

# Real-Time Fault Detection in Robotic Manufacturing Using High-Bandwidth Event Vision-Based Tactile Sensing

Eslam Sherif<sup>a,b</sup>, Akram Khairi<sup>a,b,\*</sup>, Hussain Sajwani<sup>a,b</sup>, Abdullah Solayman<sup>a,b</sup>, Abdallah Mohammad Alkilany<sup>a,b</sup>, Ahmed Awadalla<sup>a,b</sup>, Mohamad Halwani<sup>a,b</sup>, Laith AbuAssi<sup>a,b</sup>, Dewald Swart<sup>c</sup>, Abdulla Ayyad<sup>a,b</sup> and Yahya Zweiri<sup>a,b</sup>

<sup>a</sup>Advanced Research and Innovation Center (ARIC), Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

<sup>b</sup>Department of Aerospace Engineering, Khalifa University, Abu Dhabi, United Arab Emirates

<sup>c</sup>Research and Development, Strata Manufacturing PJSC, Al Ain, UAE

---

## ARTICLE INFO

### Keywords:

Event vision-based tactile sensing; Robotic Machining; In-process fault monitoring; Transfer learning; Industry 4.0; Aerospace manufacturing

---

## ABSTRACT

Early detection of machining faults is critical to avoid damage to high-value workpieces, prevent tool failure, and mitigate safety risks in automated manufacturing environments. While robotic automation has advanced across manufacturing, aerospace machining remains difficult to automate due to strict quality requirements and the lack of intelligent, real-time fault monitoring. During machining, skilled human operators rely on their tactile perception to detect subtle faults such as tool wear, misalignment, or insufficient feed force through vibration cues. Replicating this level of high-resolution tactile fidelity in robotic systems remains a key challenge. We present an event Vision-Based Tactile Sensor (EVBTS) that enables robots to perceive and interpret machining vibrations with human-like sense of touch. The sensor uses an event camera to observe a deformable, marker-embedded membrane, capturing fine-grained spatiotemporal deformation patterns with microsecond latency. This high-fidelity, biomimetic signal stream allows robotic systems to detect faults in machining dynamics. We evaluate EVBTS on a robotic drilling setup for aerospace nutplate installation, spanning 12 distinct machining conditions. A lightweight convolutional neural network, integrated into a real-time pipeline with Exponential Moving Average (EMA) filtering, achieves 98.56% classification test accuracy, a 98.11% test F1 score, and <100 ms inference latency. This pipeline demonstrates closed-loop feedback, successfully halting faulty operations mid-process to prevent defects. These results demonstrate that EVBTS enables real-time, high-resolution fault detection and intervention, allowing for early correction, much like a skilled human operator, supporting safer, more precise, and autonomous manufacturing.

---

## 1. Introduction

### Background and Motivation

Industry 4.0 marks the current pinnacle of the industrial revolution that began over a hundred years ago. This swift advancement represents a paradigm shift towards highly automated, intelligent, and data-driven manufacturing. This enhanced the integration of sophisticated quality assurance practices, such as digital twins[1, 2], machine vision [3], and sensor technologies [4], leading to more effective methods on production lines [5, 6, 7]. Transitioning from manual machinery to completely automated factories, the industry consistently seeks highly-efficient practices [8]. While automation is advancing, many precision tasks still rely on the nuanced perception of skilled human operators, who intuitively use visual, auditory, and tactile cues (primarily vibrations) to monitor process health and intervene before defects occur. Replicating this high-fidelity, real-time diagnostic capability in robotic systems remains a significant barrier to full automation [9, 10].

A key challenge in intelligent manufacturing is moving beyond reactive, post-process inspection, which only identifies defects after it has already occurred, leading to costly rework, scrapped parts, and potential safety hazards [9, 11]. The clear limitations of this approach establish a need for in-process monitoring systems that can track process health in real-time. To this end, a variety of sensors, such as accelerometers, acoustic emission sensors, and

---

\*Corresponding author

 (E. Sherif); (A. Khairi); (H. Sajwani); (A. Solayman); (A.M. Alkilany); (A. Awadalla); (M. Halwani); (L. AbuAssi); (D. Swart); (A. Ayyad); (Y. Zweiri)

ORCID(s):

dynamometers, are commonly employed to detect anomalies as they happen. However, these conventional in-process sensors often suffer from inherent physical limitations or insufficient data bandwidth to capture the complex, high-frequency dynamics critical for comprehensive fault analysis [11]. More fundamentally, they are largely limited to signaling that a fault has occurred, rather than providing a real-time diagnosis of the fault's nature or its propagation mechanism, a capability crucial for timely intervention and intelligent manufacturing. [11]

A distinct approach that evolved for high-fidelity inspection known as Vision-Based Tactile Sensing (VBTS) [12], offers a path toward more advanced in-process monitoring systems. VBTS excels at capturing detailed, high-resolution maps of the entire contact area [13, 14, 15], moving far beyond the single-point measurements provided by many traditional sensors. Because of its spatial-sensing capability, there is significant interest in adapting VBTS for real-time, in-process monitoring. However, this transition is hindered by the low frame rates of the conventional cameras used in these systems, which are insufficient to capture the high-frequency dynamic events common in manufacturing scenarios [11]. The solution to this temporal bottleneck is the replacement of the conventional camera with an event camera. While conventional frame-based cameras capture full images at fixed intervals, event-based cameras operate asynchronously, reporting only pixel-level brightness changes with microsecond latency [16, 17]. This innovation unlocks the potential of VBTS, leading to the concept of an Event Vision-Based Tactile Sensor (EVBTS), a technology capable of providing both the high spatial and temporal resolution required for true in-situ diagnosis of high-frequency dynamics in realistic manufacturing settings.

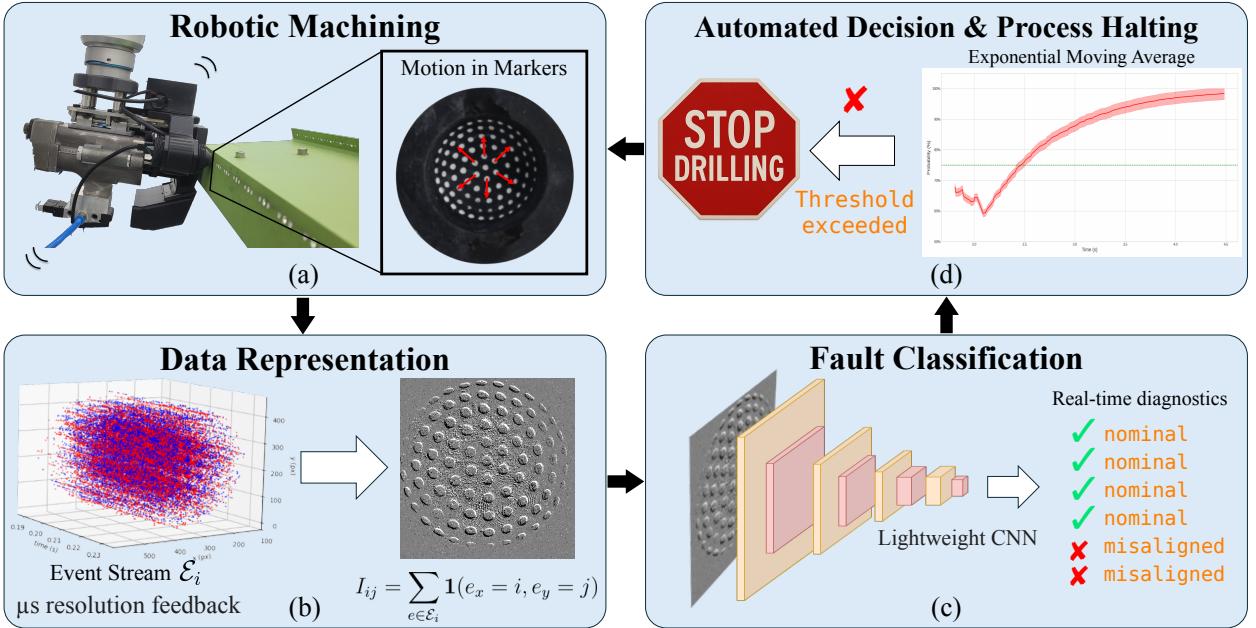
This paper introduces a new paradigm for in-situ monitoring of manufacturing processes by realizing the potential of EVBTS. While the event camera provides the necessary high-bandwidth data, the nature of the event-based data stream is fundamentally different from conventional camera frames, making traditional image processing techniques incompatible. This necessitates the development of a lightweight perception pipeline tailored for event-based data, which is a key part of our contribution. We present an end-to-end EVBTS system, designed for real-time fault detection and diagnosis in robotic machining, as illustrated in Figure 1. We demonstrate its effectiveness on a robotic drilling task, where the system can identify a wide range of nominal and fault machining conditions in real-time and intervene to halt the process before a fault leads to an excessively damaged part. The main contributions of this paper are as follows:

1. We present an end-to-end event vision-based tactile sensing system for fault detection in robotic machining that emulates the high-resolution, low-latency perception of a skilled human operator, while offering objective, quantifiable diagnostics that surpass the limits of human subjectivity.
2. We develop a lightweight machine learning pipeline capable of extracting rich diagnostics from event-based tactile signals with sub-100 ms latency. A transfer learning framework is incorporated to ensure robustness to sensor wear and adaptability across different machining setups.
3. We validate the system on a robotic drilling task for aerospace nutplate installation, covering 12 distinct machining conditions, including healthy operation, tool misalignment, and improper feed force. The system demonstrates high classification accuracy, real-time inference, and the ability to interrupt faulty operations mid-process in realistic aerospace manufacturing scenarios.

## Related Work

The foundation of quality assurance in manufacturing has traditionally rested on reactive, post-process inspection. While methods such as post-process visual checks, manual gauging, or non-destructive testing (NDT) like Eddy Current sensing provide definitive, ground-truth assessment of final part quality [18], their reactive nature is identified as an inefficient and costly option for process optimization [11]. While essential for final verification, these methods involve categorizing fault causes only after damage has occurred, resulting in scrapped high-value parts, costly rework, and potential safety hazards [9], or detecting anomalies without any real-time comprehension of their root causes [19].

To overcome the limitations of offline inspection, a variety of sensors are employed for in-process monitoring. Common choices include accelerometers, acoustic emission (AE) sensors, dynamometers (for forces and torque), microphones, strain gauges, temperature sensors, optical systems, and displacement sensors [20, 21]. However, while conceptually advantageous, many conventional in-process sensors face practical limitations that hinder their real-time diagnostic fidelity. For instance, while dynamometers and strain gauges provide valuable direct force measurements, they can be costly, complex to integrate, require periodic calibration to maintain accuracy [22], and often present a fundamental trade-off where achieving high sensitivity requires reducing the structural stiffness of integrated components, compromising overall machine tool precision and stability [23]. Similarly, optical and thermal



**Figure 1:** Workflow of the instantaneous fault detection system. **(a)** Process-induced vibrations are captured by the EVBTS as **(b)** a sparse, microsecond-resolution event stream. The stream is converted into frames and fed into **(c)** a lightweight CNN for real-time classification. **(d)** The resulting class probabilities are temporally smoothed by an Exponential Moving Average (EMA), enabling robust fault detection and automated intervention when a predefined threshold is exceeded.

systems, though effective for certain measurements, are challenged by the harsh cutting environment, often requiring a clear line-of-sight that is obstructed by chips and coolant, making continuous in-process measurement difficult [24]. Furthermore, sensors that capture process dynamics through vibration and sound, such as microphones and acoustic emission sensors, are widely used for their sensitivity, but their diagnostic fidelity can be compromised by high susceptibility to ambient factory noise and structural vibrations, and computationally intensive processing demands that hinder real-time analysis [25]. Hence, this entire class of methodologies is largely limited to fault detection; they can signal that a fault has occurred, but they fail to provide a real-time diagnosis of the fault's nature or its propagation mechanism.

Accelerometers, while ubiquitously employed for general machine health assessment, condition-based maintenance, and detecting gross failures like chatter, their utility is significantly diminished for the high-fidelity, real-time diagnostic tasks required in precision robotic applications due to several inherent issues. Their readings are vulnerable to temperature fluctuations and electromagnetic interference, which distort signals and lead to false readings [26, 27]. Furthermore, their performance is a difficult compromise between high-sensitivity sensors that often cannot withstand the high-shock operations common in machining and robust sensors that lack the necessary resolution [28, 29]. In multi-axis robotic machining, drift introduces further errors, and their constrained bandwidth makes it difficult to capture both subtle, high-frequency chatter and large-amplitude shocks simultaneously with a single sensor [30, 29]. Perhaps most fundamentally, their performance is highly sensitive to mounting location. As they are typically placed away from the cutting zone, they provide an indirect, noisy, and delayed measurement of the critical dynamics occurring at the tool-workpiece interface [11].

A rising avenue for acquiring rich diagnostic data that moves beyond single-point sensors is tactile sensing. It offers a holistic source of information about interaction forces and contact states by capturing data across a distributed area, a capability enabled by its high spatial resolution. While various modalities exist, including piezoresistive [31], capacitive [32], and piezoelectric [33], a particularly prominent category is Vision-Based Tactile Sensing (VBTS) [12]. Its primary advantage lies in its ability to capture a high-resolution map of the entire contact interface. Furthermore, their fabrication is often accessible, leveraging techniques like 3D printing for molds and direct elastomer casting to create the deformable membrane [14, 34, 35]. This combination of rich data acquisition and accessible fabrication

has led to the development of sensors like GelSight [13], TacTip [14], and DIGIT [15]. The potential of these high-resolution tactile sensors is already being leveraged in robotic manufacturing. For instance, custom VBTS designs have been integrated into robotic end-effectors to ensure critical geometric tolerances, such as normality control [34] and robotic deburring [36], while continuous-scanning designs like GelBelt [37] have been developed for large-area surface inspection. However, despite these successes, the reliance of conventional VBTS on standard, frame-based cameras imposes a fundamental limitation for monitoring dynamic processes. The typical frame rates (e.g., 30-100 Hz) are insufficient to capture the high-frequency vibrations (often in the kHz range) that are indicative of important process anomalies in machining, creating a critical bottleneck for real-time applications [11].

To overcome this critical bandwidth issue, the field is turning towards event-based cameras. Unlike conventional cameras that capture and transmit entire image frames at a fixed rate (e.g., 30 times per second), event cameras feature independent pixels that operate asynchronously. Each pixel reports an "event", a small data packet, only when it detects a change in brightness. This results in a sparse, high-temporal-resolution data stream (microsecond-level) that avoids redundancy and is ideal for capturing fast dynamics [16, 17]. Indeed, event-based vision is already demonstrating significant potential for high-speed manufacturing tasks, including precise positioning for robotic drilling [38], high-speed inspection of features in large structures like countersinks [39], and capturing rich tactile signals for contact angle prediction and normality control [40, 41]. This success has logically led to the development of Event Vision-Based Tactile Sensors (EVBTS), which fuse the high spatial resolution of tactile sensing with the high temporal resolution of event-based vision. It moves beyond simple measurement, enabling a biomimetic perception of dynamic events akin to a human operator's sense of touch. While the underlying technologies have been explored for tasks like slip detection [42, 43], this work represents the first ever application of this concept to the demanding domain of real-time fault detection and diagnosis in robotic manufacturing. This technology's potential extends beyond fully automated systems, opening possibilities for compact, handheld assistive tools that could augment a human operator's tactile perception in real-time.

The rest of the article is organized as follows: Section 2 outlines the hardware fabrication, signal processing scheme, data collection methods, and learning model structure. Section 3 describes all the components into the experimental setup during drilling. Section 4 explains in detail the data set and the sensor insights along with neural network performance, accuracy of recognition, and computational efficiency in such application. Summary of findings along with future work and limitations are all placed into Section 5.

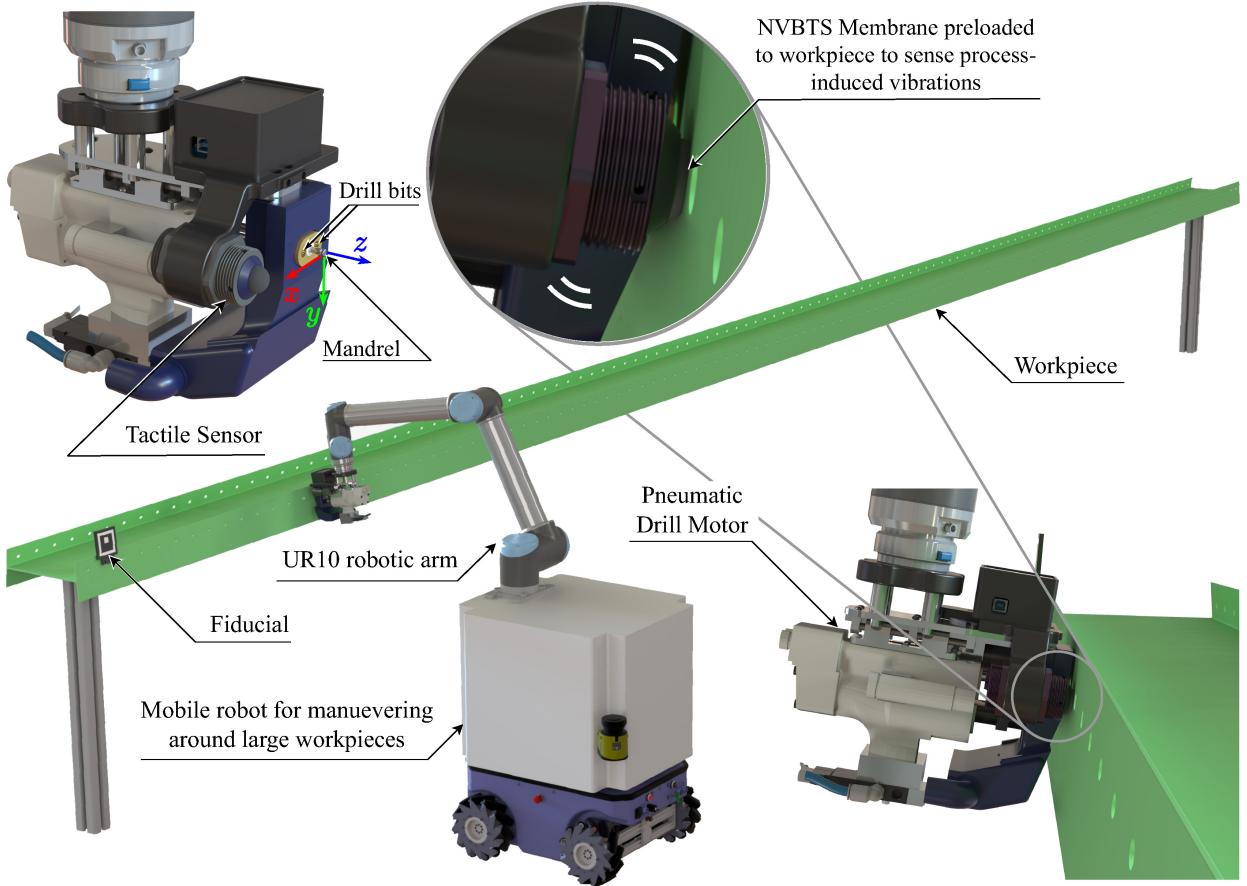
## 2. Methodology

Figure 2 illustrates our integrated drilling and condition-monitoring platform. A Universal Robots UR10 arm executes programmed drilling trajectories while carrying a custom end-effector that houses three subsystems: (i) a pneumatic drill motor; (ii) an RGB camera for ArUco-based fiducial detection, TCP alignment, and precise hole positioning [36]; and (iii) an event vision-based tactile sensor (EVBTS) for capturing process-induced vibrations. During each drilling cycle, the robot first uses the RGB camera to locate the fiducial marker on the test coupon and determine the target hole position. It then approaches the workpiece, inserts the drill into the hole [36], and brings the EVBTS membrane into preloaded contact beside the drill motor. Once drilling begins, the DVXplorer Micro inside the EVBTS captures asynchronous events generated by micro-deformations of the elastomer markers. These event streams are transmitted to a host computer, where they are preprocessed and classified by a lightweight convolutional neural network to detect anomalous (off-nominal) conditions in real time.

### 2.1. Hardware

#### 2.1.1. Operating Principle

Upon precisely localizing the target hole, the UR10 advances the pneumatic drill's mandrel into the hole, bringing the tactile sensing element into direct, preloaded contact with the workpiece and causing it to deform. As drilling begins, process-induced vibrations travel through the workpiece into the preloaded tactile sensing element, causing micro-deformations of its markers. The event camera, due to its high temporal resolution, is capable of capturing these micro-scale displacements as sparse, asynchronous ON/OFF events with microsecond-level precision. Each machining condition, either nominal or fault condition produces a characteristic spatio-temporal event pattern, which is analyzed in real-time to detect and classify off-nominal states.

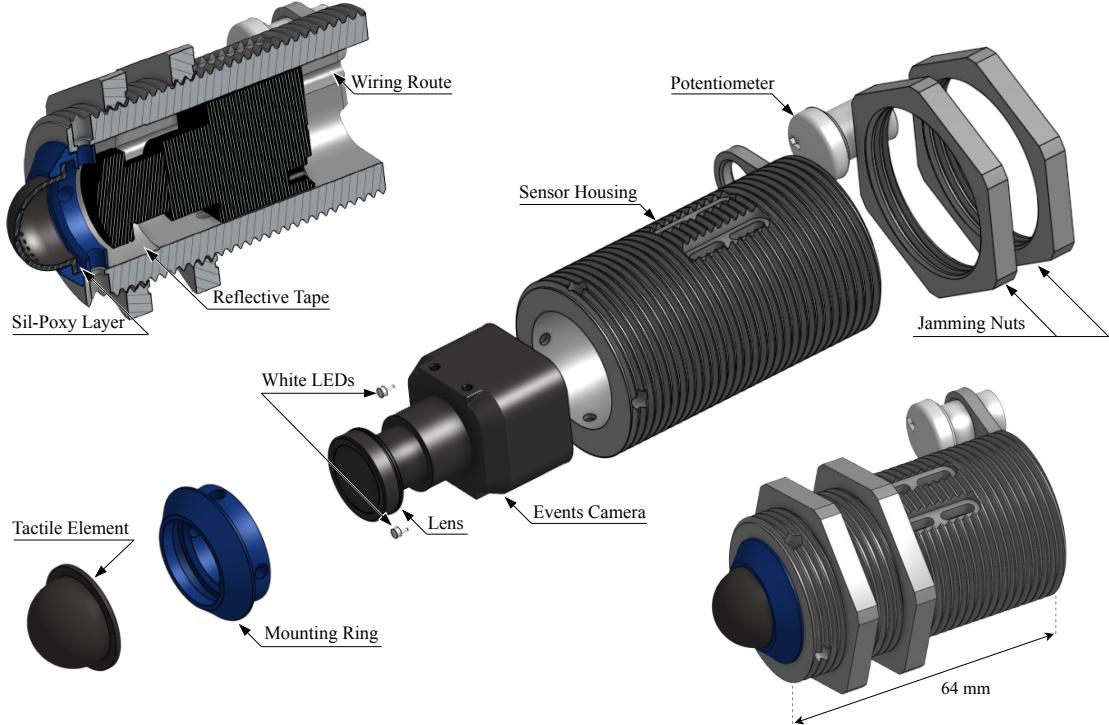


**Figure 2:** In-situ fault detection and monitoring for robotic machining. A UR10 robotic arm carries a custom end-effector integrating a pneumatic drill motor, the EVBTS tactile sensor and RGB camera. The zoom inset highlights the EVBTS elastomer membrane preloaded against the workpiece, where micro-scale marker motion is captured as asynchronous events encoding the process vibration signature.

### 2.1.2. EVBTS Sensor Design

The proposed event Vision-Based Tactile Sensor (EVBTS) was designed and developed for multiple purposes, namely, monitoring the condition of machining processes through vibration signature and perpendicularity estimation for precise machining processes [36]. The sensor assembly is illustrated in Figure 3. The sensor is designed to be integrated with a robotic end-effector to enable in-situ monitoring of machining processes such as drilling, countersinking, or deburring. The core of the sensor comprises a custom-designed housing featuring an external M36x2 mm thread fabricated using an SLA 3D printer for its high precision to preserve the thread features. This thread engages with two jamming nuts. The sensor can be integrated with the robotic end-effector either using these jamming nuts or using an internally threaded adapter that conforms to the outer thread of the sensor housing and a locking nut. This mechanism provides precise, fine-grained control over the protrusion depth of the sensor's tactile sensing element relative to its mounting base. By adjusting and locking the nuts, the depth to which the tactile sensing element presses against the workpiece surface can be accurately set and maintained, ensuring consistent contact conditions critical for repeatable vibration signature monitoring.

Internally, the housing securely mounts an iniVation DVXplorer Micro event-based camera. This camera was selected for its miniature size, high temporal resolution, and sparse data output, which are hypothesized to be advantageous for capturing the high-frequency dynamics characteristic of machining vibrations. The camera is fitted with an M12 lens with a 4 mm fixed focal length, chosen to provide an appropriate field of view of the tactile element's internal markers. The tactile sensing element is positioned directly in front of the camera, within its focal



**Figure 3:** Exploded and cross-sectional views of the EVBTS sensor.

plane. This tactile element makes direct contact with the workpiece and deforms under the influence of pressing against it during the insertion of the machining tool. We hypothesize that machining-induced vibrations propagate through the workpiece to this tactile sensing element, and these vibrations cause micro-deformations that cause the markers on the internal surface of the membrane to move in a characteristic pattern. The event-based camera, due to its high temporal resolution, captures this characteristic motion that could instantaneously be correlated to the condition of the machining process. The tactile sensing interface is secured to the main sensor housing using a custom-designed mounting ring, highlighted in blue in Figure 3. This mounting ring features an internal recession that conforms to the base geometry of the tactile element. The base of the element is seated within this groove, where it is bonded using Sil-Poxy (Smooth-On, Inc.), a silicone-based adhesive. This provides a durable chemical bond that is robust against the sensor's primarily compressive operational loads, which avoid the peeling or high-shear forces that typically cause failure for this adhesive. Moreover, this adhesive integrates the tactile element with the mounting ring, preventing slippage or detachment during dynamic interactions, while allowing the domed sensing surface of the tactile element to deform freely in response to tactile stimuli. The mounting ring's geometry further enhances this integrity by mechanically constraining the elastomer base, creating a redundant fixture that prevents slippage or detachment during dynamic interactions while allowing the domed sensing surface to deform freely. The elastomer itself is subject to material softening (Mullins effect) from cyclic fatigue, which can alter its response over time; for example, its elastic modulus can drop by around 76% after 2000 compression cycles [44]. Therefore, the mounting ring is designed for modularity, enabling easy exchange of the tactile element to address such wear. This sub-assembly is attached to the main sensor housing via a circular pattern of four M2 screws that engage with corresponding brass heat-set inserts embedded in the mounting ring, ensuring a robust and repeatable connection.

Illumination of the tactile element's underside is provided by two miniature white LEDs integrated within the housing. To ensure uniform, diffused illumination and minimize specular highlights, which could negatively impact camera performance, the inner cylindrical wall of the housing section between the LEDs and the tactile element is lined with a reflective tape. The intensity of the LEDs is adjustable via a potentiometer mounted externally on the rear of the sensor housing. This allows for manual tuning of the illumination level to optimize contrast and event generation from the camera.

**Table 1**

Key Parameters of the EVBTS Tactile Sensing Element.

Parameter	Value / Description
Outer Base Diameter (Flange)	26 mm
Inner Dome Diameter (at base)	16 mm
Overall Height (Flange to Apex)	12 mm
Membrane Wall Thickness	1 mm
Shape	Hemispherical Dome with Integrated Flange
Number of Markers	91
Marker Arrangement	Concentric Circular Rows
Marker Diameter	0.8 mm
Marker Depth	0.6 mm
Marker Shape	Recessed Cylindrical Indentations
Material	Ecoflex 00-30
Hardness (Durometer)	Shore 00-30
Fabrication Method	Cast in SLA 3D-Printed Two-Part Mold

### 2.1.3. EVBTS Sensor Fabrication

The tactile element is the primary sensing interface of the proposed sensor. It comprises a hemispherical shell fabricated from black-pigmented silicone elastomer (Ecoflex 00-30, SmoothOn, Inc.) and cylindrical recessions distributed radially on the inner surface that represent the markers. Table 1 illustrates the sensor design parameters. The manufacturing process, illustrated in Figure 4, begins by creating a CAD model of the mold designed specifically to produce the desired shape of the tactile element. The mold is fabricated using Stereolithography (SLA) 3D printer, which offers resolution up to  $25 \mu\text{m}$ . This is essential to preserve small-scale cylindrical protruding features that are used later to produce the desired markers features. The mold is composed of two parts, cavity and core, with two guiding pins in the core and locating holes in the cavity. The silicone base, curing agent, and pigment are mixed and stirred in a 50:50:1 weight ratio, degassed under vacuum to remove trapped air, and then poured into the mold cavity. After curing for four hours at room temperature, the elastomer is demolded. In the final step, the recessed markers on the inner surface are filled with white, silicone-based ink and heat-treated in a convection oven at  $80^\circ\text{C}$  for 10 minutes to create a durable, high-contrast bond with the silicone shell.

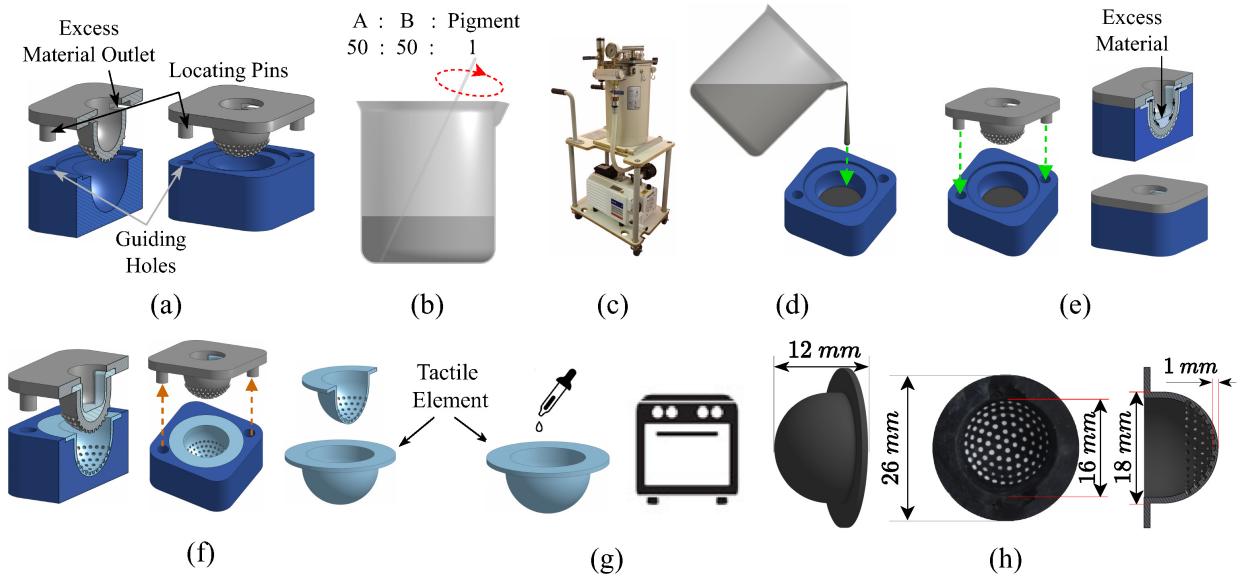
The material and its geometry were chosen to balance durability with sensitivity to high-frequency vibrations. The Ecoflex 00-30 Shore hardness was empirically selected; silicones with lower hardness (e.g: Ecoflex 00-10) lacked the necessary durability, while harder grades (e.g: Ecoflex 00-50) were too stiff to respond effectively to subtle vibration signatures. The dynamic response and recovery time were further tuned via the element's 1 mm wall thickness, which provided the rapid responsiveness required to capture high-frequency events while ensuring mechanical robustness.

## 2.2. Signal Processing

The raw data from the event camera is a stream of asynchronous events, which requires specific processing to be utilized by conventional machine learning models for instantaneous fault detection in robotic machining. This section details the working principles of the integrated event camera and the preprocessing steps applied to convert the event stream into suitable representations.

### 2.2.1. Event Camera Working Principles

Unlike conventional cameras that capture intensity information across all pixels at fixed frame rates, event cameras, specifically the iniVation DVXplorer Micro event camera integrated into our EVBTS, employs a fundamentally distinct, bio-inspired sensing paradigm [16]. Each of its  $640 \times 480$  pixels operates independently and asynchronously, responding dynamically to changes in illumination. Specifically, a pixel monitors the logarithm of the incident light intensity,  $L(x, y, t) = \log(I(x, y, t))$ . An event  $e_k = (x_k, y_k, t_k, p_k)$  is generated at coordinates  $(x_k, y_k)$  only when the change in this logarithmic intensity,  $\Delta L$ , relative to its value at the time of the last event from that pixel, surpasses a predefined contrast threshold,  $\pm C$ . The generated event includes a high-resolution timestamp  $t_k$  (typically at microsecond precision) and a polarity  $p_k \in \{0, 1\}$ , signifying the direction of the brightness change (decrease or



**Figure 4:** Fabrication workflow of the EVBTS tactile sensing element. The process includes (a) the two-part SLA-printed mold, (b-d) silicone mixing, degassing, and pouring, (e-f) curing and demolding, (g) inking the markers on the final elastomer, and (h) top view of the tactile sensor

increase, respectively). Following event generation, the pixel’s internal reference brightness level is updated, priming it for subsequent intensity variations.

The output of the EVBTS is thus not a sequence of frames, but a sparse, continuous stream of asynchronous  $(x, y, t, p)$  tuples reflecting the dynamic changes in the tactile element’s internal marker positions due to vibration. When reading out data from the event camera, events generated within a short accumulation period on the camera hardware are typically bundled together. For our system, these packets of events are received by the host computer, forming the basis for subsequent processing. The unique characteristics of the event camera offer substantial benefits for capturing the high-frequency dynamics characteristic of machining vibrations. Its inherent high temporal resolution allows for sensing rapid marker movements without motion blur. Furthermore, event-based cameras benefit from low latency and a high dynamic range, which is advantageous for reducing computational load due to data sparsity and maintaining consistent performance despite potential variations in internal illumination within the sensor housing.

### 2.2.2. Preprocessing: Event Representations

The raw, asynchronous event stream from the EVBTS is a collection of individual events, each characterized by its pixel location, timestamp, and polarity. Following common practice in event-based vision literature [42, 45], we represent the event stream as a set of events  $\mathcal{E} = \{e_k\}_{k=1}^{N_e}$ , where  $e_k = (x_k, y_k, t_k, p_k)$  denotes the  $k$ -th event, with  $(x_k, y_k)$  being the pixel coordinates,  $t_k$  the timestamp, and  $p_k \in \{0, 1\}$  the polarity. This sparse, high-temporal-resolution data is then transformed into a structured format suitable for input to conventional deep learning architectures like CNNs. Our preprocessing pipeline focuses on generating event frames that summarize event activity over defined windows.

As detailed in Section 2.1.3 and illustrated in Figure 4, the tactile sensor is fabricated from black-pigmented silicone with white, silicone-based ink used for the internal markers. This choice of materials is deliberate, designed to maximize the contrast between the markers and the elastomer background. When the elastomer deforms due to vibrations, the white markers move against the black background. This movement causes significant changes in brightness at the pixel locations corresponding to the edges of the markers, thereby reliably triggering events. Figure 5(a) conceptually illustrates how the movement of a high-contrast white marker against a dark background generates both ON (brightness increase) and OFF (brightness decrease) events at its leading and trailing edges, respectively.

For processing, the segment of the event stream corresponding to the actual material engagement phase (drilling and/or countersinking) is first identified. This is achieved by post-processing a histogram of the number of events over

time for each recorded hole operation, allowing for manual or semi-automated segmentation of the relevant "drilling phase" and exclusion of approach and retraction phases.

Within this segmented "drilling phase," events are accumulated into frames not based on fixed time windows, but on fixed event counts. Specifically, a frame is generated from a window of 30,000 consecutive events, with a stride of 5,000 events between successive windows. This event-count based windowing ensures that each frame contains a consistent amount of event information, which can be more robust to variations in event rates during different phases or intensities of vibration.

We investigated three different event representation strategies for the CNN input, whose efficacy is further explored in the ablation study (Section 4.3.2), as illustrated in Figure 5(b).

**Combined Event Frames ( $F_C$ ):** These frames accumulate all events within a given window  $\mathcal{W}_i$ , irrespective of their polarity. Each pixel value represents the total count of events at that pixel location:

$$F_C(x, y) = \sum_{e_k \in \mathcal{W}_i} \mathbb{I}(x_k = x, y_k = y) \quad (1)$$

where  $\mathbb{I}(\cdot)$  is the indicator function. This effectively creates an activity map visualizing the aggregate marker deformation and movement.

**Polarity Event Frames ( $F_P$ ):** For these frames, event polarities are explicitly considered. The pixel value is the sum of polarities of events occurring at that location, potentially reflecting directional changes in brightness:

$$F_P(x, y) = \sum_{e_k \in \mathcal{W}_i} p_k \cdot \mathbb{I}(x_k = x, y_k = y) \quad (2)$$

ON and OFF events are treated as positive (+1) and negative (-1) values, respectively, and are accumulated in a high-precision intermediate grid. The resulting frame, containing a range of positive and negative counts, is then normalized. For visualization purposes (as seen in Figure 5(b)(ii)), this maps the highest positive count (most net ON events) to white (255), the highest negative count (most net OFF events) to black (000), and a zero count to gray (127).

**Time Surface Frames ( $F_{TS}$ ):** In this representation, event time surfaces,  $T(x)$ , that retain the grayscale intensity of the scene are generated. Given an event packet,  $\mathcal{E}$ , received at time  $t$ , and  $N_e$  is the total number of events, the pixel  $x$  of  $T_k$  is updated for each event  $e_k$  as follows:

$$T_k(x) = \begin{cases} T_0 \in \mathbb{R}^{h \times w}, & \text{if } k = 0 \\ T_{k-1}(x) \exp\left(-\frac{t_k - t_0}{\tau}\right) + p_k \lambda \delta(x - x_k), & \text{if } k > 0 \end{cases} \quad (3)$$

where  $T_0$  is the event time surface image at  $t_0$  initialized with null values of 0, and  $\tau$  and  $\lambda$  are hyperparameters defined as the decay factor and event contribution, respectively.

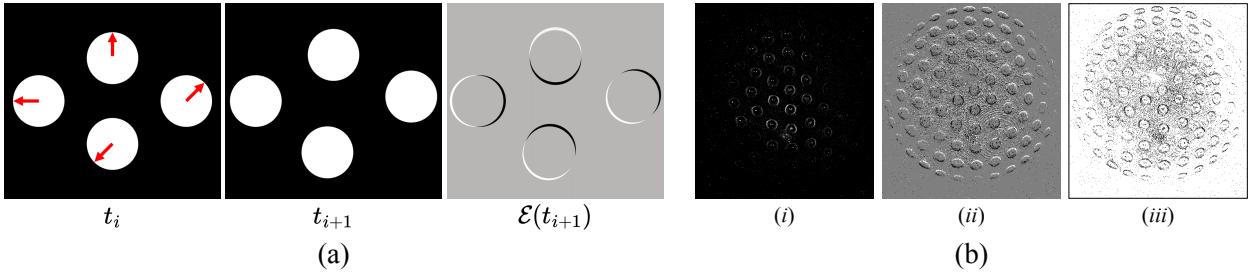
Noise from light leakage and fabrication errors is mitigated by both design and processing. The sensor's sealed, opaque housing with black pigmented silicone and controlled internal white LEDs (Section 2.1.2) prevents ambient light interference. Furthermore, the preprocessing step of thresholding high pixel counts, as detailed below, effectively filters out noise like "hot pixels" or spurious events from minor imperfections. Subsequently, the data is spatially cropped to a Region of Interest (ROI) focused on the active marker area, discarding peripheral data which could contain noisy events.

To construct these frames, for a given window  $\mathcal{W}_i$ , a 2D grid (image) initialized to zeros is created, with dimensions corresponding to the sensor resolution (e.g.,  $H \times W$ , which is  $480 \times 640$  for our case before any cropping for the network). These raw event count frames are then processed by thresholding high pixel counts (e.g., counts above 10 are set to 0 or capped) to mitigate "hot pixel" effects or extremely high-frequency local activity, followed by normalization (e.g., scaling pixel values to a range of [0, 1] for network input).

These event frames serve as the primary input to our CNN-based learning models. This transformation allows us to leverage well-established CNN architectures for image-based pattern recognition. Each generated frame is then labeled with the ground truth machining condition corresponding to the hole operation from which it was derived.

### 2.3. Data Collection Protocol

Having established the method for converting raw event streams into processed event frames in Section 2.2.2, this section details the experimental protocol designed to collect these initial event streams. The protocol encompasses a variety of controlled machining conditions, including a nominal (healthy) baseline and several deliberately induced fault conditions relevant to robotic machining.



**Figure 5:** (a) High-contrast marker motion generates ON/OFF events at leading/trailing edges. Movement from time  $t_i$  to  $t_{i+1}$  generates the events  $\mathcal{E}(t_{i+1})$ . (b) These sparse events are accumulated into event frame representations, including (i) Combined Event Frames, (ii) Polarity Frames, and (iii) Time Surface Frames.

### 2.3.1. Nominal Drilling Case Definition

The nominal drilling and countersinking case was established using parameters aligned with standard aerospace practices for the materials and tools involved [36]. Defining characteristics included: securely fixing the aluminum test coupon to ensure workpiece stability (as shown in Figure 6(i) and described in Section 3), and maintaining end-effector stability through rigid robot control with accurate weight input and approximate center of mass estimation. The drill motor mandrel was calibrated to ensure perpendicularity to the workpiece surface. The drill feed rate was set to a nominal value of 10 via a Kinecheck valve, empirically determined to ensure proper countersink depth for flush rivet installation. Drilling pressure (torque) was maintained at 90 psi, correlating to nominal drilling torque per industry standards. Countersink depth was precisely adjusted to achieve a 100° angle with a 4.85 mm outer diameter, ensuring a standard rivet head would sit flush. Finally, a healthy, sharp tool condition was ensured by using high-speed steel drill bits with a 2.45 mm drill diameter and a 6.36 mm maximum countersink diameter. Operations performed under these tightly controlled conditions were labeled as nominal (healthy) and served as the baseline parameters.

### 2.3.2. Fault Conditions Definitions and Induction

We systematically induced multiple distinct fault machining conditions. In total, we investigated 11 fault condition variations, in addition to the nominal baseline.

**Unstable Workpiece Fixture.** To simulate an improperly secured workpiece, the clamping force on the test coupon was intentionally reduced. As shown in Figure 6(a) with labeled bolts, this was achieved by selectively loosening M4 bolts ④ and ⑤ by 2 turns to simulate instability of the test coupon with relation to the fixture and loosening bolts ⑥ and ⑦ securing the ArUco fiducial to the coupon by 1 turn (to preserve the fiducial's alignment). On the other hand, the main M8 fixture bolts (labeled ①–③) remained fully tightened. This method reproduces fixture instability while maintaining overall workpiece positioning for consistent vibration-signature capture.

**Unstable/Loose End-Effector (Simulated).** To simulate a loose or compliant end-effector mounting, the UR10 robot was switched to "free drive" mode just before starting the drilling operation. The robot's configured payload was set to 0 kg in this mode to maximize its responsiveness to external forces, simulating a lack of rigidity of the end effector or a compensating torque from the robotic arm joints.

**End-Effector Pressurefoot Misalignment.** The end-effector's pressurefoot, and hence the drill bits, orientation was deliberately offset from the workpiece normal. The following angles represent the maximum misalignment that still allowed for a drilling operation to occur. An **X-axis Tilt** of  $\pm 6^\circ$  was applied around the pressurefoot's X-axis (perpendicular to the primary feed direction, causing an "up" or "down" tilt of the drill relative to the feed) as shown in Figures 6(b) and 6(c). Additionally, a **Y-axis Tilt** of  $+4^\circ$  was applied around the pressurefoot's Y-axis (sideways tilt) as shown in Figure 6(d). A  $-4^\circ$  tilt was not used as it prevented the tactile sensor from properly touching the workpiece.

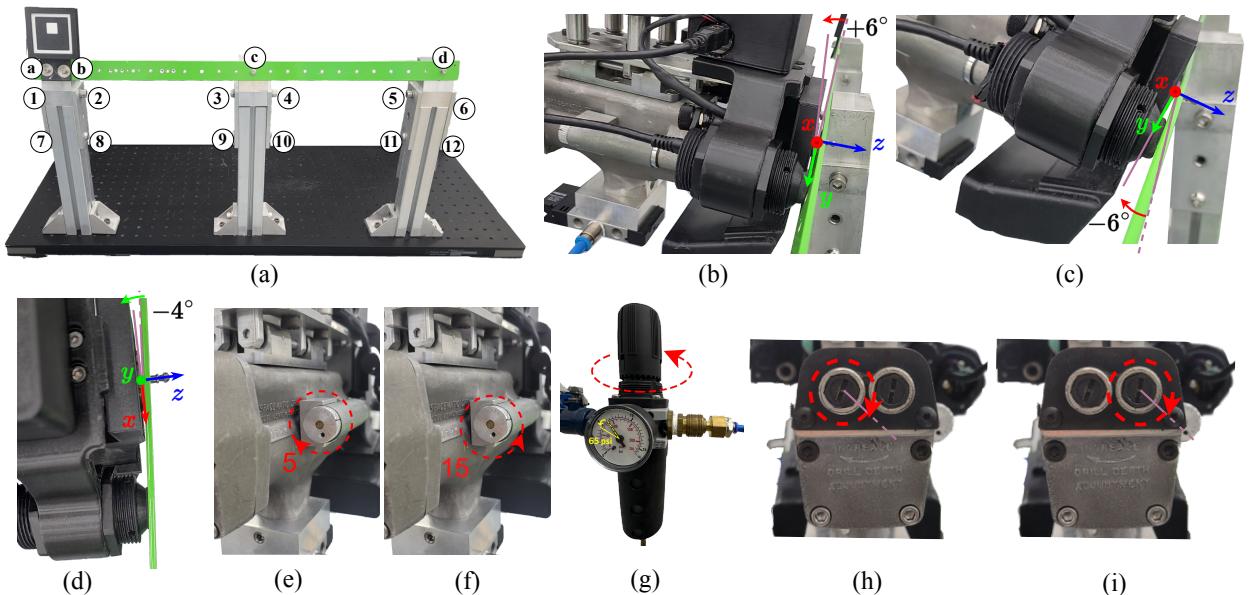
**Drill Feed Rate Variation.** The drill motor's feed rate setting was varied from the nominal value of 10. This included a **Higher Drill Feed Rate** by adjusting the knob to 5 (less damping, faster insertion), as seen in Figure 6(e), and a **Lower Drill Feed Rate** by adjusting the knob to 15 (more damping, slower insertion), as seen in Figure 6(f).

**Reduced Drilling Torque.** The air pressure supplied to the drill motor was reduced from the nominal 90 psi to simulate lower torque conditions. A lower pressure case was induced where the pressure was set to 65 psi as depicted in Figure 6(g).

**Incorrect Countersink Depth.** The drill motor's countersinking depth was altered from the nominal setting. This was done for both the **Left Spindle**, by adjusting the left bit forward by 0.8 mm (corresponding to one complete turn of the spindle adjustment thimble) as shown in Figure 6(h), and the **Right Spindle**, by adjusting the right bit forward by 0.8 mm as shown in Figure 6(i).

**Worn Drill Bit.** A drill bit that was deliberately worn was used. The drill bit was blunted by using it in drilling a scrap steel specimen continuously for few minutes and scratching the drill bit's cutting edge with the scrap specimen 15-20 times making about 10 cm scratches cutting edge until the cutting edge was worn out and became round.

Each fault condition was induced individually while other parameters were kept at their nominal healthy settings.



**Figure 6:** Adjustments used to induce specific fault conditions during data collection. (a) Unstable/loosened workpiece fixture, (b) X-axis tilt  $+6^\circ$ , (c) X-axis tilt  $-6^\circ$ , (d) Y-axis tilt  $-4^\circ$ , (e) Higher feed rate (Kinecheck adjustment knob rotated clockwise from nominal '10' to '5'), (f) Lower feed rate (Kinecheck adjustment knob rotated counter-clockwise from '10' to '15'), (g) Reduced drilling torque (air pressure regulator set to 65 psi), (h) Deeper left countersink (left bit advanced forward by 0.8 mm using spindle adjustment thimble), (i) Deeper right countersink (right bit advanced forward by 0.8 mm using spindle adjustment thimble).

### 2.3.3. Data Acquisition Procedure

Data was collected for a total of 1 nominal + 11 unique fault condition machining variations. For each case, a dedicated aluminum test coupon was used; each coupon had 11 holes to be drilled and countersunk.

The EVBTS sensor was active throughout each drilling and countersinking operation. The raw event stream from the DVXplorer Micro camera within the EVBTS was recorded. The data labeling method involved post-processing the recorded event streams. For each hole, a histogram of the number of events over time was generated. Then, the segment of the event stream corresponding to the actual material engagement phase (drilling and/or countersinking) was manually identified and segmented, excluding the approach and retraction phases. The set of frames derived from a single hole operation was then labeled with the corresponding ground truth machining condition.

## 2.4. Learning Model Architecture

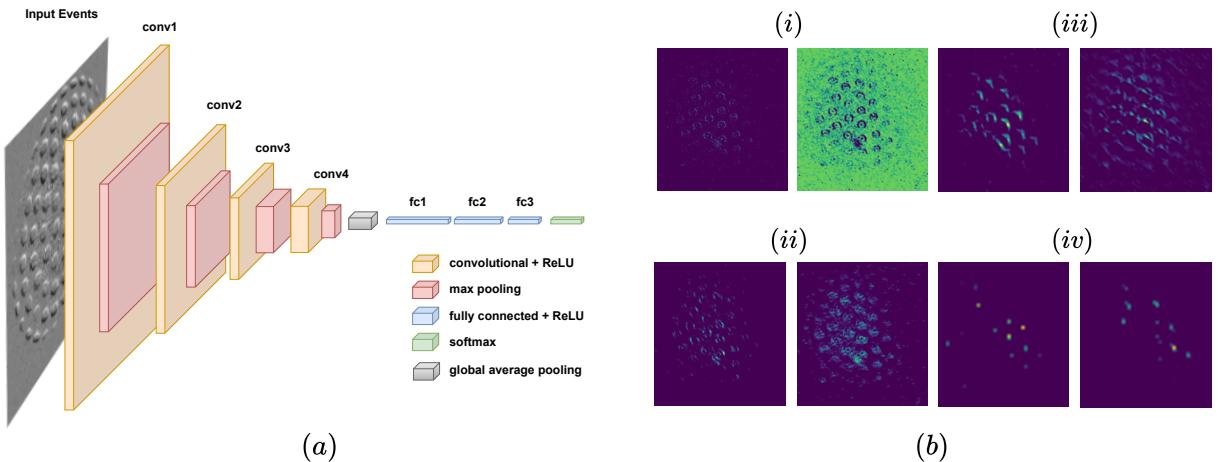
This section details the architecture of the Convolutional Neural Network (CNN) employed for fault monitoring of the drilling process. The design aims for a balance between effective feature extraction from the spatial patterns in the event count histograms and computational efficiency suitable for potential real-time application.

The CNN architecture takes a single-channel 2D event frame, processed from a raw event stream, as input. These input frames, representing event counts, are first cropped from the original sensor resolution of  $480 \times 640$  pixels to  $480 \times 440$  pixels to focus on the active region of the tactile sensor where the markers are visible. The lightweight

**Table 2**

Key Parameters of the Proposed CNN Model Architecture.

Component	Parameter Value / Description
Input Frame Size (Cropped)	$1 \times 480 \times 440$ (Channels $\times$ Height $\times$ Width)
<b>Convolutional Blocks (4 total)</b>	
For each block ( <i>Conv1</i> , <i>Conv2</i> , <i>Conv3</i> ).	
Output Channels	16
Kernel Size	$5 \times 5$
Activation	ReLU
Max Pooling Kernel/Stride	$2 \times 2 / 2$
For <i>Conv4</i> (final conv block):	
Output Channels	32
Kernel Size	$5 \times 5$
Activation	ReLU
Max Pooling Kernel/Stride	$2 \times 2 / 2$
Global Average Pooling Output	$1 \times 1$ (spatial)
<b>Fully Connected Head</b>	
FC1 Output Features / Activation	128 / ReLU
FC2 Output Features / Activation	64 / ReLU
FC3 Output Features	Number of Classes (12)

**Figure 7:** (a) Lightweight CNN architecture for fault monitoring (b) Visualization of selected feature map activations from (i) conv1 (ii) conv2 (iii) conv3 (iv) conv4.

network consists of a feature extraction backbone composed of four convolutional layers, each followed by ReLU activation and max pooling, and then a classification head comprising three fully connected layers. The number of classes corresponds to the total number of distinct healthy and unhealthy machining conditions (12 in our primary experiments). The key parameters of this architecture are summarized in Table 2. Moreover, this architecture is illustrated conceptually in Figure 7(a).

The CNN model was trained using a NVIDIA RTX3060 GPU. The dataset, consisting of event frames generated as described in Section 2.2.2 and labeled with their corresponding machining condition, was split into training and validation sets using an 80/20 ratio. To improve model generalization and robustness, various data augmentation was applied to the training data. The raw event count frames were subjected to the following transformations, each applied with a corresponding probability (denoted as  $p$ ). This involved random rotations within  $\pm 10^\circ$  ( $p_{rotate} = 0.5$ ), random brightness adjustments by a factor between 0.7 and 1.3 ( $p_{bright} = 0.7$ ), random translations of up to 10 pixels

horizontally and vertically ( $p_{translate} = 0.5$ ), and the addition of Gaussian noise (mean 0, standard deviation 2.0) to event counts ( $p_{noise} = 0.3$ ). Empty areas created by geometric transformations were filled with zeros, and pixel values were clipped to maintain valid ranges. No augmentations were applied to the validation set. While this augmentation strategy was critical for simulating variations such as minor shifts in sensor position, orientation, and illumination to ensure the model generalized well from the collected data, it cannot fully replicate the subtle complexities of real-world conditions. Expanding the dataset with more physical repetitions is therefore a valuable direction for future work to further enhance the model's robustness against these real-world variations.

The models were trained for 100 epochs using the Adam optimizer [46] with a learning rate of 0.001. The loss function employed was Cross-Entropy Loss, suitable for multi-class classification. A batch size of 32 was used. The dataset instances were downsampled if the number of frames exceeded a predefined limit (2500 frames).

During training, performance was monitored on both the training and validation sets. While overall accuracy provides a general measure of performance, it can be misleading in fault detection scenarios where "healthy" or nominal states may be more prevalent than specific fault conditions, potentially creating an imbalanced dataset. Therefore, we also tracked precision, recall, and particularly the F1-score. The F1-score, being the harmonic mean of precision and recall ( $F1 = 2 \cdot (\text{Precision} \cdot \text{Recall}) / (\text{Precision} + \text{Recall})$ ), offers a more robust evaluation, especially when the costs of both false positives (e.g., unnecessarily halting a nominal process, leading to productivity loss) and false negatives (e.g., failing to detect a genuine fault, potentially resulting in workpiece damage, accelerated tool wear, or even machine damage) are significant. For model selection, we specifically prioritized the F1-score for the nominal class when distinguished from all faulty classes combined. A high F1-score in this context indicates that the model is both effective at correctly identifying most nominal operations (high recall for nominal) and that its nominal predictions are reliable (high precision for nominal), thereby minimizing unnecessary interruptions while ensuring robust identification of the desired operational state. The model checkpoint achieving the best F1-score for the healthy class on the validation set was typically saved as the best model for further evaluation and real-time deployment.

## 2.5. Real-time Inference Pipeline

For real-time application in a robotic machining environment, the trained CNN model is integrated into a continuous monitoring pipeline that processes event streams from the EVBTS and provides immediate feedback on the drilling process health. The overall real-time inference flowchart is depicted in Figure 8.

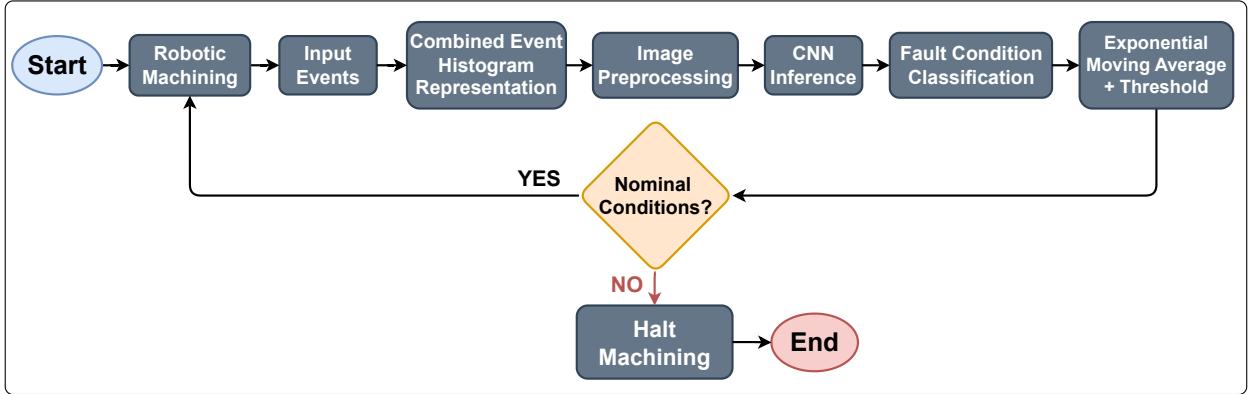
The real-time inference process unfolds in a series of steps. The process begins with the robotic system performing the drilling or countersinking operation. Simultaneously, the EVBTS continuously captures raw event data from the deformation of its tactile membrane caused by machining vibrations. As described in Section 2.2.2, these raw event streams are converted into Combined Event Histogram frames in real-time. This involves accumulating events within a sliding window of fixed event counts (e.g., 30,000 events with a 5,000 event stride). Each generated event frame then undergoes the same preprocessing steps as during training, including cropping and normalization, to prepare it for the CNN input.

The preprocessed event frame is fed into the trained CNN model, which outputs a vector of logits for each of the 12 machining conditions. The logits are transformed into probability vectors via a softmax function, representing the model's confidence for each class at that instant. To mitigate the impact of instantaneous noisy predictions and ensure robust decision-making, the probability vectors are subjected to an Exponential Moving Average (EMA). This temporal filter smooths the predictions over time, giving more weight to recent classifications. For a given class  $c$ , the EMA of its probability  $\bar{P}_c(t)$  at time  $t$  is calculated as  $\bar{P}_c(t) = \alpha \cdot P_c(t) + (1-\alpha) \cdot \bar{P}_c(t-1)$ , where  $P_c(t)$  is the instantaneous probability of class  $c$  at time  $t$ , and  $\alpha$  is the smoothing factor ( $0 < \alpha < 1$ ), which controls the responsiveness to new data.

A critical decision logic is then applied based on this smoothed data. The machining process is considered *not nominal* (i.e., faulty) and requires halting if the Exponential Moving Average of any specific fault condition class (e.g., loose workpiece, misalignment, worn drill bit) consistently crosses a predefined threshold, determined empirically. In our current implementation, if the moving average of any unhealthy class's probability exceeds 70% for a continuous duration, the system issues a halt command. This evaluation of the EMA serves as the nominal conditions check. If the moving average indicates that the machining conditions are **Nominal** (i.e., no fault condition's EMA crosses the threshold), the process loops back, and robotic machining continues. Conversely, if the moving average indicates **Non-Nominal Conditions**, the flow proceeds to halt machining. Upon detection of a non-nominal condition, a command is issued to the robotic system to immediately halt the machining operation, preventing potential damage

to the workpiece, tool, or robot. The process then terminates upon successful halting or completion of the machining operation.

This real-time pipeline ensures that the high-frequency tactile information from the EVBTS, combined with a robust learning-based classification and temporal filtering, enables instantaneous and reliable fault detection during robotic machining. The latency of this pipeline is discussed in Section 4.5 to evaluate the viability of this pipeline for real-time applications.



**Figure 8:** Flowchart of the real-time fault detection system for robotic machining.

## 2.6. Tactile Transfer Learning

To evaluate the adaptability of our EVBTS-based fault detection system, we employed a transfer learning strategy. This is critical for real-world deployment where sensors may undergo wear and tear, be replaced by new units with slight variations, or where the system needs to be adapted to new machining processes. This approach leverages our lightweight CNN model, pre-trained on the extensive original dataset (53,694 samples), and fine-tunes it on a new, smaller dataset of 17,445 samples (approximately 32.5% of the original size) collected after a period of sensor use.

The core idea behind our transfer learning strategy is to selectively unfreeze and fine-tune layers of the pre-trained lightweight CNN model based on the hierarchical nature of feature learning in CNNs. Early layers in a CNN typically learn generic, low-level features (e.g., edges, simple textures), while deeper layers learn more complex, task-specific features [47]. We hypothesized that changes due to sensor wear, tear, or minor shift in the position/orientation of the tactile sensing element's mounting ring would primarily affect the appearance of low-level features, while the underlying physics of how marker vibrations correlate with drilling faults would remain largely consistent, making higher-level learned features still relevant. An analysis of the feature maps learned by our pre-trained model, exemplified in Figure 7(b), informed our layer selection for fine-tuning.

The core idea behind our transfer learning strategy is to selectively unfreeze and fine-tune layers of the pre-trained lightweight CNN model, leveraging the hierarchical nature of feature learning in CNNs. Our strategy involved four main steps. First, the CNN model, pre-trained on the extensive original 12-class dataset, was loaded. The first two convolutional layers were then unfrozen and fine-tuned with a small learning rate ( $1 \times 10^{-4}$ ). Our visualizations (Figure 7(b)(i), 7(b)(ii)) show these layers learned fundamental, low-level features such as event occurrences, marker edges, and shapes. Fine-tuning these layers allowed the model to recalibrate to new, slightly different input characteristics resulting from sensor wear or minor shifts, without losing its core ability to detect marker-related events. Subsequent convolutional layers (conv3 and conv4, shown in Figure 7(b)(iii), 7(b)(iv)), and the first fully connected layer were kept frozen. These layers are hypothesized to learn more complex and abstract spatial patterns and feature combinations specific to the original dataset. Keeping them frozen preserved the robust, higher-level feature extraction capabilities learned from the extensive original dataset, as the overall task and fundamental physics of marker deformation remained consistent. Finally, the original final fully connected layers were initialized randomly and trained to learn the mapping from the extracted features to the specific classes and data characteristics of the new, smaller dataset. This strategy effectively balances adaptation to new sensor inputs and refined class definitions with the retention of robust, generalizable knowledge from the larger, original dataset.

This targeted strategy is highly significant as it provides a clear and data-efficient pathway for adapting the network. It enables not only the recalibration of a single tactile sensor as it accumulates wear but also suggests a method for adapting the network to variations across different tactile sensors or even extending the system to entirely different machining processes with minimal new data collection.

### 3. Experimental Setup

The experiments were conducted on a robotic drilling testbed designed to emulate aerospace manufacturing tasks. The overall system, depicted in Figure 9, comprises a collaborative robotic arm, a custom-designed drilling end-effector, and a specialized workpiece fixturing system. Both the robot and the workpiece fixture were mounted on separate optical breadboard tables to ensure stability and minimize external vibrations.

A Universal Robots UR10 collaborative robotic arm (payload capacity 10 kg, repeatability  $\pm 0.1$  mm) was utilized for all experiments. The UR10 was mounted on a fixed optical breadboard table and was responsible for the precise positioning of the end-effector and the execution of programmed drilling trajectories. The robot first localizes the workpiece and specific hole locations using the end-effector's vision system, then performs the drilling and countersinking operations. A detailed description of the robot's operational program for hole detection and drilling automation is presented in [36].

A custom-designed end-effector was mounted to the UR10's tool flange. As illustrated in Figure 9, the end-effector comprises of: the drilling system, EVBTS, and vision system. The drilling system consists of a pneumatic double-wing nutplate drill motor automated via a solenoid valve and features: adjustable feed rate (controlled by a Kinecheck feed control valve), effective torque regulation (controlled by air pressure from 65-90 psi), and countersink depth adjustment thimbles. A swarf collector is also included to manage debris, and the operational frame  $F_{Pressure\foot}$  is defined at the drill motor's pressure foot center. The EVBTS, detailed in Section 2.1.2, is the primary sensor for capturing process vibrations. Positioned approximately 65 mm laterally from the drill axis to avoid interference, it provides a fixed pose relative to the TCP ( $F_{TCP}$ ). Lastly, the vision system features an industrial RGB camera primarily for pre-process workpiece localization using an ArUco fiducial marker and accurate hole localization. Its pose relative to the TCP ( ${}^{TCP}T_{RGB}$ ) was determined via standard hand-eye calibration, enabling precise robot positioning based on observed features.

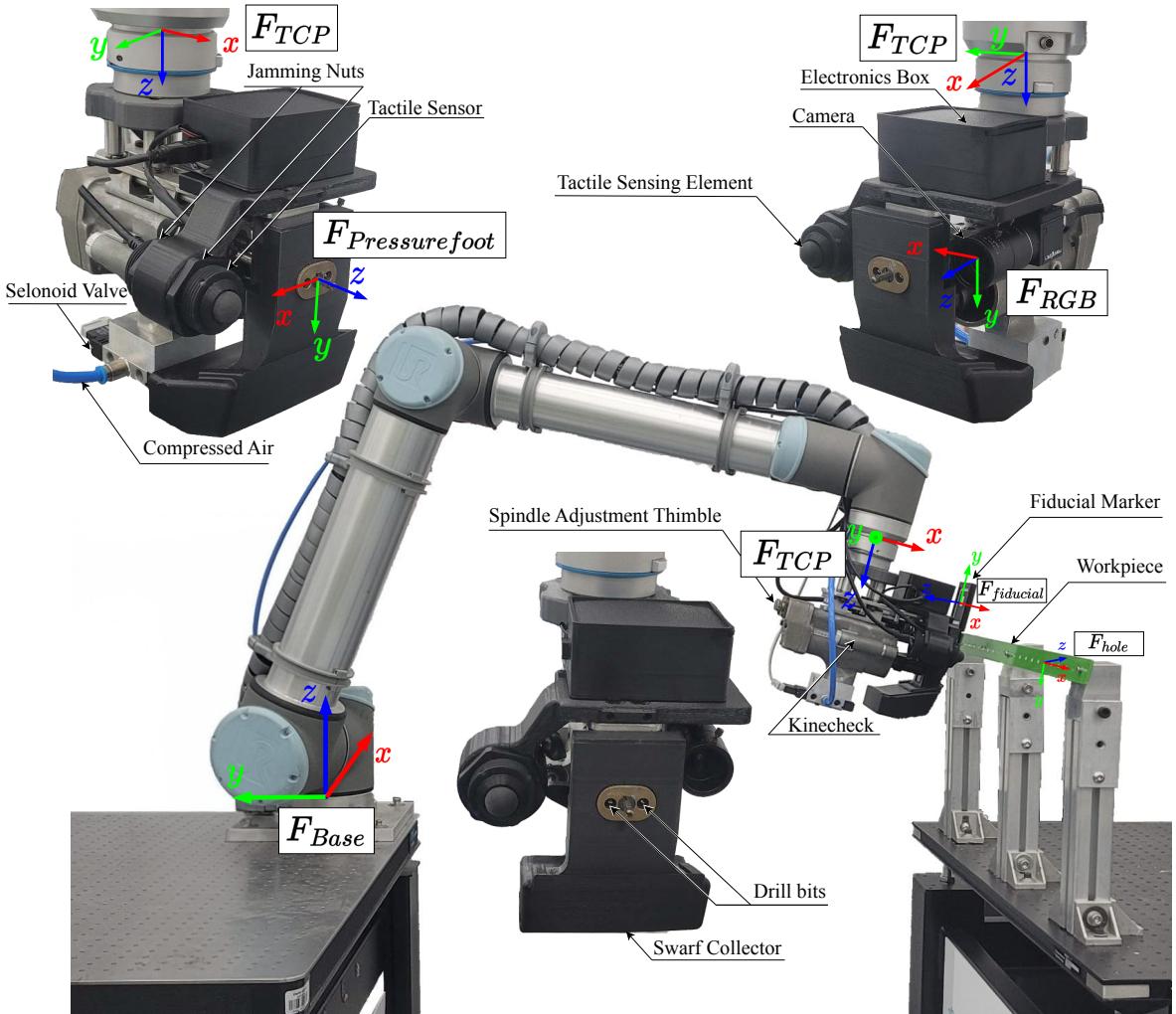
The workpiece fixture, illustrated in Figure 6(i), was designed to securely hold test coupons and replicate aspects of aerospace spar structures. It was mounted on a separate optical breadboard table. The fixture comprises three vertical T-slot aluminum extrusions, each topped with a precision-machined, threaded steel block. These blocks are angled at 15° to simulate a key geometric feature in aerospace spar profiles. The ArUco marker fiducial frame labeled  $F_{fiducial}$  represents the workpiece, and is established relative to the robot base frame  $F_{base}$ . Test coupons were fabricated from Aluminum 6000 Series sheets of 3 mm thickness. Each coupon was secured to the angled steel blocks at three points using four M4 bolts, ensuring rigid fixation under nominal conditions. An ArUco fiducial marker was affixed to the leading edge of each test coupon, aligned with the coupon's inclination. This marker enabled pre-drill pose estimation of the coupon using the end-effector's RGB camera, allowing the robot to accurately adjust its drilling path for each new coupon and ensure consistent alignment for the nominal cases.

## 4. Results and Discussion

### 4.1. Dataset Overview

Table 3 summarizes the number of hole repetitions and material-engagement durations for the twelve machining conditions: one baseline (nominal) condition and eleven induced fault conditions. Raw EVBTS event streams were segmented into classification samples by accumulating fixed-size windows of 30,000 events with 5,000-event overlap, to capture both transient and steady-state vibrations. This framing procedure yielded a total of 53,694 event-frame samples across all conditions.

We randomly stratified the set of 30,000-event frames into a training set (80 %) and a validation set (20 %), preserving the overall frequency of each class. We then applied a uniform threshold of 2,500 frames per class (see Table 3 for original and down-sampled counts). Any condition with >2,500 frames was randomly down-sampled, while classes below the threshold remained unchanged. After down-sampling, we trained on 22,782 frames and validated on 5,696 frames.



**Figure 9:** Experimental robotic drilling setup. A UR10 collaborative robot, mounted on an optical breadboard, carries a custom end-effector that integrates a pneumatic drill motor (pressure-foot reference  $F_{\text{Pressurefoot}}$ ), an EVBTS for vibration sensing, an RGB camera (frame  $F_{\text{RGB}}$ ) for ArUco-based workpiece localization (frame  $F_{\text{fiducial}}$ ). The drilling target is denoted by  $F_{\text{hole}}$ .

By first pooling all frames from all repetitions of a given condition before performing random sampling, we ensure that the resulting dataset maintains a representative temporal distribution of the entire drilling signature for each class, including its various sub-stages. Moreover, by balancing at the frame level, we prevent the classifier from simply exploiting the over-representation of long-duration classes and instead force it to focus on the shape of the markers borne from each corresponding vibration signature.

## 4.2. Sensor Data Insights

### 4.2.1. Spatial Vibration Signatures

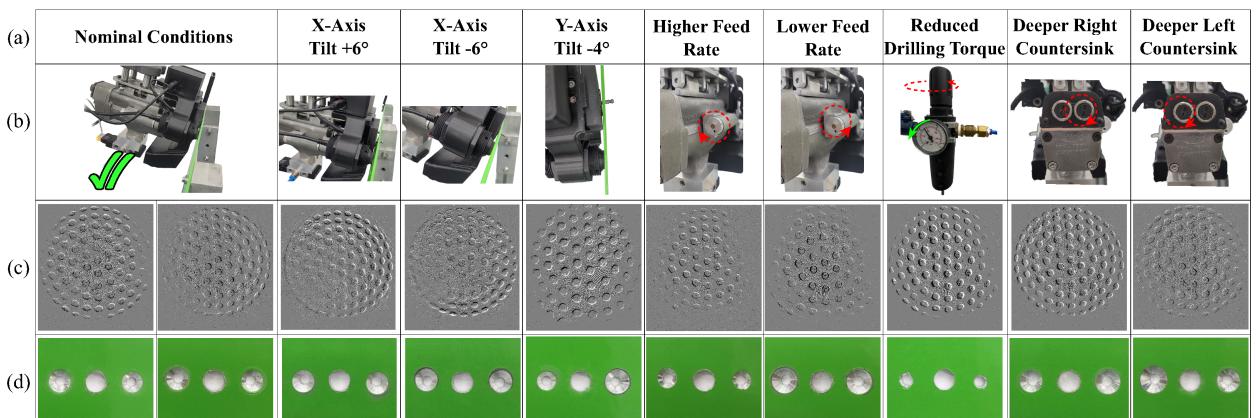
To illustrate how each fault condition affects both the EVBTS spatial response and the resulting hole geometry, Figure 10 shows, for each condition, the end-effector adjustment in 10(b), a representative Polarity Event Frame in 10(c), and the final hole geometry in 10(d). Note that while Polarity Event Frames are displayed here for their visual clarity, our final classification model utilizes Combined Event Frames, which, as demonstrated in the ablation study (Section 4.3.2), provide superior performance for the CNN. Under nominal drilling, events formed a concentric ring, reflecting uniform vibrations and resulting in a symmetrical countersink with clean edges. Tool misalignments, such as X-Axis Tilt ( $\pm 6^\circ$ ) and Y-Axis Tilt ( $-4^\circ$ ), caused event concentrations in specific quadrants,

**Table 3**

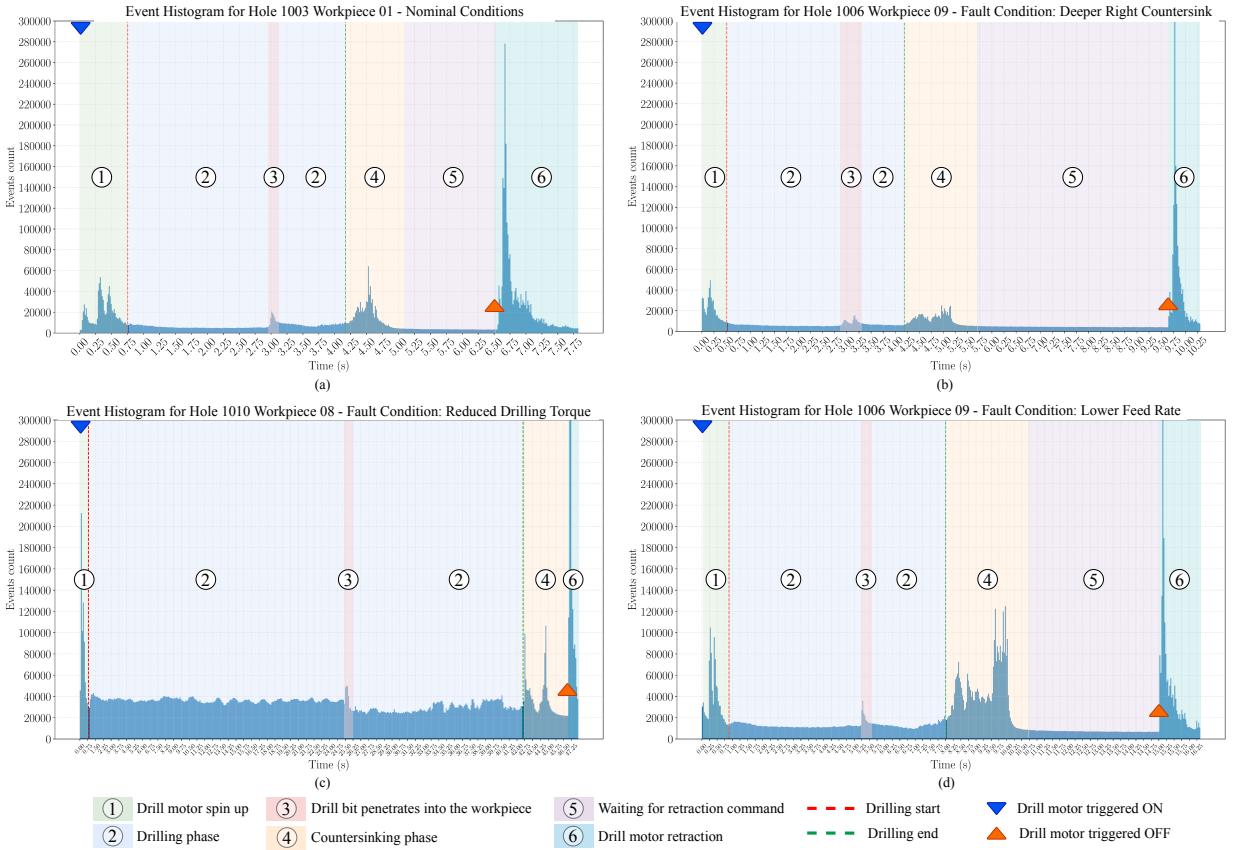
Dataset composition for robotic drilling conditions. This table summarizes all twelve machining conditions, one nominal and eleven induced fault conditions, showing, for each: the number of drilled holes, the approximate material-engagement duration range per hole, and the resulting event-frame sample counts both before and after applying a uniform down-sampling of 2,500 frames per class. By comparing original vs. down-sampled counts, the table documents our dataset balancing strategy for robust model training.

Index	Machining Condition	Approximate Duration per Hole (s)	Number of Repetitions	No. of frames (Original)	No. of frames (Down-sampled)
1	Healthy (Nominal) Drilling	4.0 – 6.0	28	7,974	2,500
2	Unstable Workpiece Fixture	3.5 – 6.0	11	3,569	2,500
3	Unstable/Loose End-Effector (Simulated)	4.0 – 5.0	10	3,040	2,500
4	Tool Misalignment: X-Axis Tilt (+6°)	4.5 – 6.0	11	2,959	2,500
5	Tool Misalignment: X-Axis Tilt (-6°)	4.5 – 6.0	11	4,237	2,500
6	Tool Misalignment: Y-Axis Tilt (+4°)	3.5 – 4.0	11	2,014	2,014
7	Higher Drill Feed Rate	2.0 – 3.0	11	1,464	1,464
8	Lower Drill Feed Rate	7.5 – 9.0	10	5,638	2,500
9	Reduced Drilling Torque (65 psi)	14.0 – 40.0	5	13,570	2,500
10	Incorrect Countersink Depth: Left Bit Forward (+0.8 mm)	4.0 – 5.0	10	2,630	2,500
11	Incorrect Countersink Depth: Right Bit Forward (+0.8 mm)	4.0 – 6.0	11	2,742	2,500
12	Worn Drill Bit	7.0 – 9.0	10	3,857	2,500

directly corresponding to off-center countersinks due to altered load distribution. Feed rate variations also influenced the patterns: a higher feed rate resulted in events concentrating near the dome center with fainter outer rings (shallow countersink), while a lower feed rate yielded a similar central focus but with higher event counts and sharper outer rings (deeper countersink). Reduced drilling torque led to thickened marker edges due to increased contact time and friction, producing a shallow countersink or merely a chamfered entry. Incorrect countersink depths (Deeper Left or Right Countersink) manifested as asymmetric countersink diameters, indicating closer or greater distance from the respective drill bit. These condition-specific spatial patterns demonstrate that the EVBTS captures precise, interpretable signatures for each fault condition.



**Figure 10:** Visual summary of evaluated machining conditions. (a) Machining Conditions (b) Corresponding end-effector adjustment (c) Resulting EVBTS Polarity Event Frame (used for visualization) (d) Images of the final, resulting drilled and countersunk holes.



**Figure 11:** EVBTS events histogram for four representative drilling trials. Total EVBTS event counts (ON + OFF) in 500 time bins for: (a) Nominal Drilling, (b) Deeper Right Countersink fault, (c) Reduced Drilling Torque fault, and (d) Lower Drill Feed Rate fault. The interval between the red and green dashed lines indicates the drilling phase used for analysis.

#### 4.2.2. Temporal Vibration Signatures

Having established how each condition imprints a unique spatial vibration pattern, we now examine the corresponding temporal signatures. Figure 11 presents event-count histograms over the full machining cycle for four representative cases, one nominal (a) and three fault conditions (b-d), revealing distinct, phase-specific features that further validate our segmentation and classification approach. Each cycle can be segmented into distinct phases, including 1) motor spin-up, 2) a steady drilling plateau, 3) a penetration spike, 4) a countersinking phase, and 5) drill retraction.

In terms of temporal signatures, distinct phase-specific features were observed in the event-count histograms for each condition. Under nominal drilling (Figure 11(a)), following the initial motor spin-up, the event count settled into a stable plateau, signifying steady cutting forces, interrupted only by a single spike upon workpiece penetration. Conversely, introducing an over-depth countersink (Figure 11(b)) yielded two distinct penetration spikes: the first occurring when the forward-offset right drill bit breached the material, and the second when the nominal-depth left bit followed. Reducing the drilling torque (Figure 11(c)) markedly elevated the steady drilling plateau to approximately twice its nominal level and significantly prolonged the drilling phase. This effect is attributed to increased tool-workpiece contact time and friction under sub-optimal torque, which intensifies tactile marker micro-deformations and generates additional asynchronous events. Similarly, slowing the feed rate (Figure 11(d)) also resulted in an elevated event-count plateau compared to the nominal case, though less pronounced than with reduced torque, and amplified the penetration spike. This increase arises because a slower feed rate prolongs the cutter's engagement with the workpiece per unit of material removed, leading to a greater number of movements in the tactile markers and, consequently, more events captured by the camera. These phase-specific deviations confirm that each fault mode encodes a unique, repeatable temporal signature on the event stream.

### 4.3. Ablation Studies

To validate our architectural choices for the proposed CNN model and the selected event representation, we conducted two sets of ablation studies. The first investigates the impact of removing specific layers from our proposed four-convolutional-block architecture. The second compares the performance of the full CNN architecture when trained on different input event representations. All models in these studies were trained for 30 epochs for comparative purposes, and performance was primarily evaluated based on the F1-score for the nominal class on the validation set, as well as overall validation accuracy.

#### 4.3.1. Architectural Ablation

We systematically evaluated the contribution of different components of the model architecture by training several ablated versions. Our baseline model architecture (referred to as Original in Figure 12(a) and Table 4) consists of four convolutional blocks followed by a global average pooling layer and three fully connected layers. The ablated versions were created by progressively removing convolutional layers from the deeper end of the network or by simplifying the fully connected head. Specifically, we tested the following model architectures: **NoConv234**: Only the first convolutional block was retained. **NoConv34**: Only the first two convolutional blocks were retained. **NoConv4**: The first three convolutional blocks were retained, removing the final convolutional block. **NoFC12**: The full four convolutional blocks were used, but the first two fully connected layers were removed, directly connecting the output of global average pooling (after flattening) to a single output fully connected layer. **NoFC2**: The full four convolutional blocks and fully connected layer 1 were used, but the second fully connected layer was removed, connecting FC1 directly to FC3.

The training loss curves for these architectural variants are presented in Figure 12(a). It is evident that reducing the depth of the convolutional backbone significantly impacts learning, with NoConv234 (only one convolutional block) exhibiting the highest training loss and slowest convergence. As more convolutional layers are added (NoConv34, NoConv4), the training loss decreases more rapidly and reaches lower final values, indicating improved feature extraction capabilities. The original proposed CNN model architecture demonstrates the best training loss profile among the convolutional ablations. Table 4a summarizes the key performance metrics for these architectural variants. The results clearly indicate that the Original architecture achieves the highest validation F1-score for the nominal class (94.70%) and the highest final validation accuracy (97.75%). Removing convolutional layers progressively degrades performance, with NoConv234 performing the worst. Simplifying the fully connected head by removing FC2 (NoFC2) results in a marginal decrease in performance, while removing both FC1 and FC2 (NoFC12) leads to a more substantial drop in the healthy F1-score. This suggests that while the four convolutional blocks are crucial for effective feature extraction, the depth of the fully connected head also contributes, albeit less, to the model's discriminative power. Our proposed architecture with four convolutional blocks and a three-layer fully connected head provides a good balance, achieving high accuracy and F1-score, justifying its selection for further experiments.

#### 4.3.2. Event Representation Ablation

To determine the most effective way to represent the event stream for our CNN model, we compared three different event frame generation strategies (as detailed in Section 2.2.2), all using the original proposed CNN model architecture.

Figure 12(b) shows the training loss curves for the model trained with these three different event representations. The Combined Event Frames representation demonstrates the fastest convergence and achieves the lowest training loss, suggesting it provides a highly learnable input for the CNN. Polarity Event Frames show slightly slower convergence and a higher final loss, while Time Surface Frames exhibit the slowest convergence and highest training loss among the three.

The performance metrics summarized in Table 4b further confirm these observations. Using Combined Event Frames yields the best validation healthy F1-score (96.60%) and the highest overall validation accuracy (97.01% at best, 96.92% final). Polarity Event Frames and Time Surface Frames result in lower F1-scores (90.12% and 89.35%, respectively) and lower overall accuracies. This suggests that for our task and architecture, the simple aggregation of all event activity (Combined Event Frames), which effectively highlights regions of marker movement, provides a more discriminative input than representations that explicitly encode polarity or fine-grained temporal recency or history within the frame. Therefore, Combined Event Frames were selected as the input representation for all other experiments in this study.

**Table 4**

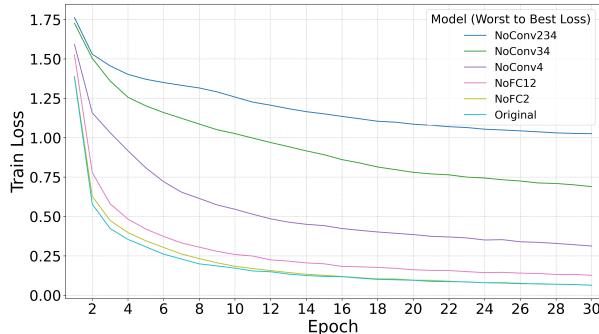
Performance metrics for ablation studies. Comparison of validation performance for (a) architectural ablations, and (b) event representation ablations

(a) Architectural Ablation

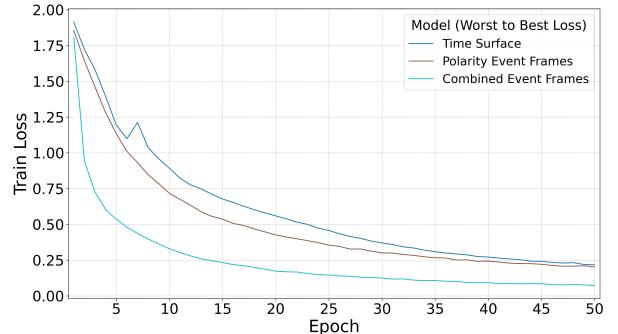
Model	Best Validation Nominal F1-Score	Final Validation Accuracy
<b>Original</b>	<b>94.70%</b>	<b>97.75%</b>
NoFC2	94.42%	97.51%
NoFC12	88.24%	92.99%
NoConv4	72.38%	88.73%
NoConv34	36.22%	77.03%
NoConv234	29.21%	62.18%

(b) Event Representation Ablation

Representation	Best Validation Healthy F1-Score	Best Validation Overall Accuracy
<b>Combined Events</b>	<b>96.60%</b>	<b>97.01%</b>
Polarity Events	90.12%	91.91%
Time Surface Events	89.35%	92.09%



(a)



(b)

**Figure 12:** Ablation study training loss curves. (a) Comparison of architectures with varying convolutional depth (b) Comparison of different event representations

#### 4.4. Accuracy of Recognition per Class

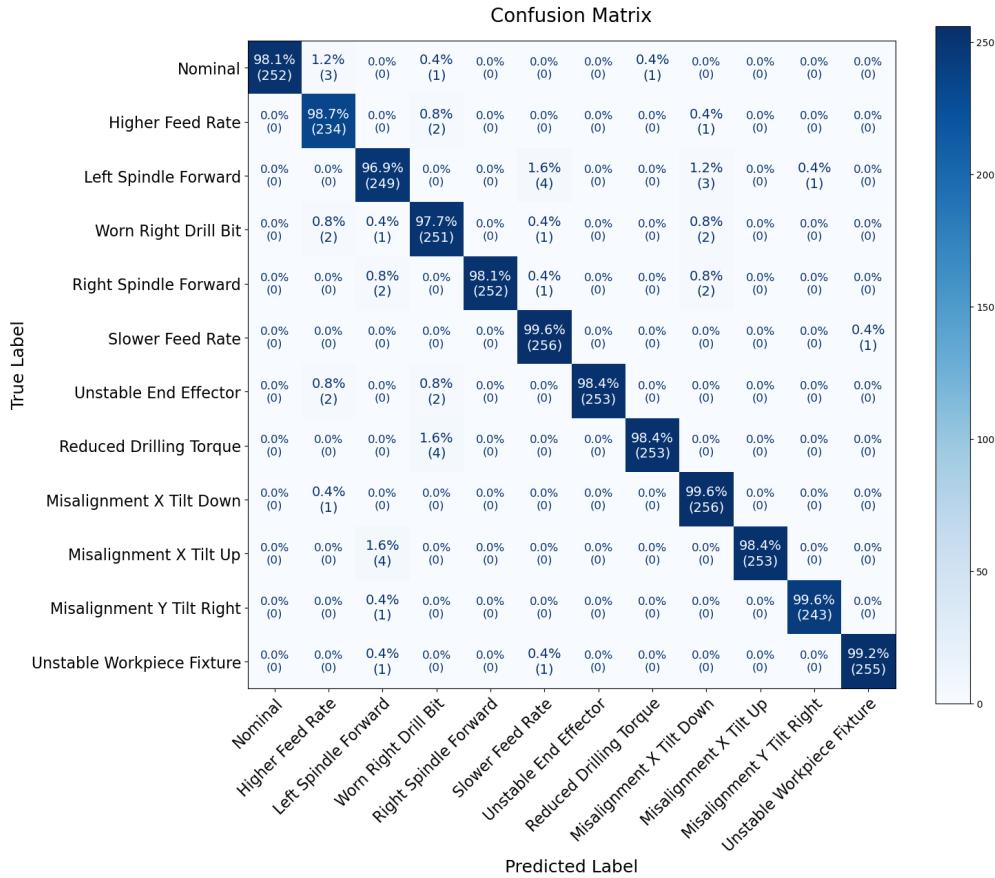
To gain a more granular understanding of the classification model's performance, we analyzed its ability to correctly classify each of the individual nominal and faulty machining conditions. This was evaluated on a held-out test set (10% of the original dataset), collected after training and was not used during training or model selection, ensuring a blind assessment of generalization performance.

Figure 13 presents the confusion matrix for the model's predictions on this test set. The matrix visualizes the true versus predicted labels, with the diagonal elements representing correctly classified instances for each class. The strong diagonal presence indicates high classification accuracy across most conditions. Minor confusions are observed, for instance, a small number of Nominal instances being misclassified as Higher Feed Rate and some Left Spindle Forward instances being misclassified as Slower Feed Rate. These minor off-diagonal values suggest some subtle similarities in the vibration signatures generated by these distinct conditions that occasionally challenge the classifier.

The model demonstrated high accuracy across all individual classes, with most exceeding 98%. The highest accuracies were achieved for: Lower Feed Rate (99.61%), Misalignment X Tilt Down +6° (99.61%), and Misalignment Y Tilt Right-4° (99.59%), indicating their highly distinct vibration signatures. The lowest accuracy was for Deeper Left Countersink at 96.89%, which aligns with the minor confusions observed in the confusion matrix, suggesting a subtle similarity in its vibration pattern to other conditions. Overall, the strong per-class performance underscores the model's capability to build a robust and discriminative feature space from the EVBTS data.

The overall test accuracy achieved by the CNN model on this unseen data was 98.56%, with a corresponding F1-score of 98.11%. This high overall accuracy, coupled with strong per-class performance, underscores the model's capability to effectively distinguish between various healthy and unhealthy drilling states based on the event frame representations derived from the EVBTS. The training history with final validation accuracy of 98.63% and healthy

F1-score of 98.70% on the validation split of the full training data indicated stable learning and good generalization, which is corroborated by these strong test set results.



**Figure 13:** Confusion matrix for the classification model on the held-out test set. Rows represent the true labels, and columns represent the predicted labels.

#### 4.5. Processing Time and Computational Efficiency

Real-time fault detection in robotic machining requires the monitoring system to operate with low latency and high computational efficiency. We evaluated these aspects for our proposed EVBTS-based CNN system, including both the event frame generation and the CNN inference. All latency measurements were conducted on a system equipped with an NVIDIA RTX3060 GPU.

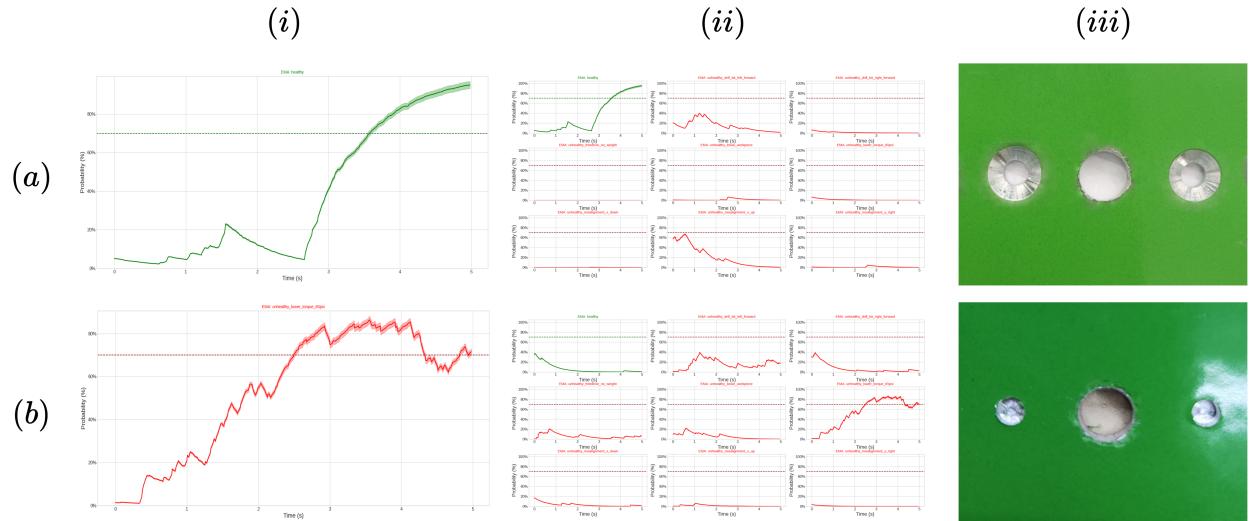
**Computational Complexity:** The proposed classification model's computational complexity was analyzed, providing insights into its parameter count and Multiply-Accumulate Operations (MACs) and Floating Point Operations (FLOPs). The network comprises a total of 39,340 trainable parameters. The total number of Multiply-Adds (MAdds), a common measure of computational workload, is approximately 0.514 GMacs, which translates to 1.045 GFLOPs. This relatively low complexity for a deep learning model is essential for achieving real-time performance on embedded or robotic platforms. The input size for the CNN inference (single event frame) is 0.84 MB, with the total estimated memory footprint (forward/backward pass, parameters) being around 36.09 MB.

**Real-time Latency Analysis:** The total end-to-end latency for our real-time fault detection pipeline is composed of two main components. The first is the Event-to-Frame Latency, which is the time required to accumulate and preprocess a 30,000-event frame, averaging 85.43 ms. The second is the CNN Inference Latency for a single frame, which averaged 11.22 ms on our hardware. Combining these two primary components for a single inference cycle, the total end-to-end latency from the arrival of raw events to a classification output is approximately  $85.43 \text{ ms} + 11.22 \text{ ms} = 96.65 \text{ ms}$ . This latency is well within the typical requirements for real-time process monitoring in manufacturing.

applications, where a response time of under 100-200 ms is often considered viable for detecting and mitigating faults [48]. The lightweight nature of the model architecture and the efficient event-to-frame conversion process are critical factors enabling this real-time performance.

#### 4.6. EMA-Based Real-time Fault Detection and Intervention Performance

To demonstrate the efficacy of the real-time fault detection and intervention pipeline detailed in Section 2.5, we analyzed the smoothed probability outputs for representative nominal and faulty drilling operations. Figure 14 illustrates the system's performance. For these tests, the Exponential Moving Average (EMA) smoothing factor was set to  $\alpha = 0.03$ , and the decision threshold was set at 70% probability. A halt command was triggered if any fault condition's EMA crossed this threshold and remained above it for one second or longer, a duration determined empirically to ensure robust decisions and prevent premature stops. The top row of Figure 14 shows a nominal (healthy) operation. The EMA of the nominal class probability (Figure 14(a)(i)) rises steadily, crossing the 70% threshold and remaining high, confirming the nominal state and allowing the process to continue. Crucially, the EMAs for all other fault classes remain low (Figure 14(a)(ii)), demonstrating the classifier's specificity. The result is a successfully completed, high-quality countersunk hole, as shown in Figure 14(a)(iii). Conversely, the bottom row of Figure 14 depicts a faulty operation with reduced drilling torque. The EMA for the Reduced Drilling Torque class (Figure 14(b)(i)) rapidly crosses the 70% threshold. After remaining above the threshold for more than one second, the system triggers a halt command. The EMAs for all other classes remain well below the threshold (Figure 14(b)(ii)), again confirming the correct fault is identified. The resulting workpiece (Figure 14(b)(iii)) shows an incomplete hole, where the machining operation was stopped mid-process. This intervention is critical: by halting the operation, the system prevents the creation of a fully formed but defective hole (like the drilled holes shown in Figure 10(d)), saving valuable machining time and preventing damage to a high-value workpiece. These results illustrate that the EMA filter effectively smooths predictions for stable decision-making, while the threshold and time-based logic provide a clear and reliable mechanism for intervention.



**Figure 14:** Real-time fault detection & intervention pipeline performance. The first row (a) shows a nominal (healthy) drilling operation, while the second row (b) shows a faulty Reduced Drilling Torque operation. Column (i) EMA of the true class probability over time. Column (ii) EMAs for nine other classes, demonstrating classifier specificity by remaining below the threshold. Column (iii) The resulting workpiece state.

#### 4.7. Transfer Learning Training and Results

The transfer learning process was conducted for 100 epochs using the Adam optimizer [46] and CrossEntropyLoss. The new dataset was split into 80% for training and 20% for validation. Similar to the original dataset, the frame

counts were downsampled accordingly and the same data augmentation was applied. The fine-tuned model achieved a validation accuracy of 93.49% and, importantly, a validation F1-score for the nominal class of 91.40%.

This performance on a significantly smaller dataset demonstrates the effectiveness of the transfer learning strategy. By selectively fine-tuning the early feature extractors and the initial part of the classifier while retraining a new head, the model successfully adapted to the new conditions while leveraging the rich feature representations learned from the original, more extensive dataset. This indicates that the core knowledge about vibration patterns is transferable, and targeted fine-tuning can account for sensor variations and different classification needs for other machining processes.

## 5. Conclusion

In this paper, we presented the first system that employs event-based vision-tactile sensing for real-time vibration-signature analysis in robotic machining. Specifically, we designed and developed a custom Event Vision-Based Tactile Sensor (EVBTS), integrated with a UR10 manipulator and pneumatic drill, to capture high-frequency vibration signatures during drilling and countersinking operations. Our rigorous experimental evaluation encompassed twelve distinct machining conditions (one nominal and eleven induced faults), resulting in the development of a comprehensive event vision-based dataset. This dataset, processed into 30,000-event windows, formed the basis for training a lightweight Convolutional Neural Network (CNN).

The findings of this paper establish a new paradigm for lightweight, data-efficient, and in-situ condition monitoring and fault diagnosis in robotic manufacturing. The novel design and application of the EVBTS, combined with a practical and validated end-to-end pipeline, provide a foundational and extendable framework for intelligent, data-driven manufacturing across a range of tasks. This includes drilling, deburring, and other contact-rich manufacturing processes relevant for high-value sectors like aerospace, where quantifiable, objective diagnostics are paramount.

The proposed EVBTS lays the foundation for intelligent, data-driven robotic machining and has significant potential in other domains. Its compact design is well-suited for integration into smart, handheld tools. By leveraging edge deployment for onboard processing, such tools could offer real-time diagnostic feedback, augmenting a skilled operator's perception to detect subtle vibrations indicative of process quality, thus reducing reliance on subjective experience. This paradigm also extends to applications like medical robotics, where instruments equipped with such sensors could provide critical feedback on tissue properties through micro-vibration analysis during surgical procedures.

To enable these broader applications, future research will focus on several key areas. For instance, edge deployment as aforementioned. The complete processing pipeline should be implemented on a resource-constrained system, with verification of real-time latency and power consumption to ensure practical applicability in robotic platforms. Advanced data processing algorithms for the perception of raw event data should be employed to enhance temporal resolution and reduce latency. Finally, physics-based modeling and simulation will include the development of a high-fidelity Finite Element Analysis (FEA) model of the sensing chain to simulate how vibrations propagate through the workpiece, are transferred into the elastomer, and ultimately influence the movement of internal markers.

## CRediT authorship contribution statement

**Eslam Sherif:** Writing - review & editing, Writing - original draft, Resources, Methodology, Software, Investigation, Formal analysis, Conceptualization. **Akram Khairi:** Writing - review & editing, Writing - original draft, Resources, Methodology, Software, Investigation, Formal analysis, Conceptualization. **Hussain Sajwani:** Methodology, Software, Investigation, Writing - review & editing. **Abdullah Solayman:** Writing - original draft, Writing - review & editing, Conceptualization. **Abdallah Mohammad Alkilany:** Methodology, Investigation, Formal analysis. **Ahmed Awadalla:** Validation, Investigation, Conceptualization. **Mohamad Halwani:** Writing - review & editing, Methodology, Conceptualization. **Laith AbuAssi:** Validation, Project administration, Methodology. **Dewald Swart:** Project administration, Formal analysis, Conceptualization. **Abdulla Ayyad:** Writing - review & editing, Project administration, Formal analysis, Conceptualization. **Yahya Zweiri:** Writing - review & editing, Project administration, Supervision, Funding acquisition, Conceptualization.

## References

- [1] J. Rymarczyk, Technologies, opportunities and challenges of the industrial revolution 4.0: Theoretical considerations, *Entrepreneurial Business and Economics Review* 8 (1) (2020) 185–198. doi:10.15678/EBER.2020.080110.

- URL <https://doi.org/10.15678/EBER.2020.080110>
- [2] J. Leng, J. Guo, J. Xie, X. Zhou, A. Liu, X. Gu, D. Mourtzis, Q. Qi, Q. Liu, W. Shen, L. Wang, Review of manufacturing system design in the interplay of industry 4.0 and industry 5.0 (part ii): Design processes and enablers, *Journal of Manufacturing Systems* 79 (2025) 528–562. doi:10.1016/j.jmsy.2025.02.005.  
URL <http://dx.doi.org/10.1016/j.jmsy.2025.02.005>
- [3] M. Javaid, A. Haleem, R. P. Singh, S. Rab, R. Suman, Exploring impact and features of machine vision for progressive industry 4.0 culture, *Sensors International* 3 (2022) 100132. doi:<https://doi.org/10.1016/j.sintl.2021.100132>.  
URL <https://www.sciencedirect.com/science/article/pii/S266635112100053X>
- [4] M. Javaid, A. Haleem, R. P. Singh, S. Rab, R. Suman, Significance of sensors for industry 4.0: Roles, capabilities, and applications, *Sensors International* 2 (2021) 100110. doi:<https://doi.org/10.1016/j.sintl.2021.100110>.  
URL <https://www.sciencedirect.com/science/article/pii/S2666351121000310>
- [5] S. M. Noble, M. Mende, D. Grewal, A. Parasuraman, The fifth industrial revolution: How harmonious human-machine collaboration is triggering a retail and service [r]evolution, *Journal of Retailing* 98 (2) (2022) 199–208. doi:10.1016/j.jretai.2022.04.003.  
URL <https://doi.org/10.1016/j.jretai.2022.04.003>
- [6] J. Leng, D. Wang, W. Shen, X. Li, Q. Liu, X. Chen, Digital twins-based smart manufacturing system design in industry 4.0: A review, *Journal of Manufacturing Systems* 60 (2021) 119–137. doi:<https://doi.org/10.1016/j.jmsy.2021.05.011>.  
URL <https://www.sciencedirect.com/science/article/pii/S0278612521001151>
- [7] M. Ramadan, A. Youssef, A. Ayyad, L. AbuAssi, O. A. Hay, M. Salah, B. Moyo, Y. Zweiri, Y. Abdulrahman, Vision-guided robotic system for aero-engine inspection and dynamic balancing, *Scientific Reports* 14 (1) (2024) 30742. doi:10.1038/s41598-024-80540-w.  
URL <https://doi.org/10.1038/s41598-024-80540-w>
- [8] B. Caiazzo, M. D. Nardo, T. Murino, A. Petrillo, G. Piccirillo, S. Santini, Towards zero defect manufacturing paradigm: A review of the state-of-the-art methods and open challenges, *Computers in Industry* 134 (2021) 103548. doi:10.1016/j.compind.2021.103548.  
URL <https://doi.org/10.1016/j.compind.2021.103548>
- [9] Z. Lai, C. Yang, S. Lan, L. Wang, W. Shen, L. Zhu, Bearingfm: Towards a foundation model for bearing fault diagnosis by domain knowledge and contrastive learning, *International Journal of Production Economics* 275 (2024) 109319. doi:10.1016/j.ijpe.2024.109319.  
URL <http://dx.doi.org/10.1016/j.ijpe.2024.109319>
- [10] A. Diez-Olivan, J. Del Ser, D. Galar, B. Sierra, Data fusion and machine learning for industrial prognosis: Trends and perspectives towards industry 4.0, *Information Fusion* 50 (2019) 92–111. doi:10.1016/j.inffus.2018.10.005.  
URL <http://dx.doi.org/10.1016/j.inffus.2018.10.005>
- [11] A. Shokrani, H. Dogan, D. Burian, T. D. Nwabueze, P. Kolar, Z. Liao, A. Sadek, R. Teti, P. Wang, R. Pavel, T. Schmitz, Sensors for in-process and on-machine monitoring of machining operations, *CIRP Journal of Manufacturing Science and Technology* 51 (2024) 263–292. doi:<https://doi.org/10.1016/j.cirpj.2024.05.001>.  
URL <https://www.sciencedirect.com/science/article/pii/S1755581724000592>
- [12] U. H. Shah, R. Muthusamy, D. Gan, Y. Zweiri, L. Seneviratne, On the Design and Development of Vision-based Tactile Sensors, *Journal of Intelligent & Robotic Systems* 102 (4) (2021) 82. doi:10.1007/s10846-021-01431-0.  
URL <https://doi.org/10.1007/s10846-021-01431-0>
- [13] W. Yuan, S. Dong, E. H. Adelson, Gelsight: High-resolution robot tactile sensors for estimating geometry and force, *Sensors* 17 (12) (2017). doi:10.3390/s17122762.  
URL <https://www.mdpi.com/1424-8220/17/12/2762>
- [14] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, N. F. Lepora, The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies, *Soft Robotics* 5 (2) (2018) 216–227, pMID: 29297773. arXiv:<https://doi.org/10.1089/soro.2017.0052>, doi:10.1089/soro.2017.0052.  
URL <https://doi.org/10.1089/soro.2017.0052>
- [15] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, D. Jayaraman, R. Calandra, Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation, *IEEE Robotics and Automation Letters* 5 (3) (2020) 3838–3845. doi:10.1109/LRA.2020.2977257.
- [16] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, D. Scaramuzza, Event-based vision: A survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44 (1) (2022) 154–180. doi:10.1109/TPAMI.2020.3008413.
- [17] M. Salah, A. Ayyad, M. Humais, D. Gehrig, A. Abusafieh, L. Seneviratne, D. Scaramuzza, Y. Zweiri, E-calib: A fast, robust, and accurate calibration toolbox for event cameras, *IEEE Transactions on Image Processing* 33 (2024) 3977–3990. doi:10.1109/TIP.2024.3410673.
- [18] N. Zhang, C. Ye, L. Peng, Y. Tao, Novel array eddy current sensor with three-phase excitation, *IEEE Sensors Journal* 19 (18) (2019) 7896–7905. doi:10.1109/JSEN.2019.2919661.  
URL <http://dx.doi.org/10.1109/JSEN.2019.2919661>
- [19] R. Li, H. Ma, R. Wang, H. Song, X. Zhou, L. Wang, H. Zhang, K. Zeng, C. Xia, Application of unsupervised learning methods based on video data for real-time anomaly detection in wire arc additive manufacturing, *Journal of Manufacturing Processes* 143 (2025) 37–55. doi:10.1016/j.jmapro.2025.03.113.  
URL <http://dx.doi.org/10.1016/j.jmapro.2025.03.113>
- [20] S. Rezvani, C.-J. Kim, S. S. Park, J. Lee, Simultaneous clamping and cutting force measurements with built-in sensors, *Sensors* 20 (13) (2020) 3736. doi:10.3390/s20133736.  
URL <http://dx.doi.org/10.3390/s20133736>

- [21] L. Cen, S. N. Melkote, J. Castle, H. Appelman, A wireless force-sensing and model-based approach for enhancement of machining accuracy in robotic milling, *IEEE/ASME Transactions on Mechatronics* 21 (5) (2016) 2227–2235. doi:10.1109/tmech.2016.2567319.  
URL <http://dx.doi.org/10.1109/TMECH.2016.2567319>
- [22] Effect of Sensor Failure on Dynamometry Calibration, Vol. Volume 1: Acoustics, Vibration, and Phononics of ASME International Mechanical Engineering Congress and Exposition. arXiv:<https://asmedigitalcollection.asme.org/IMECE/proceedings-pdf/IMECE2020/84478/V001T01A009/6636593/v001t01a009-imece2020-24264.pdf>, doi:10.1115/IMECE2020-24264.  
URL <https://doi.org/10.1115/IMECE2020-24264>
- [23] B. Denkena, K. M. Litwinski, D. Brouwer, H. Boujnah, Design and analysis of a prototypical sensory z-slide for machine tools, *Production Engineering* 7 (1) (2012) 9–14. doi:10.1007/s11740-012-0419-1.  
URL <http://dx.doi.org/10.1007/s11740-012-0419-1>
- [24] J. Huang, D. T. Pham, C. Ji, Z. Zhou, Smart cutting tool integrated with optical fiber sensors for cutting force measurement in turning, *IEEE Transactions on Instrumentation and Measurement* 69 (4) (2020) 1720–1727. doi:10.1109/tim.2019.2916240.  
URL <http://dx.doi.org/10.1109/TIM.2019.2916240>
- [25] F. Xie, J. Ren, Z. Chen, Q. Feng, Vibration-displacement measurements with a highly stabilised optical fiber michelson interferometer system, *Optics amp; Laser Technology* 42 (1) (2010) 208–213. doi:10.1016/j.optlastec.2009.06.010.  
URL <http://dx.doi.org/10.1016/j.optlastec.2009.06.010>
- [26] B. Maslah, M. B. S. H., M. S. H., M. W., M. H., M. B., M. S., M. T., Battery impedance spectroscopy for li-ion batteries: A comparative study, *Measurement* 218 (2023) 113168. doi:10.1016/j.measurement.2023.113168.  
URL <https://doi.org/10.1016/j.measurement.2023.113168>
- [27] Z. Fan, Y. Xu, P. Liu, X. Li, R. Zhang, T. Yang, W. Wu, Y. Li, L. Chen, G. Zhang, Ssa: Microsecond-level clock synchronization based on machine learning for iot devices, *IEEE Transactions on Instrumentation and Measurement* 72 (2023) 1–12. doi:10.1109/TIM.2023.3298422.  
URL <https://doi.org/10.1109/TIM.2023.3298422>
- [28] N. Jaber, S. Ilyas, O. Shekhah, M. Eddaoudi, M. I. Younis, Resonant gas sensor and switch operating in air with metal-organic frameworks coating, *Journal of Microelectromechanical Systems* 27 (2) (2018) 156–163. doi:10.1109/JMEMS.2018.2794546.  
URL <https://ieeexplore.ieee.org/document/8283611>
- [29] C. Ge, E. Cretu, Polymeric piezoelectric accelerometers with high sensitivity, broad bandwidth, and low noise density for organic electronics and wearable microsystems, *Microsystems amp; Nanoengineering* 10 (1) (May 2024). doi:10.1038/s41378-024-00704-6.  
URL <http://dx.doi.org/10.1038/s41378-024-00704-6>
- [30] E. K. Fredrickson, W. S. D. Wilcock, M. J. Harrington, G. Cram, J. Tilley, D. Martin, J. Burnett, The self-calibrating tilt accelerometer: A method for observing tilt and correcting drift with a triaxial accelerometer, *Earth and Space Science* 11 (11) (Oct. 2024). doi:10.1029/2024ea003909.  
URL <http://dx.doi.org/10.1029/2024EA003909>
- [31] A. Solayman, H. Sajwani, O. Abdul Hay, R. Chang, L. AbuAssi, A. Ayyad, Y. Zweiri, Y. Abdul Samad, Piezosight: Coupling vision based tactile sensor neural network processing and piezoresistive stimulation for enhanced piezo-vision hybrid sensing, *Materials amp; Design* 253 (2025) 113899. doi:10.1016/j.matdes.2025.113899.  
URL <http://dx.doi.org/10.1016/j.matdes.2025.113899>
- [32] M. Alshawabkeh, H. Alagi, S. E. Navarro, C. Duriez, B. Hein, H. Zangl, L.-M. Faller, Highly stretchable additively manufactured capacitive proximity and tactile sensors for soft robotic systems, *IEEE Transactions on Instrumentation and Measurement* 72 (2023) 1–10. doi:10.1109/tim.2023.3250232.  
URL <http://dx.doi.org/10.1109/TIM.2023.3250232>
- [33] X. Huang, Z. Ma, W. Xia, L. Hao, Y. Wu, S. Lu, Y. Luo, L. Qin, G. Dong, A high-sensitivity flexible piezoelectric tactile sensor utilizing an innovative rigid-in-soft structure, *Nano Energy* 129 (2024) 110019. doi:10.1016/j.nanoen.2024.110019.  
URL <http://dx.doi.org/10.1016/j.nanoen.2024.110019>
- [34] I. M. Zaid, M. Halwani, A. Ayyad, A. Imam, F. Almaskari, H. Hassanin, Y. Zweiri, Elastomer-based visuotactile sensor for normality of robotic manufacturing systems, *Polymers* 14 (23) (2022). doi:10.3390/polym14235097.  
URL <https://www.mdpi.com/2073-4360/14/23/5097>
- [35] A. Solayman, M. Halwani, I. M. Zaid, R. B. Ambade, R. Chang, A. Ayyad, F. AlMaskari, Y. Zweiri, Y. Abdul Samad, Mechanical behaviour, contact pose estimation, and finite element analysis of vision based tactile sensors fabricated by molding and direct ink writing: A comparative study, *Advanced Engineering Materials* 26 (22) (May 2024). doi:10.1002/adem.202400630.  
URL <http://dx.doi.org/10.1002/adem.202400630>
- [36] M. Halwani, A. Ayyad, L. AbuAssi, Y. Abdulrahman, F. Almaskari, H. Hassanin, A. Abusafieh, Y. Zweiri, A novel vision-based multi-functional sensor for normality and position measurements in precise robotic manufacturing, *Precision Engineering* 88 (2024) 367–381. doi:<https://doi.org/10.1016/j.precisioneng.2024.02.015>.  
URL <https://www.sciencedirect.com/science/article/pii/S0141635924000515>
- [37] M. Mirzaee, H. Huang, W. Yuan, Gelbelt: A vision-based tactile sensor for continuous sensing of large surfaces, *IEEE Robotics and Automation Letters* 10 (2) (2025) 2016–2023, publisher Copyright: © 2016 IEEE. doi:10.1109/LRA.2025.3527306.
- [38] A. Ayyad, M. Halwani, D. Swart, R. Muthusamy, F. Almaskari, Y. Zweiri, Neuromorphic vision based control for the precise positioning of robotic drilling systems, *Robotics and Computer-Integrated Manufacturing* 79 (2023) 102419. doi:10.1016/j.rcim.2022.102419.  
URL <http://dx.doi.org/10.1016/j.rcim.2022.102419>
- [39] M. Salah, A. Ayyad, M. Ramadan, Y. Abdulrahman, D. Swart, A. Abusafieh, L. Seneviratne, Y. Zweiri, High speed neuromorphic vision-based inspection of countersinks in automated manufacturing processes, *Journal of Intelligent Manufacturing* 35 (7) (2024) 3067–3081.

- doi:10.1007/s10845-023-02187-0.  
URL <https://doi.org/10.1007/s10845-023-02187-0>
- [40] H. Sajwani, A. Ayyad, Y. Alkendi, M. Halwani, Y. Abdulrahman, A. Abusafieh, Y. Zweiri, Tactigraph: An asynchronous graph neural network for contact angle prediction using neuromorphic vision-based tactile sensing, Sensors 23 (14) (2023) 6451. doi:10.3390/s23146451. URL <http://dx.doi.org/10.3390/s23146451>
- [41] M. Salah, I. Mohamed Zaid, M. Halwani, H. Sajwani, A. Solayman, A. Ayyad, R. Azzam, A. Abusafieh, Y. Zweiri, Neutac: Zero-shot sim2real measurement for neuromorphic vision-based tactile sensors, IEEE Transactions on Instrumentation and Measurement 73 (2024) 1–15. doi:10.1109/TIM.2024.3451581.
- [42] N. Funk, E. Helmut, G. Chalvatzaki, R. Calandra, J. Peters, Evtac: An event-based optical tactile sensor for robotic manipulation, IEEE Transactions on Robotics 40 (2024) 3812–3832. doi:10.1109/TRO.2024.3428430.
- [43] B. Ward-Cherrier, N. Pestell, N. F. Lepora, Neurotac: A neuromorphic optical tactile sensor applied to texture recognition, in: 2020 IEEE International Conference on Robotics and Automation (ICRA), 2020, pp. 2654–2660. doi:10.1109/ICRA40945.2020.9197046.
- [44] A. Solayman, M. Halwani, I. M. Zaid, R. B. Ambade, R. Chang, A. Ayyad, F. AlMaskari, Y. Zweiri, Y. Abdul Samad, Mechanical behaviour, contact pose estimation, and finite element analysis of vision based tactile sensors fabricated by molding and direct ink writing: A comparative study, Advanced Engineering Materials 26 (22) (2024) 2470059. arXiv:<https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/adem.202470059>, doi:<https://doi.org/10.1002/adem.202470059>
- [45] F. Baghaei Naeini, A. M. AlAli, R. Al-Husari, A. Rigi, M. K. Al-Sharman, D. Makris, Y. Zweiri, A novel dynamic-vision-based approach for tactile sensing applications, IEEE Transactions on Instrumentation and Measurement 69 (5) (2020) 1881–1893. doi:10.1109/TIM.2019.2919354.
- [46] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization (2017). arXiv:1412.6980.  
URL <https://arxiv.org/abs/1412.6980>
- [47] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks?, in: Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, K. Weinberger (Eds.), Advances in Neural Information Processing Systems, Vol. 27, Curran Associates, Inc., 2014. URL [https://proceedings.neurips.cc/paper\\_files/paper/2014/file/532a2f85b6977104bc93f8580abbb330-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2014/file/532a2f85b6977104bc93f8580abbb330-Paper.pdf)
- [48] R. Teti, K. Jemielniak, G. O'Donnell, D. Dornfeld, Advanced monitoring of machining operations, CIRP Annals 59 (2) (2010) 717–739. doi:<https://doi.org/10.1016/j.cirp.2010.05.010>.  
URL <https://www.sciencedirect.com/science/article/pii/S0007850610001976>