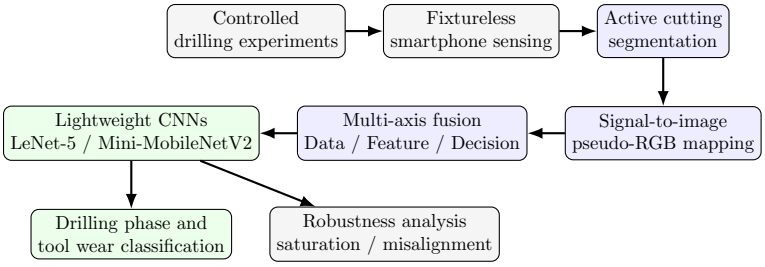
While the above studies demonstrate the effectiveness of advanced sensing and learning methods, most rely on relatively controlled experimental conditions and dedicated measurement systems. Smartphones offer a different sensing paradigm because they integrate low-cost MEMS sensors and convenient data logging in a widely available platform. Tools such as phyphox have also demonstrated the potential of smartphones for sensor-based experiments and data acquisition [34]. These features make smartphones attractive for rapid and non-invasive shop-floor monitoring.

However, smartphone-based vibration monitoring introduces important technical limitations. Embedded MEMS accelerometers typically provide lower sampling rates and less controlled sensor mounting than industrial accelerometers. In machining applications, where tool wear and fracture can generate high-frequency vibration and acoustic-emission components, these limitations can obscure or distort the original physical signal. Prior vibration-based studies have shown that tool wear can be reflected in measurable dynamic responses [35, 37], while classical signal-processing theory explains how undersampled high-frequency components may fold into the observable baseband [36, 38]. These observations motivate further investigation into whether bandwidth-limited smartphone signals can still preserve useful diagnostic information for TCM.

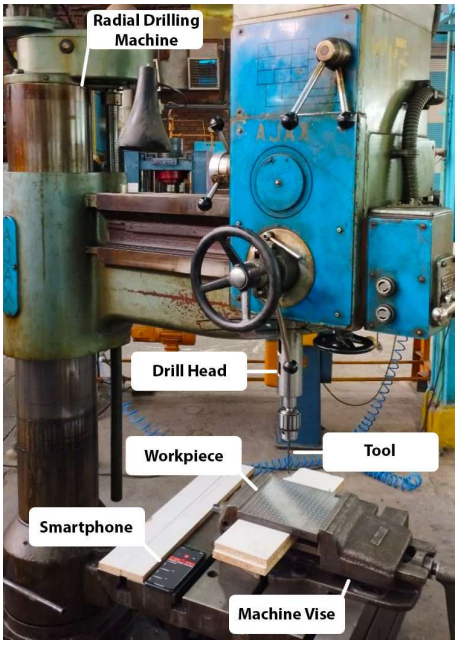
2.5 Research Gap and Positioning of the Present Study

Overall, the literature shows strong progress in vibration-based TCM, ML classification, deep learning, transfer learning, and signal-to-image representations. Nevertheless, several gaps remain. First, many existing approaches depend on high-frequency industrial sensors and rigid mounting, which limits rapid deployment. Second, relatively limited attention has been given to fixtureless smartphone-based vibration acquisition in machining environments. Third, few studies systematically examine how multi-axis fusion strategies behave when low-cost sensor data are affected by practical disturbances such as saturation and angular misalignment.

The present study addresses these gaps by evaluating a smartphone-based, fixtureless vibration monitoring framework for drilling. It specifically examines whether bandwidth-limited, potentially aliased smartphone vibration signals can still retain diagnostic information related to drilling state and tool wear. The framework combines active cutting segmentation, signal-to-image transformation, lightweight CNN classification, and a systematic comparison of Data-Level, Feature-Level, and Decision-Level fusion. In addition, robustness tests under simulated sensor saturation and an-



(a) Overview of the proposed smartphone-based tool condition monitoring workflow. The process begins with fixtureless tri-axial vibration acquisition using a freely placed smartphone and proceeds through active cutting segmentation, signal-to-image pseudo-RGB transformation, and multi-axis fusion. The resulting representations are analyzed with lightweight CNN architectures for drilling-state identification and drill-wear classification. The workflow also incorporates robustness evaluation under simulated sensor saturation and angular misalignment to assess the practical reliability of the proposed monitoring framework.



(b) Physical experimental setup with smartphone placement.

Figure 1: Experimental framework for smartphone-based tool condition monitoring. Subfigure (a) summarizes the workflow from fixtureless vibration acquisition to fusion-based CNN classification and robustness evaluation. Subfigure (b) illustrates the physical drilling setup and the freely placed smartphone used for in-process vibration data acquisition.

gular misalignment are included to assess the practical reliability of the proposed framework under non-ideal deployment conditions.

3 Materials and Methods

The proposed TCM framework introduces a low-cost, non-invasive, and data-driven approach based on smartphone-embedded sensing. The central idea is to examine whether severely band-limited smartphone vibration signals, including potentially aliased wear-related components, can be transformed into 2D spatial features suitable for CNN classification. The methodology is structured into four primary stages: (i) controlled machining experiments, (ii) smartphone-based vibration data acquisition under non-rigid mounting conditions, (iii) signal-to-image transformation of multi-axis data, and (iv) supervised deep learning-based classification for drilling phase detection and discrete wear assessment. Figure 1a summarizes the proposed framework's overall workflow, linking the experimental drilling campaign, fixtureless smartphone-based vibration acquisition, signal-to-image conversion, multi-axis fusion strategies, lightweight CNN classification, and robustness evaluation.

3.1 Experimental Design and System Overview

Unlike conventional TCM setups that rely on expensive, high-frequency industrial sensors (e.g., piezoelectric accelerometers) and dedicated data acquisition (DAQ) hardware, the proposed framework uses a commercial smartphone as a standalone vibration-sensing device. The embedded tri-axial MEMS accelerometer is used to capture structural vibrations induced by the tool–workpiece interaction in a non-invasive manner.

Commercial smartphone digitizers generally lack ideal analog anti-aliasing filters. One hypothesis investigated in this study is that this hardware characteristic, although traditionally viewed as a limitation, may still preserve useful diagnostic information. Under these conditions, high-frequency acoustic emissions and friction-induced vibrations (typically $> 1\,\text{kHz}$) associated with severe tool wear may be partially aliased into the observable low-frequency baseband (e.g., $< 100\,\text{Hz}$). This possibility motivates using smartphone measurements for low-cost diagnostic feature extraction, while acknowledging that the recorded signals are not faithful broadband vibration measurements.

To validate this approach, the experimental campaign simulated realistic drilling operations under constant, controlled machining parameters (spindle speed and feed rate). The device recorded vibration data directly for subsequent transformation and deep learning analysis. To ensure robust supervised learning, the ground truth for tool wear states was rigorously established using optical imaging after each drilling cycle.

3.1.1 Experimental Setup and Data Acquisition

To evaluate the proposed TCM framework, we designed a series of controlled drilling experiments. The setup emulates realistic industrial conditions while using the smartphone sensor's hardware constraints as part of the investigated low-cost sensing scenario.

Machining Parameters and Materials Experiments were conducted on a heavy-duty industrial radial drilling machine. This machine was selected for its high structural rigidity, which reduces ambient environmental noise and helps ensure that the primary vibrations captured by the sensor originate from the cutting process. A full-factorial design of experiments (DoE) was implemented under dry cutting conditions to accelerate tool wear and eliminate the damping effects of cutting fluids.

The selected operating parameters comprised three cutting speeds (V_c = 9, 16, and 24 m/min, corresponding to spindle speeds N = 474, 830, and 1270 rpm) and three feed rates (f = 0.10, 0.20, and 0.30 mm/rev). The cutting tools were 6 mm diameter, 5% cobalt-alloyed high-speed steel (HSS-Co) twist drills (DIN 338, 118° point angle, 30° helix angle). The workpiece material was hot-work tool steel (AISI H13 / DIN 1.2344), selected for its high mechanical strength, thermal resistance, and ability to induce distinct tool wear during dry machining. For each parameter combination, a new drill bit was used and operated until failure or severe wear was visually confirmed.

Non-Invasive Sensing and Constructive Aliasing Tri-axial vibration signals (X, Y, and Z) were acquired using a commercial smartphone (POCO X3 Pro) equipped with an internal MEMS

accelerometer. Data logging was facilitated by the Phyphox application [34] at a nominal requested sampling rate of 300 Hz. However, due to hardware constraints, the effective sampling rate was empirically observed at approximately $f_s \approx 200$ Hz.

To align with the study's low-cost, non-invasive objective, the smartphone was placed freely on the machine table adjacent to the workpiece vise, without any rigid mounting fixture (Figure 1b). The setup was deliberately designed to emulate practical shop-floor scenarios in which ideal sensor mounting may be infeasible because of operational constraints, space limitations, or the need for rapid deployment. Therefore, this unconstrained mechanical coupling was selected to evaluate the system under non-ideal sensor installation conditions.

This data acquisition strategy also allows the study to examine the possible role of constructive aliasing. The effective sampling rate establishes a Nyquist limit of roughly 100 Hz. Given the spindle speeds (7.9 Hz to 21.1 Hz) and the two-flute drill geometry, the fundamental cutting frequencies (15.8 Hz to 42.2 Hz) and their immediate harmonics naturally fall within this baseband. However, friction-induced vibrations and structural acoustic emissions caused by progressive flank wear often occur at higher frequencies (frequently > 1 kHz). Because smartphone sensor digitizers generally lack stringent analog anti-aliasing filters, these high-frequency micro-dynamics may not be fully rejected and may partially fold back into the 0–100 Hz observable spectrum. If this aliasing interpretation holds, tool wear may appear as broadband energy elevation or altered low-frequency texture. Under this assumption, the CNN could extract relevant wear features without requiring expensive, high-frequency industrial DAQ systems, although other factors, such as structural resonances and nonlinear machine vibrations, may also contribute to the observed spectral changes.

Optical Ground Truth Establishment To generate reliable labels for the supervised deep learning models, we visually inspected the cutting tools offline after successive drilling cycles (see Figure 2). The drill tips were examined using an optical microscope at a fixed 15× magnification. High-resolution digital micrographs were captured to document progressive morphological changes on the tool flank and chisel edge (Figure 3). This visual evidence served as the ground truth, allowing the continuous vibration datasets to be categorized into discrete tool-health states for subsequent classification tasks.

3.2 Signal Preprocessing and Dataset Formulation

Raw vibration logs inherently contain extended periods of non-machining states, including idle spindle rotation, tool approach, and retraction. To prevent the deep learning models from learning irrelevant mechanical noise and to isolate the temporal window in which continuous tool–workpiece contact generates wear-related vibration signatures, a systematic signal segmentation protocol was implemented.

3.2.1 Envelope-Based Active Phase Segmentation

To extract the active drilling phase, we adopted an envelope analysis based on the Hilbert transform. Initially, the tri-axial acceleration components (X, Y, and Z) were fused into a single resultant magnitude signal to capture the total vibratory energy with reduced dependence on the

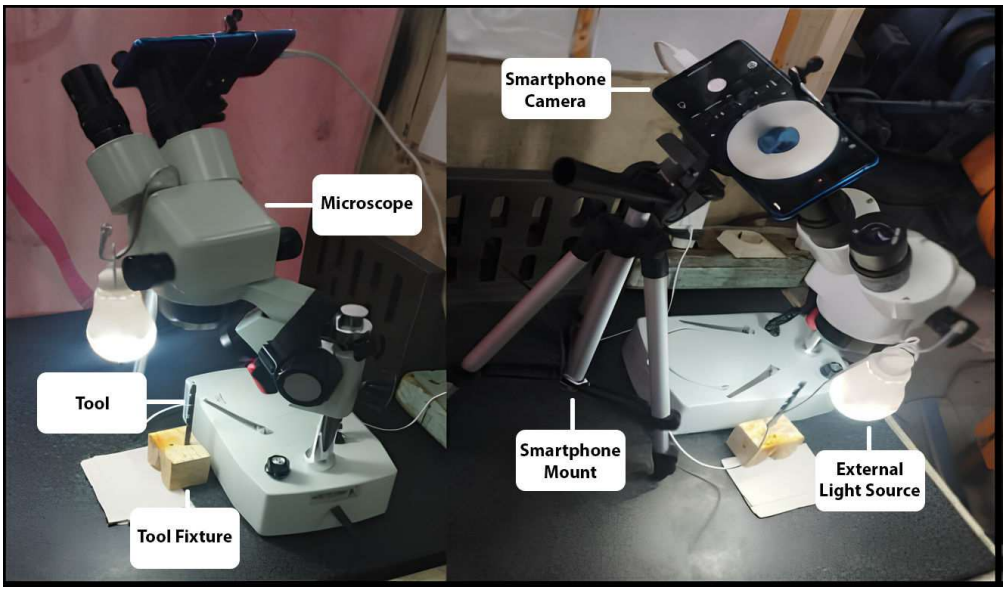
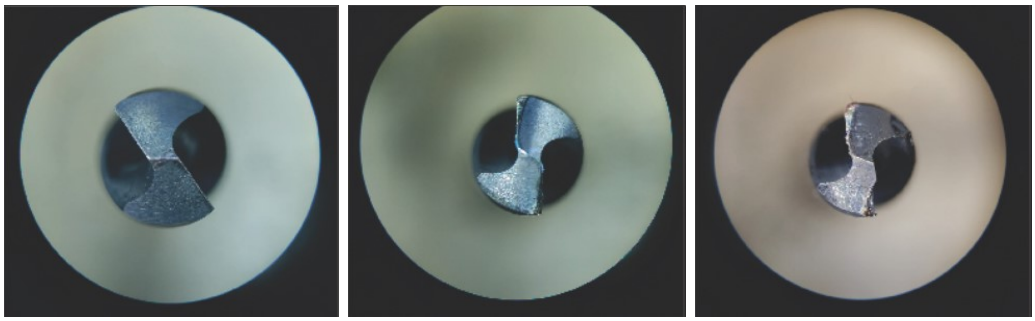
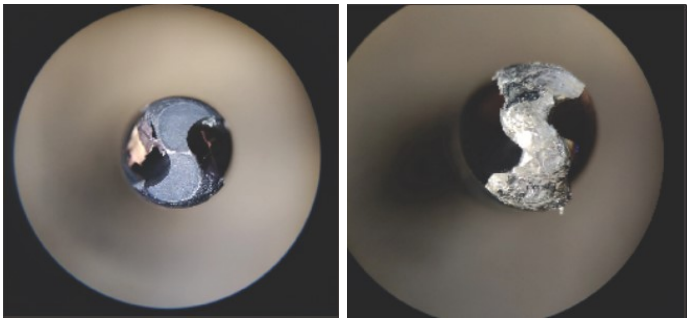


Figure 2: Optical inspection setup used for ground-truth data collection. A smartphone is mounted to the eyepiece of a stereomicroscope, with auxiliary illumination, to capture high-resolution micrographs of drill bit flank wear after machining cycles.



Undamaged Partially Worn Heavily Worn



Fully Worn Worn Out

Figure 3: Representative optical micrographs illustrating the progressive morphological degradation of the drill bit. The visual ground truth captures the evolution of flank wear and plastic deformation across distinct physical stages.

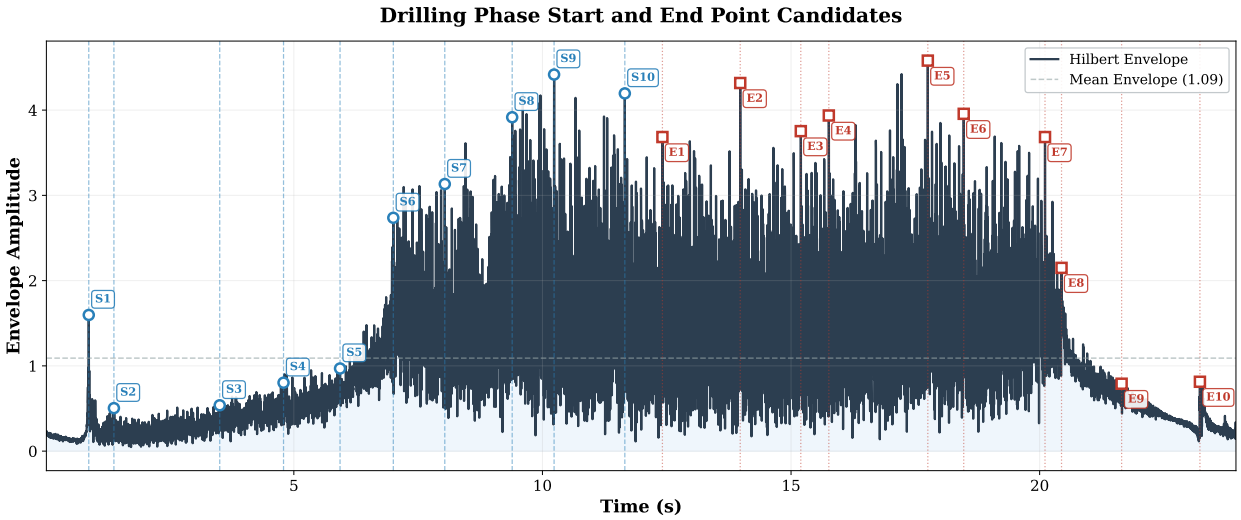


Figure 4: Representative envelope analysis of the fused tri-axial vibration signal used for active drilling phase segmentation. The horizontal solid gray line defines the mean energy threshold. Vertical dashed blue (S1–S10) and dotted orange (E1–E10) lines indicate the algorithmically proposed candidate start and end points of the cutting process, respectively, enabling precise semi-automated extraction of the continuous machining interval.

smartphone's spatial orientation:

$$A(t) = \sqrt{x(t)^2 + y(t)^2 + z(t)^2} \tag{1}$$

The analytic representation of A(t) was computed via the Hilbert transform to extract the low-frequency amplitude envelope. Because the continuous cutting phase exhibits a sustained elevation in high-energy vibrations compared to air-cutting, the mean envelope energy served as a baseline threshold. A semi-automated peak-detection algorithm identified the top candidate boundaries for the start and end of the cut. These localized boundaries were graphically verified (Figure 4) to eliminate transient entry/exit anomalies, effectively isolating the pure active drilling segment.

3.2.2 Construction of Task-Specific Databases

To address the dual objectives of this study–real-time process monitoring and discrete tool condition assessment–the segmented data was structured into two distinct, non-overlapping databases:

- Database I (Drilling State Identification): Designed as a binary classification task to evaluate if aliased smartphone vibrations can reliably differentiate between machining states. It comprises two classes: Active Drilling (the isolated segments) and non-drilling (idle/air-cutting segments).
- Database II (Tool Wear Classification): Designed to map active drilling vibrations to physical drill-bit degradation. To establish rigorous, data-driven ground truth labels, the optical micrographs of the tool flank (captured post-machining) were processed using a custom computer vision model trained via the Roboflow platform. The model segmented and quantified the exact pixel area of tool wear.

To eliminate class imbalance–a common bias in TCM datasets–and avoid arbitrary subjective wear thresholds, the extracted geometrical wear areas were quantified and divided into three equal tertiles. This generated a perfectly balanced, three-class dataset: Undamaged (Class 0) corresponding to the lowest relative wear area, Partially Worn (Class 1) for the intermediate tertile, and Heavily Worn (Class 2) for the highest optical wear thresholds. As this study constitutes a proof-of-concept investigation, the tertile-based geometric labeling is deliberately chosen: it enforces a balanced and objective classification scheme that is strictly optimized for machine learning evaluation, rather than relying on absolute physical wear standards (e.g., VBmax) which would require a different metrological scope.

3.3 Sub-Nyquist Signal Interpretation

3.3.1 Aliasing-Based Wear Encoding

The dynamic signature of a drilling process is fundamentally governed by the spindle speed (N_{spindle}, in RPM) and the number of cutting flutes (z). According to foundational machining theory [35], the fundamental tool-passing frequency (f_{TP}) is defined as:

$$f_{\text{TP}} = \frac{N_{\text{spindle}} \times z}{60} \tag{2}$$

Under healthy cutting conditions, the dominant vibration energy is concentrated at f_{TP} and its lower-order harmonics, which typically fall well within the operational bandwidth of the smartphone's MEMS accelerometer.

However, as tool wear progresses–manifesting as severe flank wear, micro-chipping, or built-up edge (BUE) formation–the tribological conditions at the tool-workpiece interface deteriorate. Tool condition monitoring literature indicates that such degradation can generate high-frequency acoustic emissions and broadband structural excitation components (f_{wear}) [47].

To motivate the hypothesis that a low-sampling-rate smartphone sensor can still contain information related to high-frequency tool wear phenomena, the concept of constructive aliasing is mathematically formulated. According to the fundamental principles of discrete-time signal processing established by Oppenheim and Schafer [36], aliasing is a spectral phenomenon that directly dictates the shape of the sampled discrete-time signal. When a continuous-time signal $x_c(t)$ is sampled at intervals of T, the spectrum of the resulting discrete-time signal, $X_s(j\omega)$, is expressed as an infinite sum of shifted replicas of the original continuous-time spectrum, $X_c(j\omega)$:

$$X_s(j\omega) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X_c\left(j(\omega - k\omega_s)\right) \tag{3}$$

where ω is the angular frequency, ω_s is the sampling angular frequency ($\omega_s = 2\pi/T$), and k is an integer ($k = 0, \pm1, \pm2, \ldots$) representing the harmonic shift index. Converting the angular frequencies into cyclic frequencies in Hertz ($f = \omega/2\pi$ and $f_s = \omega_s/2\pi$, where f_s is the smartphone's sampling rate), the equation is rewritten as:

$$X_s(f) = f_s \sum_{k=-\infty}^{\infty} X_c(f - kf_s) \tag{4}$$

In the context of the machining process, the continuous-time mechanical vibration may contain high-frequency energy signatures generated by tool flank wear, friction, and micro-fractures. Let the dominant frequency of this wear-induced signature be denoted as f_{wear}. Because the smartphone's internal IMU operates at a constrained sampling rate (e.g., $f_s \approx 400\,\text{Hz}$), the Nyquist criterion is violated ($f_{\text{wear}} \gg f_s/2$) [37].

Due to the infinite summation in the mathematical formulation, if high-frequency wear-related energy is present, it would not necessarily be entirely lost under ideal sampling assumptions. Instead, it could be periodically shifted by integer multiples of f_s. While this shift is mathematically defined in the frequency domain, its physical manifestation would occur in the time domain. The shifted frequency that folds back into the positive observable baseband can be expressed as:

$$f_{\text{alias}} = |f_{\text{wear}} - k f_s| \tag{5}$$

where the integer k ensures that the alias frequency falls within the sensor's physical bandwidth ($0 \leq f_{\text{alias}} \leq f_s/2$).

Under the assumption that this aliasing phenomenon is detectable in the present experimental setup, the proposed methodology could capitalize on how these potentially aliased frequencies manifest in the discrete time domain. Specifically, high-frequency wear (f_{wear}) might modulate the raw discrete time-series $x[n]$ into complex, lower-frequency waveforms (f_{alias}). By mapping these raw 1D temporal sequences directly into 2D spatial matrices (RGB images), any such spectral fold-back waveforms would, if present, be translated into deterministic spatial textures. The deployed CNNs could then extract and learn these structural time-domain textures. Thus, the framework operates under the working hypothesis that constructive aliasing–if it occurs–may provide a pathway to bypass both expensive high-frequency DAQ hardware and computationally heavy frequency-domain transformations. The empirical validation of this hypothesis is examined in the following section.

3.3.2 Spectral Exploration Metric: power spectral density (PSD)

To empirically validate the constructive aliasing hypothesis formulated in Section 3.3.1, a robust frequency-domain analysis of the raw vibration signals is required. Given that machining vibration signals are highly stochastic, non-stationary, and contaminated with structural noise, a standard FFT periodogram produces a spectral estimate with high variance, which can mask the underlying aliasing trends [38]. Therefore, Welch's method is employed for Power Spectral Density (PSD) estimation.

Welch's method mitigates spectral variance by dividing the discrete 1D time-series signal, $x[n]$, into K overlapping segments of length L. A windowing function, $w[n]$, is applied to each segment to reduce spectral leakage. The modified periodogram for the k-th segment, $P_k(f)$, is mathematically defined as [36]:

$$P_k(f) = \frac{1}{f_s L U} \left| \sum_{n=0}^{L-1} x_k[n] w[n] e^{-j2\pi f n / f_s} \right|^2 \tag{6}$$

where f_s is the sampling frequency and U is the window power normalization constant, calculated as $U = \frac{1}{L} \sum_{n=0}^{L-1} |w[n]|^2$, which compensates for the energy loss introduced by the windowing function.

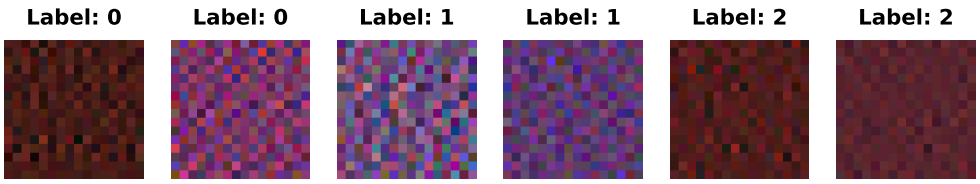


Figure 5: Representative pseudo-RGB images generated from segmented tri-axial vibration signals and their corresponding class labels.

The final robust PSD estimate, $S_{xx}(f)$, is obtained by averaging the modified periodograms of all K segments [36]:

$$S_{xx}(f) = \frac{1}{K} \sum_{k=1}^{K} P_k(f) \tag{7}$$

PSD estimation was configured according to the smartphone acquisition constraints. The effective sampling frequency was $f_s \approx 200\,\text{Hz}$, yielding a Nyquist limit of $f_{\text{Nyquist}} \approx 100\,\text{Hz}$. To balance frequency resolution and variance reduction, a Hann (Hanning) window was applied. The segment length was set to $L = 256$ data points with a 50% overlap between consecutive segments (128 points). This configuration yields a frequency resolution of $\Delta f = \frac{f_s}{L} \approx 0.8\,\text{Hz}$, which is sufficient for examining broadband energy accumulated within the baseband while reducing sensitivity to localized noise spikes.

3.4 Signal-to-Image Transformation and Multi-Axis Fusion Strategies

To exploit the spatial pattern-recognition capability of CNNs, the 1D vibration signals (X, Y, and Z axes) were mapped into a 2D spatial domain.

3.4.1 Sequential-to-Spatial Mapping

The isolated active drilling signals were segmented into discrete frames using a sliding-window approach with a length of 256 consecutive samples and a 50% overlap, corresponding to approximately 1.24 s of machining per window. To mitigate amplitude scaling biases, each frame underwent a two-stage Min-Max normalization: initially scaled to [0,255] for 8-bit integer conversion, and subsequently divided by 255 to yield a [0,1] floating-point range required by CNN frameworks.

Each normalized 256-point 1D array was reshaped into a compact 16×16 2D spatial matrix. The spatial matrices of the X, Y, and Z axes were assigned to the red, green, and blue channels, respectively, producing a unified $16 \times 16 \times 3$ pseudo-RGB image (Figure 5). This representation preserves temporal variations as spatial texture patterns, pixel-intensity gradients, and regional contrast changes that can be learned by CNN models.

3.4.2 Fusion Topologies

In machining operations, vibrations along different axes are strongly coupled. To determine the optimal integration hierarchy for extracting these cross-axis correlations, three distinct multi-axis fusion topologies were evaluated [39]:

- Data-Level Fusion (Early Fusion): The network processes the single fused RGB tensor of size $16 \times 16 \times 3$, implicitly learning inter-axis spatial correlations from the first convolutional layer.
- Feature-Level Fusion (Intermediate Fusion): Three parallel feature extractors independently process single-channel inputs ($16 \times 16 \times 1$) from each axis. Deep, abstract feature vectors extracted from the final layers of these branches are flattened and concatenated into a unified representation before classification.
- Decision-Level Fusion (Late Fusion): Three fully independent classifiers are trained per axis, each yielding a standalone Softmax probability vector. These discrete predictions are subsequently aggregated via concatenation and a secondary meta-classifier to yield the final diagnostic decision.

3.5 Classification Models and Training Strategy

3.5.1 Deep Learning Architectures

Two families of CNNs were employed to systematically evaluate the proposed multi-axis fusion strategies: the classical LeNet-5 [40], serving as a well-established baseline, and a custom lightweight architecture inspired by MobileNetV2 [41], hereafter referred to as Mini-MobileNetV2. The latter is specifically tailored for computationally constrained edge environments and resource-constrained condition-monitoring applications. A summary of the computational complexity and architectural specifications across all six model variants is presented in Table 1.

Table 1: Architectural details and computational complexity of the evaluated models under different fusion strategies.[1]

Model Architecture	Fusion Strategy	Total Params	Model Size (KB)	MACs (M)	FLOPs (M)
LeNet-5	Data-Level	21,091	82.4	0.17	0.34
LeNet-5	Feature-Level	77,995	304.7	0.30	0.60
LeNet-5	Decision-Level	63,208	246.9	0.29	0.57
Mini-MobileNetV2	Data-Level	13,091	51.1	1.57	3.14
Mini-MobileNetV2	Feature-Level	50,915	198.9	4.50	9.00
Mini-MobileNetV2	Decision-Level	15,846	61.9	4.50	8.99

[1]MACs (multiply-accumulate operations) and FLOPs (floating-point operations) are reported for a single forward pass with one input sample. Model size represents H5-format storage. The reduction ratio compares the most efficient model (Mini-MobileNetV2 Data-Level) with the most computationally expensive one (Mini-MobileNetV2 Feature-Level). All LeNet-5 variants use tanh activation, while MobileNetV2 variants employ ReLU6 activation with inverted residual blocks (expansion factor = 2). Weight sharing in Decision-Level MobileNetV2 reduces parameters by 65.5% compared to its non-shared counterpart.

The two CNN families were implemented under the same three fusion topologies defined in the preceding subsection. For Data-Level fusion, a single backbone receives the unified $16 \times 16 \times 3$ pseudo-RGB tensor. For Feature-Level fusion, three axis-specific $16 \times 16 \times 1$ branches extract representations independently before concatenation and classification. For Decision-Level fusion, three axis-specific prediction vectors are concatenated and passed to a compact meta-classifier.

The LeNet-5 variants retain the classical two-stage convolution-pooling structure with tanh activations, followed by fully connected layers. In the Data-Level configuration, the flattened feature maps are passed through 120- and 84-neuron dense layers before the Softmax classifier. In the Feature-Level configuration, the three 84-dimensional axis representations are concatenated into a 252-dimensional vector and refined through a 64-neuron dense layer with ReLU activation and 0.4 dropout. In the Decision-Level configuration, the concatenated 9-dimensional probability vector is integrated by the same 64-neuron meta-classifier.

The Mini-MobileNetV2 variants use a lightweight backbone composed of an initial 16-filter 3×3 convolution followed by four inverted residual blocks with expansion factor $t = 2$ and channel depths of 16, 24, 32, and 32. Depthwise separable convolutions, linear bottlenecks, ReLU6 activations, global average pooling, and 0.4 dropout are used to reduce computational cost and overfitting. The Data-Level and Feature-Level configurations feed the resulting representation to a 128-neuron dense layer before classification, whereas the Decision-Level configuration uses a shared-weight backbone for the three axes and combines their probability vectors through a 64-neuron meta-classifier, reducing parameters by 65.5% relative to a non-shared implementation.

3.5.2 Traditional Machine Learning Baselines

To benchmark the proposed deep learning architectures, Support Vector Machine (SVM) [42] and Random Forest (RF) [43] were implemented. For compatibility with these non-convolutional classifiers, each $16 \times 16 \times 3$ input tensor was linearized via row-major flattening into a 768-dimensional feature vector.

- Support Vector Machine (SVM) [42]: Configured with a Radial Basis Function (RBF) kernel to map non-linear decision boundaries. The regularization parameter was set to $C = 1.0$, with the kernel coefficient defined as $\gamma = 1/(n_{\text{features}} \operatorname{Var}(X))$, where $n_{\text{features}} = 768$. This setup adheres to established practices for medium-dimensional feature spaces [42, 44].
- Random Forest (RF) [43]: An ensemble of $N = 100$ bootstrap-aggregated decision trees using the Gini impurity criterion. The maximum number of features evaluated per node split was restricted to $\sqrt{n_{\text{features}}}$ [44]. Final predictions were derived via majority voting, providing inherent regularization against spatial feature redundancy.

To ensure a fair and reproducible comparison, both baselines were implemented via scikit-learn [44] and evaluated using the identical 10-fold file-level stratified cross-validation protocol detailed in Section 3.5.3, strictly preventing segment-level data leakage.

3.5.3 Dataset Partitioning and Leakage-Free Training Strategy

To ensure an unbiased evaluation, dataset partitioning was strictly executed at the file level rather than the individual segment level. This approach is essential for time-series and sensor-based signals to prevent data leakage, where temporal correlations within a single continuous recording could otherwise yield artificially inflated performance metrics.

By using stratified group k-fold cross-validation, the study ensured that all segments from a specific raw signal recording were assigned exclusively to the training, validation, or testing sets. No overlapping segments from a single file appeared in both the training and evaluation phases, forcing the models to learn generalizable features rather than memorize file-specific signatures. The dataset was partitioned as follows:

- File-Level Stratification: The raw recordings were divided into $k = 10$ folds, maintaining proportional representation of the "Undamaged", "Partially Worn", and "Heavily Worn" conditions in each fold.
- Hold-out Testing: For each iteration, we strictly reserved a subset of files for unbiased final testing, while using the remaining files for training and validation.
- Temporal Integrity: By maintaining the integrity of each file's segments during the split, the model's performance reflects its capability to identify tool conditions on entirely unseen machinery recordings.

Following the partitioning phase, the investigated architectures were optimized using the Adam optimizer with an initial learning rate of 0.001 and a categorical cross-entropy loss function. We trained each model with a batch size of 32 for up to 100 epochs, using early stopping based on validation loss to mitigate overfitting. Furthermore, the networks used rectified linear unit (ReLU) activation functions and Glorot uniform weight initialization to ensure stable, rapid convergence across all validation folds.

3.6 Robustness Evaluation and Validation Metrics

3.6.1 Robustness Evaluation Framework

In practical industrial applications, intelligent fault diagnosis systems may be exposed to unconstrained environments and hardware-related disturbances. Rather than using rigid, specialized mechanical fixtures, this study records vibration data by placing the mobile device directly on the machine surface. This fixtureless approach was adopted to expose the learning models to realistic workshop variability. In addition, we applied simulated perturbations to the test data to evaluate model performance under more severe non-ideal operating conditions.

To investigate the robustness of the proposed multi-level fusion frameworks, we designed and injected two structurally distinct sensor fault scenarios into the test dataset: multi-axis sensor saturation and dynamic sensor misalignment.

Sensor Saturation Fault Simulation In a fixtureless configuration, the vibrations and cyclic forces generated during drilling may cause the smartphone to momentarily lose solid contact or experience secondary impacts against the machine surface. These mechanical overloads can generate acceleration spikes that exceed the smartphone's MEMS accelerometer dynamic range. Consequently, the sensor's analog-to-digital converter (ADC) may saturate, reporting flattened plateaus rather than the actual harmonic peaks of the vibration signal.

To model this nonlinear physical behavior, a dynamic multi-parameter hard-limiting operator was developed. The fault-injection pipeline modifies the raw test signals in three stages:

- Temporal windowing: A specific vulnerable window within the signal duration (seq_len) is isolated using a start-position parameter (Start_Position = 0.20) and a duration factor (Saturation_Ratio = 0.25). This isolates a window spanning 25% of the total segment length, starting after the initial 20% of the sample.
- Impact multiplier: Raw values within this time frame are scaled by an intensity factor (Intensity_Factor = 2 and 3) to simulate severe mechanical shock overloads that push the signal amplitude beyond the sensor's physical bounds.
- Signal clipping: Because vibration amplitudes vary across drilling conditions (e.g., healthy, moderately worn, and severely worn classes), implementing a rigid static threshold could introduce artifacts. Thus, the saturation ceiling ($X_{\max}$) is computed dynamically as the absolute maximum value recorded in the corresponding unperturbed axis of that specific recording file. The clipped output $\tilde{x}$ is defined via a dynamic hard-limiter [45]:

$$\tilde{x} = \begin{cases} X_{\max}, & \text{if } x\mu > X_{\max}, \\ x\mu, & \text{otherwise}, \end{cases} \tag{8}$$

where μ represents the intensity factor. This clipping is applied to the Z-axis (the primary direction of drill bit penetration and thrust-force loading) before the time-series sequence is transformed into a 16×16 image grid. This causes texture degradation and loss of high-frequency spatial patterns in that channel.

Sensor Misalignment Simulation When a sensing device is placed freely on a machine without a rigid fixture, multi-directional forces and mechanical shocks can cause sensor misalignment. The smartphone may not maintain its initial orientation and can undergo micro-rotations, sliding, or angular drift across the machine surface. This physical displacement alters the sensor's spatial orientation relative to the machine's true reference frame. Consequently, the accelerometer may register cross-axis energy leakage rather than independent orthogonal accelerations, altering the contrast and structural texture of the generated 2D fusion grids.

Following established benchmarks in deep-learning-based structural health monitoring, this unconstrained orientation shift was simulated by virtually rotating the original 3-axis measurement matrix S using a 3D rotation matrix R. The perturbed measurement matrix $\tilde{S}$ from the misaligned sensor is determined by:

$$\tilde{S} = SR \tag{9}$$

The rotation matrix R is parameterized by the standard three-dimensional Euler angles: yaw (ψ), pitch (θ), and roll (ϕ), which define rotations around the Z, Y, and X axes, respectively. The composite transformation matrix is expressed as [46]:

$$R = R_z(\psi)R_y(\theta)R_x(\phi) = \begin{bmatrix} \cos\psi\cos\theta & \cos\psi\sin\theta\sin\phi - \sin\psi\cos\phi & \cos\psi\sin\theta\cos\phi + \sin\psi\sin\phi \\ \sin\psi\cos\theta & \sin\psi\sin\theta\sin\phi + \cos\psi\cos\phi & \sin\psi\sin\theta\cos\phi - \cos\psi\sin\phi \\ -\sin\theta & \cos\theta\sin\phi & \cos\theta\cos\phi \end{bmatrix} \tag{10}$$

Because the smartphone was placed directly on the machine surface, rotation within the contact plane is a plausible source of sensor misalignment. Therefore, this study investigates rotational perturbations within the XY plane, where sliding and angular drift are most likely, to quantify the fusion networks' robustness and stability against unconstrained orientation shifts.

3.6.2 Validation Metrics

To quantitatively evaluate model performance, overall accuracy and the class-based F1-score were adopted. Overall accuracy provides a global performance measure for the multiclass classification tasks and is defined as:

$$\text{Accuracy} = \frac{\sum_{i=1}^{C} TP_i}{\sum_{i=1}^{C}(TP_i + FN_i)} \tag{11}$$

where C denotes the number of classes, and TP_i and FN_i correspond to the counts of true-positive and false-negative predictions for class i, respectively.

To account for potential minor class imbalances, the F1-score was computed individually for each class c, representing the harmonic mean of precision and recall:

$$F1_c = \frac{2\,\text{precision}_c\,\text{recall}_c}{\text{precision}_c + \text{recall}_c} \tag{12}$$

where precision and recall for class c are given by:

$$\text{precision}_c = \frac{TP_c}{TP_c + FP_c}, \qquad \text{recall}_c = \frac{TP_c}{TP_c + FN_c} \tag{13}$$

In these equations, TP_c, FP_c, and FN_c represent the numbers of true-positive, false-positive, and false-negative predictions for class c, respectively.

4 Results and Discussion

This section evaluates the empirical efficacy of the proposed smartphone-based TCM framework as a proof of concept. To investigate whether the convolutional networks extract meaningful machining dynamics rather than arbitrary environmental noise, the evaluation begins with a spectral signature analysis (Section 4.1). This analysis explores the spectral footprint of the undersampled signals to examine whether wear-related high-frequency dynamics may appear as discriminative low-frequency signatures.

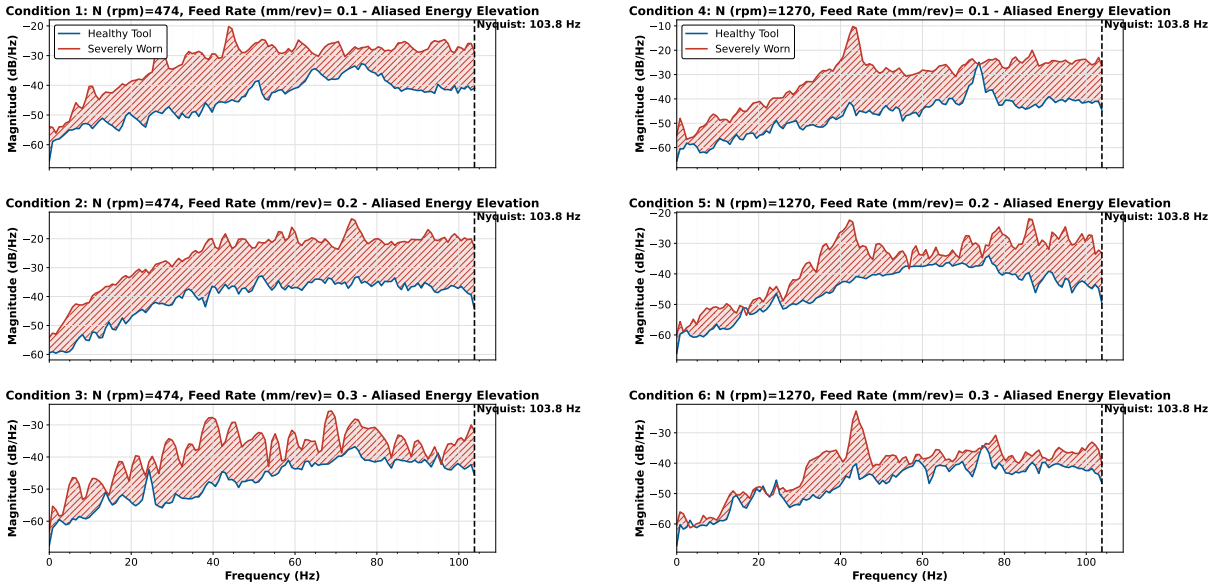


Figure 6: Comparative power spectral density (PSD) analysis of Z-axis vibrations under six cutting conditions. The left column corresponds to a spindle speed of $N = 474$ rpm with feed rates of 0.1, 0.2, and 0.3 mm/rev (Conditions 1–3), while the right column corresponds to $N = 1270$ rpm with the same feed rates (Conditions 4–6). Shaded regions indicate spectral energy elevation that separates healthy and severely worn tools across the investigated cutting conditions.

Following this initial spectral analysis, the experimental evaluation addresses three main tasks. First, the models are evaluated for (1) drilling phase identification, which distinguishes active cutting from non-drilling states, and (2) discrete tool wear classification. Within these tasks, the diagnostic performance of the lightweight Mini-MobileNetV2 is benchmarked against a standard LeNet-5 baseline and traditional ML classifiers (SVM and Random Forest), while three distinct multi-axis fusion strategies (Data-Level, Feature-Level, and Decision-Level) are systematically compared using the selected performance metrics. Next, the study evaluates robustness by subjecting the multi-axis fusion frameworks to simulated physical perturbations, specifically multi-axis sensor saturation and dynamic angular misalignment.

4.1 Frequency-Domain Analysis of Wear-Related Spectral Patterns

Before evaluating the classification performance of the CNN models, a frequency-domain analysis was conducted to examine whether the bandwidth-limited smartphone signals contain wear-related spectral patterns. Welch's method was used to estimate the PSD of the recorded vibrations.

Figure 6 presents a comparative PSD analysis of the Z-axis (thrust direction) signals, contrasting a healthy tool (blue line) against a severely worn tool (red line). The analysis spans two spindle speeds ($N = 474$ rpm and $N = 1270$ rpm) and three feed rates to examine whether tool-condition effects remain visible across operating conditions.

As shown across all subplots in Figure 6, the healthy tool's frequency spectrum remains relatively stable and low-energy across the observable baseband. In contrast, the severely worn tool shows

a clear, consistent broadband energy elevation. This is characterized by an upward shift of 10 to 20 dB/Hz across almost the entire spectrum up to the Nyquist frequency. The red hatched area visually represents this accumulated energy. This spectral gap between the healthy and worn states remains visible across the investigated spindle speeds and feed rates.

From a physical machining perspective, primary acoustic emissions and vibrations from tool wear typically occur in the kilohertz range. In this smartphone-based setup, the absence of an ideal analog anti-aliasing filter may allow some high-frequency components to appear within the low-frequency baseband. While isolating the exact contribution of aliased energy from other low-frequency dynamics is difficult, the persistence of this energy shift across cutting parameters suggests that it is a useful diagnostic feature. The observed broadband elevation is consistent with the aliasing hypothesis, although structural resonances and increased stochastic vibration under degraded cutting conditions may also contribute.

4.2 Drilling Phase Identification Results

To evaluate the system's capability in isolating the drilling phase from non-drilling operations, the proposed architectures were benchmarked across all fusion strategies against traditional machine learning baselines (SVM and RF). As summarized in Table 2, Mini-MobileNetV2 with Data-Level fusion emerged as the best-performing configuration, achieving a peak testing accuracy of 91.74% and an F1-score of 90.67%. It outperformed both its LeNet-5 counterpart (79.35% accuracy) and the RF baseline (88.53% accuracy). While Feature-Level fusion yielded the lowest validation loss for Mini-MobileNetV2, Data-Level fusion maximized overall recall, making it the most reliable configuration for this specific task.

The performance gap between the CNN models and the traditional baselines also highlights the impact of the signal-imaging strategy. In the CNN configurations, the transformed pseudo-RGB inputs preserve local spatial textures and inter-axis patterns that can be learned through convolutional filters. By contrast, SVM and RF operate on flattened 768-dimensional vectors, which retain the same numerical samples but largely discard the two-dimensional neighborhood structure introduced by imaging. This difference suggests that the main benefit of signal imaging is not only dimensional formatting, but also its ability to expose structured spatial representations that are better exploited by lightweight CNN architectures.

The confusion matrices (Figure 7) further illustrate the inter-class behavior. LeNet-5 showed more errors for the "non-drilling" class, generating 46 false positives (misclassified as drilling). In contrast, Mini-MobileNetV2 reduced these false positives to 10. This reduction highlights Mini-MobileNetV2's effectiveness in handling class imbalance, supporting its suitability for practical, unconstrained monitoring scenarios.

Further error analysis indicated that signal transformations (e.g., time stretching) and manual labeling introduced uncertainty during high-speed operational transitions. As depicted in the Z-axis spectrogram (Figure 8), these mechanical transitions manifest as spectral boundary "blurring". This temporal smearing likely encouraged the LeNet-5 model to learn generalized, less distinctive spatial patterns, increasing its error rate. Conversely, Mini-MobileNetV2 showed greater capacity to handle these noisy labels and transition ambiguities, supporting its robustness for practical applications where perfect temporal annotation of fixtureless sensor data is difficult.

Table 2: Comparison of drilling phase detection performance for LeNet-5 and Mini-MobileNetV2 under Data-Level, Feature-Level, and Decision-Level fusion, together with SVM and Random Forest baselines. Values report mean ± standard deviation; Mini-MobileNetV2 with Data-Level fusion achieved the highest accuracy and F1-score, indicating the strongest overall ability to separate active drilling from non-drilling states.

Fusion Method	**Model**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
		Mean ± Std			
Data-Level	LeNet-5	79.35±5.24	76.96±6.12	73.51±5.88	74.73±6.01
	Mini-MobileNetV2	**91.74±3.12**	90.00±3.45	91.47±2.98	**90.67±3.21**
Feature-Level	LeNet-5	77.88±4.83	78.43±5.16	77.16±4.92	77.79±5.01
	Mini-MobileNetV2	87.02±3.75	87.64±3.97	86.51±3.82	87.07±3.85
Decision-Level	LeNet-5	80.24±5.94	80.04±6.32	72.43±5.72	74.39±6.18
	Mini-MobileNetV2	84.37±4.14	82.03±4.43	81.87±4.31	81.95±4.42
Not applicable	SVM	84.86±6.85	86.24±6.02	85.32±4.92	84.55±6.19
Not applicable	Random Forest	88.53±7.53	90.59±6.12	90.14±5.81	88.92±6.77

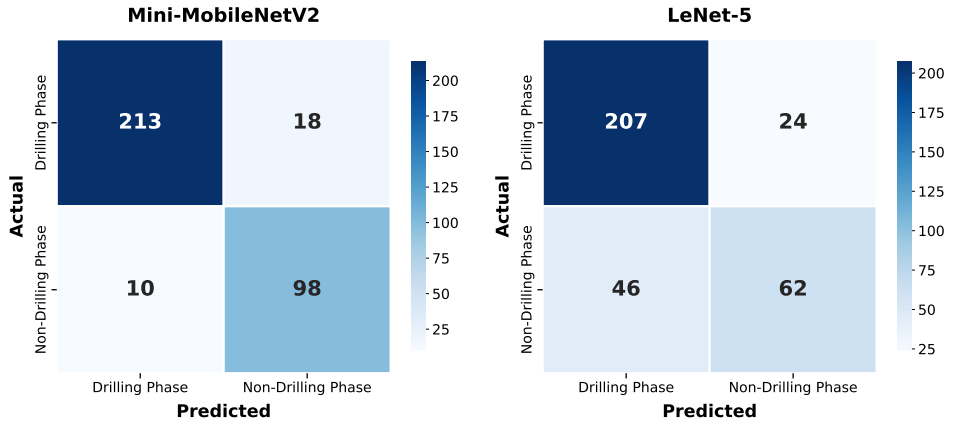


Figure 7: Confusion matrices for drilling phase identification.

4.3 Tool Wear Classification Results

A secondary set of experiments was conducted to evaluate the framework's capacity to classify progressive tool wear, categorizing the cutting state into three balanced classes: "Undamaged", "Partially Worn", and "Heavily Worn". A comprehensive comparison of all experimental configurations is provided in Table 3. Unlike the drilling phase identification task, the Feature-Level fusion strategy paired with the Mini-MobileNetV2 architecture emerged as the best-performing configuration for tool wear classification. This model achieved an overall accuracy of 89.68% and an F1-score of 89.67%, outperforming not only the LeNet-5 variants (which remained below 80%) but also the RF baseline (87.94%).

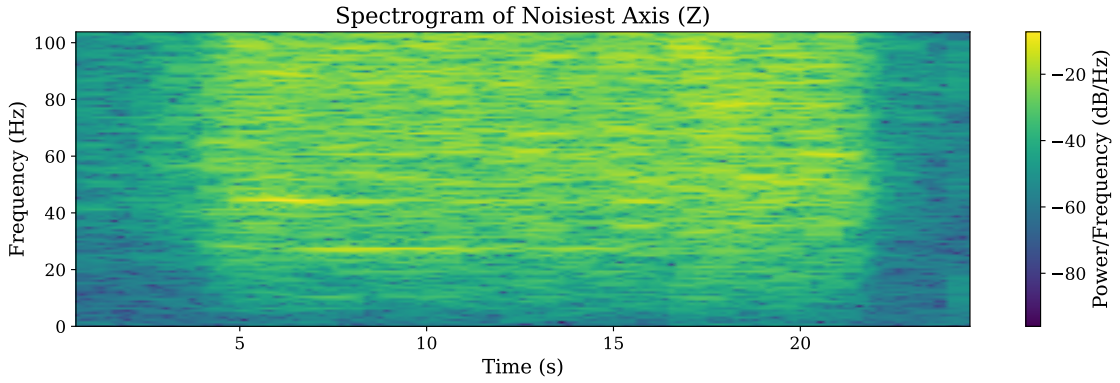


Figure 8: Z-axis spectrogram showing boundary blurring, which complicates precise manual labeling of transition states.

The performance discrepancy across fusion strategies, particularly the lower performance of Data-Level and Decision-Level approaches, can be interpreted by examining the raw multi-axis vibration signatures. As illustrated in Figure 9, during the drilling phase, the Z-axis (thrust direction) experiences the highest cutting forces, containing sharp, high-amplitude wear-related signatures (red points). Conversely, the X and Y axes often capture secondary vibrations mixed with environmental noise and machine-specific resonances.

In the Data-Level approach, concatenating these axes at the raw input stage forces the network to process the dominant Z-axis signals alongside lateral noise, potentially blurring the subtle temporal textures of progressive wear. Decision-Level fusion may also be affected when predictions from noisier axes contribute to the final decision, allowing axis-specific errors to propagate.

In contrast, Feature-Level fusion acts as a structural filter. By allocating independent convolutional branches to each axis, the Mini-MobileNetV2 extracts localized, axis-specific features before concatenation. This allows the network to isolate the dominant wear signatures from the Z-axis while mitigating cross-channel noise interference, leading to more accurate and stable classifications. LeNet-5's lower performance across fusion levels suggests that MobileNetV2's inverted residual blocks are better suited to learning the complex textures associated with progressive wear in smartphone-acquired data.

To further elucidate the internal decision-making process and visually validate the filtering capability of the proposed Feature-Level fusion, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed. Figure 10 presents a diagnostic interpretation of the model's focus across three tool-wear conditions: Undamaged, Partially Worn, and Heavily Worn.

As depicted in the combined Grad-CAM heatmaps (Figure 10c), the network does not allocate its attention randomly across the signal matrix. Instead, it learns to localize its focus on specific, high-intensity textural patterns. Comparing heatmaps across different wear states shows that the model dynamically shifts its focal intensity to isolate wear-related anomalies while suppressing background noise from the lateral axes. This visual evidence supports the hypothesis that Feature-Level fusion acts as an intelligent structural filter, ensuring classification is driven by physically meaningful diagnostic features rather than spurious machine noise.

Overall, these results extend beyond identifying a single optimal model; they support a low-cost

Table 3: Comparison of tool wear classification performance for LeNet-5 and Mini-MobileNetV2 under Data-Level, Feature-Level, and Decision-Level fusion, together with SVM and Random Forest baselines. Values report mean ± standard deviation for the three-class wear classification task involving undamaged, partially worn, and heavily worn drills; Mini-MobileNetV2 with Feature-Level fusion achieved the highest accuracy and F1-score, indicating the strongest overall ability to distinguish progressive tool wear states.

Fusion Method	**Model**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
		Mean ± Std			
Data-Level	LeNet-5	77.82±5.18	74.41±6.07	71.84±5.76	72.13±5.94
	Mini-MobileNetV2	85.09±4.87	83.03±4.43	83.72±4.18	83.26±4.39
Feature-Level	LeNet-5	76.58±4.53	73.42±4.77	74.48±4.19	73.07±4.13
	Mini-MobileNetV2	**89.68±3.38**	89.91±3.14	89.45±2.97	**89.67±3.52**
Decision-Level	LeNet-5	79.18±6.47	76.83±7.14	70.47±6.04	71.88±6.32
	Mini-MobileNetV2	82.17±3.82	78.71±4.46	76.44±3.47	76.29±3.58
Not applicable	SVM	84.04±7.07	85.55±6.26	84.72±5.18	83.41±6.54
Not applicable	Random Forest	87.94±7.91	90.03±6.29	89.70±6.07	88.20±7.11

monitoring paradigm. By showing that a smartphone–placed freely on a machine without rigid fixtures–can achieve nearly 90% classification accuracy, this study supports the feasibility of low-cost, fixtureless diagnostic sensing. It suggests that deep learning architectures can extract consistent and physically meaningful patterns from bandwidth-limited smartphone vibration data.

4.4 Robustness Against Sensor Saturation

To evaluate the framework's resilience against simulated mechanical overloads and impact-induced clipping, the dynamic hard-limiting operator (detailed in Section 3.6.1) was applied to the Z-axis at intensity factors of $\mu = 2$ and $\mu = 3$. This simulates unidirectional thrust impacts while leaving the lateral measurements (X and Y axes) unchanged. The corresponding percentage drop in classification accuracy is summarized in Figure 11.

At the moderate saturation level ($\mu = 2$), all deep learning models remained stable, with accuracy drops remaining well below 3%. However, the stronger saturation scenario ($\mu = 3$) revealed performance differences associated with the chosen multi-axis fusion strategy.

As illustrated, Data-Level fusion architectures and traditional baselines (SVM and RF) experienced larger performance degradation, with accuracy drops ranging from 13.67% to 15.65%. Because Data-Level fusion concatenates the tri-axial signals into a unified spatial tensor before feature extraction, clipping and texture loss from the Z-axis can distort early-stage spatial convolutions. Consequently, the networks struggled to disentangle the flattened Z-axis plateaus from healthy lateral dynamics.

In contrast, Feature-Level and Decision-Level fusion strategies demonstrated stronger fault tolerance. Notably, MobileNetV2 Decision-Level fusion and LeNet-5 Feature-Level fusion restricted

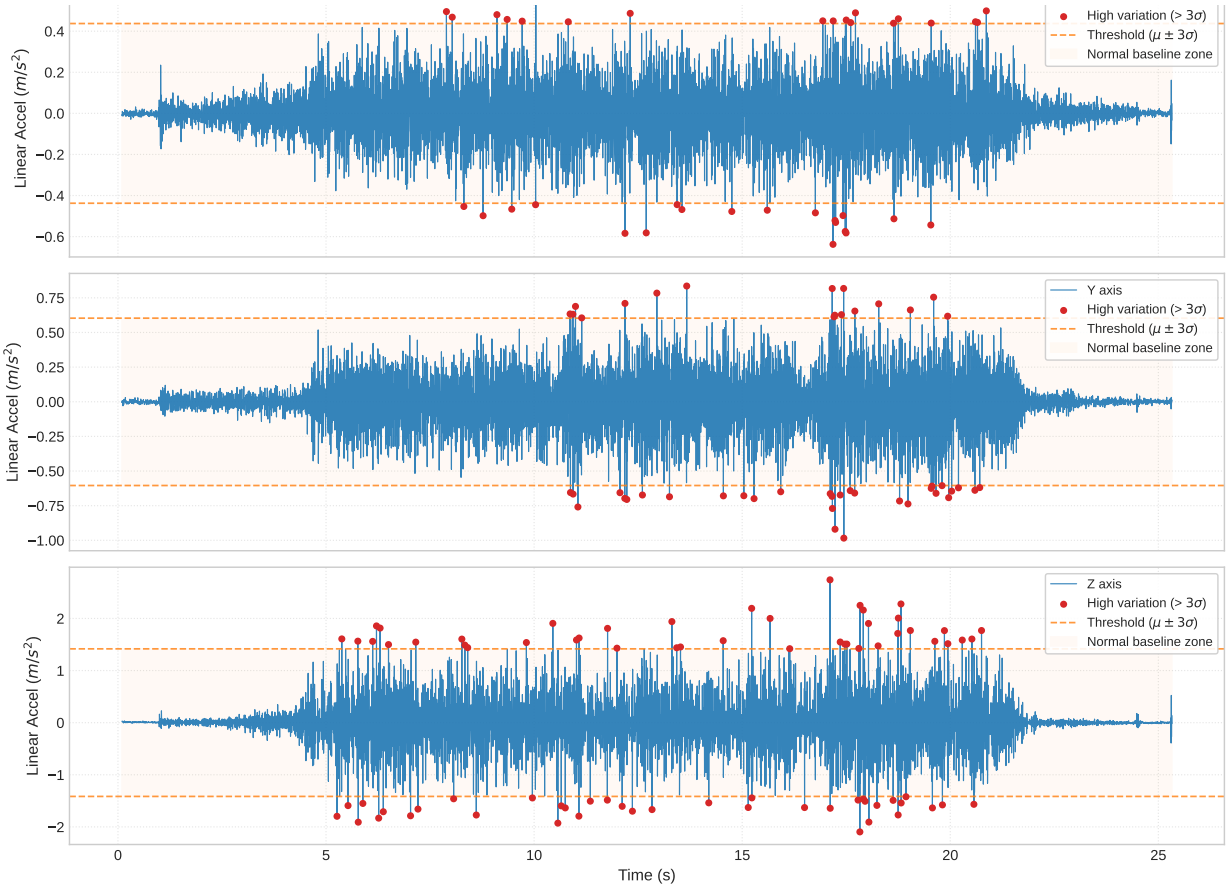


Figure 9: Time-domain representation of raw linear acceleration signals across the X, Y, and Z axes. The dashed orange lines define the statistical Three-Sigma threshold baseline, effectively establishing the normal variance boundary. Red markers highlight localized, axis-specific high-variation transients and disturbances that exceed this threshold.

their performance drops to 4.53% and 4.69%, respectively, under the stronger saturation condition. Because the simulated clipping was limited to the Z-axis, these multi-branch architectures effectively compartmentalized the signal degradation. This structural isolation allowed the networks to partly compensate for the degraded thrust-direction information by relying on the less-affected spatial patterns preserved within the independent X and Y branches.

These findings suggest that, in fixtureless environments where momentary single-axis sensor saturation may occur, independent-branch fusion strategies can improve robustness by limiting the propagation of corrupted axis-specific information.

4.5 Robustness Against Sensor Misalignment

To evaluate the framework's stability under unconstrained, fixtureless operating conditions, the models were subjected to simulated physical displacements. As detailed in Section 3.6.1, dynamic orientation shifts were induced by mathematically rotating the tri-axial measurement matrix within the XY contact plane–a plausible degree of freedom for physical sliding and angular rotation.

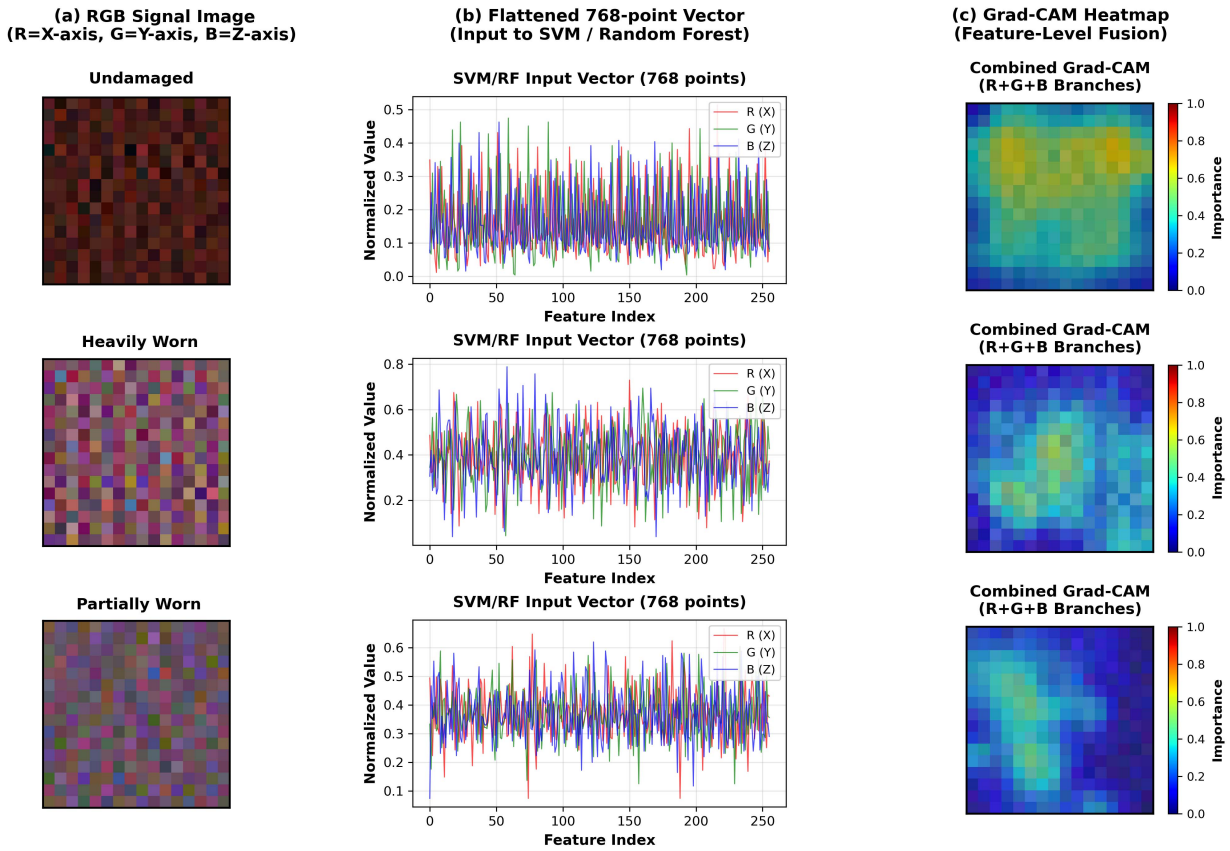


Figure 10: Visual interpretability of the Feature-Level fusion model across three tool wear conditions. (a) RGB representation of the tri-axial vibration signals (X, Y, and Z axes mapped to R, G, and B channels, respectively). (b) Corresponding flattened 768-point signal vectors illustrating amplitude variations. (c) Combined Grad-CAM heatmaps demonstrating the spatial regions of highest feature importance (warm colors) that drive the network's classification decisions.

The performance degradation across increasing rotational severities (15° to 60°) is presented in Figure 12.

The experimental results indicate an architectural vulnerability in early-fusion strategies. Under a 60° misalignment, both LeNet-5 and MobileNetV2 architectures employing Data-Level fusion showed substantial accuracy drops of 70.00% and 61.76%, respectively. Traditional SVM classifiers showed a similar trend (52.35% drop). This degradation can be attributed to angular rotation, which geometrically induces cross-axis energy leakage between the X and Y channels. In Data-Level fusion, where raw temporal signals are directly concatenated into a unified spatial grid before convolution, this leakage distorts the learned 2D structural textures and reduces the effectiveness of the learned spatial filters.

Conversely, the results highlight the resilience of late-fusion and multi-branch architectures. Specifically, the MobileNetV2 equipped with Decision-Level fusion (utilizing Soft Weighting) emerged as the most robust configuration, exhibiting minimal degradation (maximum drop of 0.59%) even under a 60° rotation.

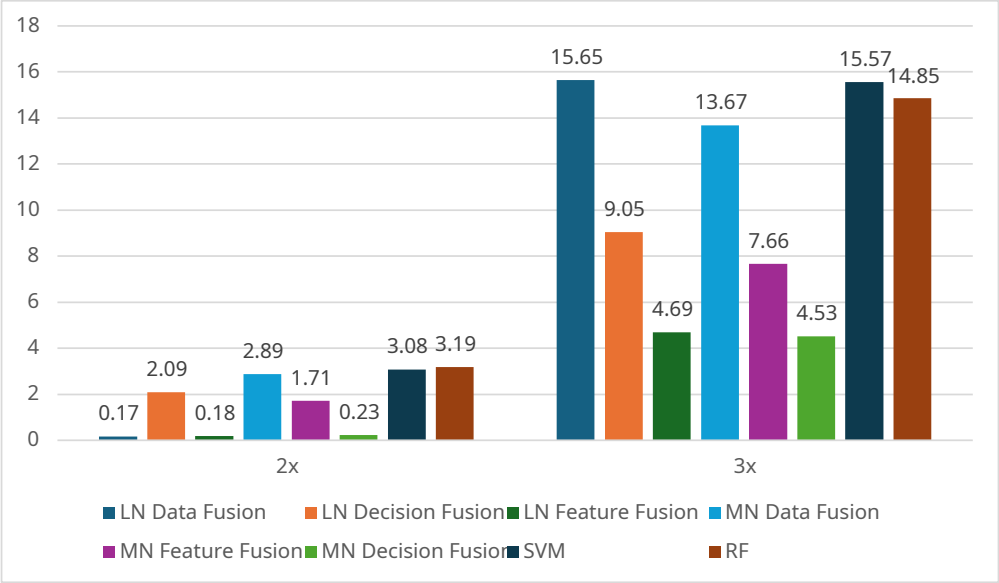


Figure 11: Percentage drop in classification accuracy under simulated sensor saturation on the Z-axis ($\mu = 2$ and $\mu = 3$). Abbreviations denote the evaluated classifiers: LN = LeNet-5, MN = Mini-MobileNetV2, SVM = Support Vector Machine, and RF = Random Forest.

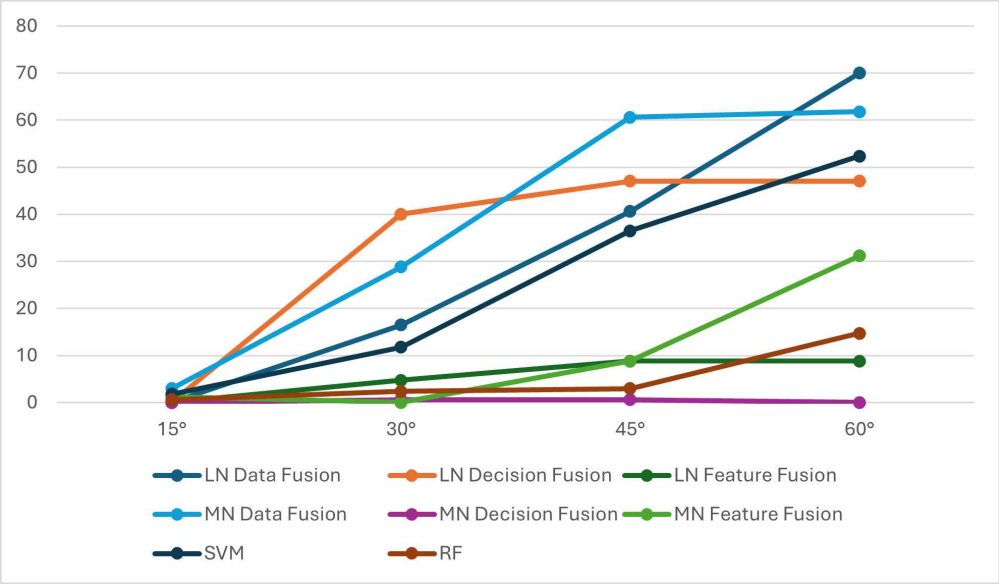


Figure 12: Percentage drop in classification accuracy under progressive sensor misalignment in the XY plane. Abbreviations denote the evaluated classifiers: LN = LeNet-5, MN = Mini-MobileNetV2, SVM = Support Vector Machine, and RF = Random Forest.

This strong fault tolerance stems from the physical structure of the simulated perturbation and the model's architectural isolation. Because the simulated misalignment occurs within the XY plane, the orthogonal Z-axis (capturing the primary thrust and penetration forces) remains physically unperturbed. By dedicating an independent classification pathway to each axis, the Decision-Level MobileNetV2 framework effectively isolates the Z-axis features. During the final ensemble

voting phase, the soft-weighting mechanism can suppress the distorted probabilistic outputs from the misaligned X and Y branches, compensating for them by relying more strongly on the stable predictions of the unperturbed Z-axis.

These findings indicate that while Data-Level fusion may offer advantages under unperturbed conditions, Decision-Level and independent Feature-Level fusion can provide greater robustness under fixtureless deployment scenarios involving sensor misalignment.

Overall, the results reveal a task-dependent tradeoff among fusion strategies. Data-Level fusion provided the strongest drilling phase identification performance, indicating that early cross-axis integration is beneficial when the objective is to separate active cutting from non-drilling states. For tool wear classification and robustness evaluation, however, Feature-Level and Decision-Level fusion were more reliable because they preserved axis-specific information and limited the influence of corrupted channels. These findings suggest that the optimal fusion topology depends not only on baseline classification accuracy, but also on the expected level of sensor disturbance in fixtureless smartphone-based monitoring.

5 General Discussion, Limitations, and Future Work

This study presents a proof-of-concept (PoC) demonstrating the feasibility of using a consumer-grade smartphone as an edge-computing sensing node for tool condition monitoring. Traditional industrial setups typically rely on high-frequency piezoelectric accelerometers and rigid data acquisition (DAQ) hardware. In contrast, empirical findings suggest that a mobile device's built-in MEMS sensors can capture diagnostically meaningful machining dynamics under specific conditions. A plausible, though not yet fully isolated, explanation for this capability is aliasing. Although the smartphone's sampling rate is inherently limited, high-frequency energy generated by tool wear may partially project into the observable sub-Nyquist baseband. This mechanism, together with low-frequency machine dynamics and structural resonances, may explain how the proposed learning pipeline extracts discriminative diagnostic features from signals that would traditionally be considered severely bandwidth-limited.

5.1 Architectural Complexity and Task Dimensionality

A key observation from the comparative analysis is the need to align model complexity with the specific diagnostic task. For identifying transient drilling phases, where the signal shows broad, high-amplitude shifts, the custom Mini-MobileNetV2's hierarchical feature extraction performed best.

Conversely, the discrete classification of progressive tool wear presented a different challenge. Tool degradation manifests as subtle cross-axis temporal textures rather than simple amplitude spikes. In this context, the inverted residual blocks of MobileNetV2 were better suited to decoding these intricate patterns. Simpler architectures, such as the LeNet-5 baseline, struggled to generalize in the multi-class wear estimation task, indicating that the captured low-frequency wear signatures require a sufficiently expressive feature space to be effectively resolved.

The comparison with SVM and Random Forest further emphasizes the importance of the signal-

imaging step. The pseudo-RGB representation does more than reshape the vibration sequence; it preserves local neighborhoods, texture-like variations, and inter-axis spatial relationships that can be exploited by convolutional filters. In contrast, the traditional baselines operate on flattened feature vectors, retaining the same sample values but losing the explicit two-dimensional neighborhood structure. Therefore, the observed CNN advantage reflects both architectural capacity and the compatibility between signal imaging and convolution-based feature extraction.

5.2 Sensor Fusion Strategies in Unconstrained Environments

The experimental results highlight the critical role of multi-axis fusion strategies, particularly in fixtureless environments where the sensing device is unconstrained. The optimal fusion topology was task-dependent. Data-level fusion provided the strongest drilling-phase identification performance, suggesting that early cross-axis integration is useful when the primary objective is to separate active cutting from non-drilling states. For tool wear classification, however, Feature-Level fusion provided the best balance between information sharing and axis isolation. By extracting features from each axis independently before concatenation, it preserved wear-sensitive patterns from the dominant thrust direction while reducing the influence of lateral noise.

The robustness analyses further clarify this tradeoff. Under ideal or mildly perturbed conditions, early fusion can capture inter-axis spatial correlations effectively. However, simulated fault injections, including sensor saturation and dynamic misalignment, revealed this approach's structural vulnerabilities. When the Z-axis experienced severe saturation, or when simulated rotation in the XY plane caused cross-axis energy leakage, Data-Level architectures exhibited significant accuracy drops. Because the raw signals are concatenated before feature extraction, noise or distortion from a single axis directly degrades the unified spatial input. Feature-Level and Decision-Level fusion were more resilient because they limited the propagation of corrupted-axis information, making them more appropriate when physical disturbance is expected in fixtureless smartphone-based monitoring.

5.3 Limitations and Future Directions

As a proof of concept, this framework operates within specific boundaries that must be acknowledged. From a sensing perspective, the empirical validation was conducted using a single smartphone model on a specific drilling machine. While this demonstrates practical feasibility and the empirical benefits of multi-branch fusion, the exact physical mechanisms underlying the hypothesized aliasing contribution require rigorous signal-processing validation using parallel high-frequency reference sensors. The system's generalizability across diverse mobile devices, MEMS noise floors, machine-tool platforms, and cutting conditions also requires broader investigation.

From a labeling perspective, we discretized progressive tool wear into balanced categories based on offline optical inspection to support supervised learning. Although this provided a functional ground truth for classification, real-world tool degradation is a continuous process governed by standardized physical metrics, such as flank wear width. Relying on discrete classes therefore simplifies the underlying wear continuum and limits direct translation to threshold-based industrial maintenance decisions.

From a signal-interpretation perspective, the captured low-frequency signals provide a useful proxy for overall state classification, but the extreme undersampling inherently masks precise spectral harmonics. Consequently, this approach cannot currently resolve specific micro-failure modes, such as distinguishing between thermal cracking and micro-chipping. Practical deployment also poses packaging challenges because consumer smartphones lack the IP-rated ruggedness required for long-term exposure to the thermal, fluidic, and chip-contaminated conditions of typical CNC environments.

Future research will focus on scaling this concept into a more generalized edge-AI solution. This includes testing the framework across heterogeneous devices and transitioning from discrete classification to continuous wear-regression models that align with standardized industrial wear metrics. Future work will also include quantizing the trained Mini-MobileNetV2 models into mobile-optimized formats (e.g., TensorFlow Lite) for real-time, on-device inference and exploring semi-supervised learning methods to reduce reliance on meticulously labeled offline data.

6 Conclusion

This work establishes the feasibility of smartphone-based vibration sensing for low-cost, fixtureless tool condition monitoring in drilling operations. Using only the built-in MEMS accelerometer of a consumer mobile device, the proposed framework provides an accessible alternative to conventional monitoring systems that depend on high-frequency industrial sensors and dedicated data acquisition hardware. The results indicate that, despite severe bandwidth limitations, smartphone-acquired vibration signals can retain diagnostically useful information for drilling phase identification and discrete tool wear classification when suitable signal representations and learning architectures are used.

The experimental results highlight the importance of matching model architecture and fusion strategy to the diagnostic task. Mini-MobileNetV2 with Data-Level fusion achieved the strongest drilling-phase identification performance, showing that early cross-axis integration effectively separates active cutting from non-drilling states. For tool wear classification, Mini-MobileNetV2 with Feature-Level fusion achieved the best performance, indicating that axis-specific feature extraction better resolves progressive wear patterns. The comparison with SVM and Random Forest further supports the value of signal imaging, because CNNs can exploit local spatial neighborhoods and inter-axis texture patterns that are not explicitly preserved after flattening.

The robustness analyses also show that fusion topology is critical in unconstrained deployments. Data-Level fusion was more vulnerable to simulated sensor saturation and angular misalignment because corrupted axis information directly affected the unified input representation. In contrast, Feature-Level and Decision-Level fusion improved robustness by limiting the propagation of axis-specific disturbances. These findings support the feasibility of smartphone-based, fixtureless monitoring while emphasizing that practical deployment requires careful fusion design.

Future work should validate the aliasing hypothesis using synchronized high-frequency reference sensors, evaluate the framework across additional smartphones and machine-tool platforms, and transition from discrete wear classification to continuous wear-regression models aligned with standardized industrial metrics. Quantizing the trained Mini-MobileNetV2 models into mobile-optimized formats, such as TensorFlow Lite, will also be necessary for real-time, on-device inference in prac-

tical edge-AI applications.

Statements and Declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgment

During the preparation of this work, the authors used OpenAI's ChatGPT-5 in order to improve the language and address potential grammar errors. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] Keller, A. Z., Kamath, A. R. R., and Perera, U. D., "Reliability analysis of CNC machine tools", Reliab. Eng., 3(6), pp. 449–473 (1982).

[2] Ye, Y., Hu, T., Zhang, C., and Luo, W., "Design and development of a CNC machining process knowledge base using cloud technology", Int. J. Adv. Manuf. Technol., 94(9–12), pp. 3413–3425 (2018).

[3] Wan, S., Li, D., Gao, J., and Li, J., "A knowledge based machine tool maintenance planning system using case-based reasoning techniques", Robot. Comput.-Integr. Manuf., 58, pp. 80–96 (2019).

[4] Joshi, A. V., Machine Learning and Artificial Intelligence, Springer International Publishing, Cham (2020).

[5] Soori, M., Arezoo, B., and Dastres, R., "Machine learning and artificial intelligence in CNC machine tools, A review", Sustain. Manuf. Serv. Econ., p. 100009 (2023).

[6] Leibner, P., Von Stillfried, A., Hampel, F., and Schindler, C., "Review of Smartphone Accelerometers for Usage in Research and Engineering", IEEE Instrum. Meas. Mag., 28(7), pp. 44–51 (2025).

[7] Yang, C., Shen, W., and Wang, X., "The Internet of Things in Manufacturing: Key Issues and Potential Applications", IEEE Syst. Man Cybern. Mag., 4(1), pp. 6–15 (2018).

[8] Siderska, J. and Jadaan, K. S., "Cloud manufacturing: a service-oriented manufacturing paradigm. A review paper", Eng. Manag. Prod. Serv., 10(1), pp. 22–31 (2018).

[9] Lee, J., Davari, H., Singh, J., and Pandhare, V., "Industrial Artificial Intelligence for industry 4.0-based manufacturing systems", Manuf. Lett., 18, pp. 20–23 (2018).

[10] Nasir, V. and Sassani, F., "A review on deep learning in machining and tool monitoring: methods, opportunities, and challenges", Int. J. Adv. Manuf. Technol., 115(9–10), pp. 2683–2709 (2021).

[11] Wang, B. and Liu, Z., "Influences of tool structure, tool material and tool wear on machined surface integrity during turning and milling of titanium and nickel alloys: a review", Int. J. Adv. Manuf. Technol., 98(5–8), pp. 1925–1975 (2018).

[12] Yeganefar, A., Niknam, S. A., and Asadi, R., "The use of support vector machine, neural network, and regression analysis to predict and optimize surface roughness and cutting forces in milling", Int. J. Adv. Manuf. Technol., 105(1–4), pp. 951–965 (2019).

[13] Nasir, V., Mohammadpanah, A., and Cool, J., "The effect of rotation speed on the power consumption and cutting accuracy of guided circular saw: Experimental measurement and analysis of saw critical and flutter speeds", Wood Mater. Sci. Eng., 15(3), pp. 140–146 (2020).

[14] Nasir, V. and Cool, J., "Intelligent wood machining monitoring using vibration signals combined with self-organizing maps for automatic feature selection", Int. J. Adv. Manuf. Technol., 108(5–6), pp. 1811–1825 (2020).

[15] Serin, G., Sener, B., Ozbayoglu, A. M., and Unver, H. O., "Review of tool condition monitoring in machining and opportunities for deep learning", Int. J. Adv. Manuf. Technol., 109(3–4), pp. 953–974 (2020).

[16] Wang, M. and Wang, J., "CHMM for tool condition monitoring and remaining useful life prediction", Int. J. Adv. Manuf. Technol., 59(5–8), pp. 463–471 (2012).

[17] Brecher, C., Esser, M., and Witt, S., "Interaction of manufacturing process and machine tool", CIRP Ann., 58(2), pp. 588–607 (2009).

[18] Quintana, G. and Ciurana, J., "Chatter in machining processes: A review", Int. J. Mach. Tools Manuf., 51(5), pp. 363–376 (2011).

[19] Zhou, Z., Yao, B., Xu, W., and Wang, L., "Condition monitoring towards energy-efficient manufacturing: a review", Int. J. Adv. Manuf. Technol., 91(9–12), pp. 3395–3415 (2017).

[20] Ravindra, H. V., Srinivasa, Y. G., and Krishnamurthy, R., "Acoustic emission for tool condition monitoring in metal cutting", Wear, 212(1), pp. 78–84 (1997).

[21] Wang, G., Yang, Y., and Li, Z., "Force Sensor Based Tool Condition Monitoring Using a Heterogeneous Ensemble Learning Model", Sensors, 14(11), pp. 21588–21602 (2014).

[22] Carden, E. P. and Fanning, P., "Vibration Based Condition Monitoring: A Review", Struct. Health Monit., 3(4), pp. 355–377 (2004).

[23] Schueller, A. and Salda–a, C., "Generalizability analysis of tool condition monitoring ensemble machine learning models", J. Manuf. Process., 84, pp. 1064–1075 (2022).

[24] Demir, H. G. and Yesilyurt, I., "A comparison of four machine learning techniques and continuous wavelet transform approach for detection and classification of tool breakage during milling process", Trans. Can. Soc. Mech. Eng., 47(1), pp. 26–42 (2023).

[25] Dey, A. and Yodo, N., "Performance improvement techniques for neural networks in tool condition monitoring", Int. J. Mechatron. Manuf. Syst., 15(2/3), p. 107 (2022).

[26] Hassan, M., Sadek, A., and Attia, H., "A Real-Time Deep Machine Learning Approach for Sudden Tool Failure Prediction and Prevention in Machining Processes", Sensors, 23(8), p. 3894 (2023).

[27] Gudelek, M. U., Serin, G., Ozbayoglu, A. M., and Unver, H. O., "An industrially viable wavelet long-short term memory-deep multilayer perceptron-based approach to tool condition monitoring considering operational variability", Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng., p. 095440892211421 (2022).

[28] Li, Y., Zhao, Z., Fu, Y., and Chen, Q., "A novel approach for tool condition monitoring based on transfer learning of deep neural networks using time–frequency images", J. Intell. Manuf. (2023).

[29] Kounta, C. A. K. A., Arnaud, L., Kamsu-Foguem, B., and Tangara, F., "Deep learning for the detection of machining vibration chatter", Adv. Eng. Softw., 180, p. 103445 (2023).

[30] Zhang, P., Gao, D., Hong, D., Lu, Y., Wu, Q., Zan, S., and Liao, Z., "Improving generalisation and accuracy of on-line milling chatter detection via a novel hybrid deep convolutional neural network", Mech. Syst. Signal Process., 193, p. 110241 (2023).

[31] Ross, N. S., Sheeba, P. T., Shibi, C. S., Gupta, M. K., Korkmaz, M. E., and Sharma, V. S., "A novel approach of tool condition monitoring in sustainable machining of Ni alloy with transfer learning models", J. Intell. Manuf. (2023).

[32] Naveen Venkatesh, S., Arun Balaji, P., Elangovan, M., Annamalai, K., Indira, V., Sugumaran, V., and Mahamuni, V. S., "Transfer Learning-Based Condition Monitoring of Single Point Cutting Tool", Comput. Intell. Neurosci., 2022, pp. 1–14 (2022).

[33] Ou, J., Li, H., Liu, B., and Peng, D., "Deep transfer residual variational autoencoder with multi-sensors fusion for tool condition monitoring in impeller machining", Measurement, 204, p. 112028 (2022).

[34] Staacks, S., H–tz, S., Heinke, H., and Stampfer, C., "Advanced tools for smartphone-based experiments: phyphox" (2018).

[35] Orhan, S., Er, A. O., Camu?cu, N., and Aslan, E., "Tool wear evaluation by vibration analysis during end milling of AISI D3 cold work tool steel with 35 HRC hardness", NDT E Int., 40(2), pp. 121–126 (2007).

[36] "Discrete time signal processing. Hauptbd.–, 2. ed, Prentice Hall, Upper Saddler River, NJ (1999).

[37] Dheeraj Simon, G. and Deivanathan, R., "Early detection of drilling tool wear by vibration data acquisition and classification", Manuf. Lett., 21, pp. 60–65 (2019).

[38] Randall, R. B., Vibration-Based Condition Monitoring: Industrial, Aerospace and Automotive Applications, Wiley, Chichester (2011).

[39] Khusheef, A. S., Shahbazi, M., and Hashemi, R., "Deep Learning-Based Multi-Sensor Fusion for Process Monitoring: Application to Fused Deposition Modeling", Arab. J. Sci. Eng., 49(8), pp. 10501–10522 (2024).

[40] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P., "Gradient-based learning applied to document recognition", Proc. IEEE, 86(11), pp. 2278–2324 (1998).

[41] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C., "MobileNetV2: Inverted Residuals and Linear Bottlenecks", 2018 IEEECVF Conf. Comput. Vis. Pattern Recognit., IEEE, Salt Lake City, UT, pp. 4510–4520 (2018).

[42] Aoyagi, K., Wang, H., Sudo, H., and Chiba, A., "Simple method to construct process maps for additive manufacturing using a support vector machine", Addit. Manuf., 27, pp. 353–362 (2019).

[43] Wu, D., Wei, Y., and Terpenny, J., "Surface roughness prediction in additive manufacturing using machine learning", Manuf. Equip. Syst., ASME 2018 13th International Manufacturing Science and Engineering Conference, MSEC 2018, American Society of Mechanical Engineers (ASME) (2018).

[44] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., and Louppe, G., "Scikit-learn: Machine Learning in Python", J. Mach. Learn. Res., 12 (2012).

[45] Talayoglu, B., Vande Velde, J., and Da Silva, B., "Lightweight AI for Sensor Fault Monitoring", Electronics, 14(22), p. 4532 (2025).

[46] Pires, I. M., Hussain, F., Marques, G., and Garcia, N. M., "Comparison of machine learning techniques for the identification of human activities from inertial sensors available in a mobile device after the application of data imputation techniques", Comput. Biol. Med., 135, p. 104638 (2021).

[47] Ahmed, Y. S., Arif, A. F. M., and Veldhuis, S. C., "Application of the wavelet transform to acoustic emission signals for built-up edge monitoring in stainless steel machining", Measurement, 154, p. 107478 (2020).