Article

Transitioning from Simulation to Reality: Applying Chatter Detection Models to Real-World Machining Data

Matthew Alberts [1,†,¶], Sam St. John [2,†,¶], Simon Odie [3,†,‡,¶], Anahita Khojandi [5,†,¶], Bradley Jared [6,†], Tony Schmitz [7,§,‡], Jaydeep Karandikar [8,‡], and Jamie Coble [4,‖,¶,*],

[1] malberts@vols.utk.edu;
[2] stjohn3@vols.utk.edu;
[3] sodie@vols.utk.edu;
[4] jamie@utk.edu;
[5] khojandi@utk.edu;
[6] bhjared@utk.edu;
[7] tony.schmitz@utk.edu;
[8] karandikarjm@ornl.gov
* Correspondence: jamie@utk.edu
† Department of Industrial and Systems Engineering, University of Tennessee, 851 Neyland Drive, Knoxville, TN, 37996,USA
‡ Manufacturing Science Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN, 37830, USA
§ Department of Mechanical, Aerospace, and Biomedical Engineering, University of Tennessee, 1512 Middle Drive, Knoxville, TN, 37996, USA
‖ Department of Nuclear Engineering, University of Tennessee, 863 Neyland Drive, Knoxville, TN, USA
¶ These authors contributed equally to this work.

Abstract: Chatter, a self-excited vibration phenomenon, is a critical challenge in high-speed machining operations, affecting tool life, product surface quality, and overall process efficiency. While machine learning models trained on simulated data have shown promise in detecting chatter, their real-world applicability remains uncertain due to discrepancies between simulated and actual machining environments. This study bridges the gap by validating a ML-based chatter detection system using both simulated and real-world data. We apply a Random Forest classification model trained on over 140,000 simulated machining datasets, incorporating techniques like Operational Modal Analysis (OMA), Receptance Coupling Substructure Analysis (RCSA), and Transfer Learning (TL) to adapt the model for real-world operational data. The model is validated against 1,600 real-world machining datasets, achieving an accuracy of 86.1%, with strong precision and recall scores. Our results demonstrate the model's robustness and potential for practical implementation in industrial settings, highlighting challenges such as sensor noise and variability in machining conditions. This work advances the use of predictive analytics in machining processes, offering a data-driven solution to improve manufacturing efficiency through more reliable chatter detection.

Keywords: real-world, predictive analytics, vibration analysis, chatter, stability, additive manufacturing

Citation: Alberts, M.; St. John, S.; Odie, S.; Khojandi, A.; Jared, B.; Schmitz, T.; Karandikar, J.; Coble, J. Transitioning from Simulation to Reality: Applying Chatter Detection Models to Real-World Machining Data. *Journal Not Specified* **2024**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

1. Introduction

Chatter, a self-excited vibration phenomenon occurring during machining processes, poses significant challenges to manufacturing efficiency and product quality [1,2]. It manifests as unwanted oscillations between the cutting tool and the workpiece, leading to poor surface finishes, dimensional inaccuracies, increased tool wear, and potential damage to machine components [3]. Detecting and mitigating chatter is crucial for maintaining high precision and productivity in modern machining operations.

Traditional methods for chatter detection have relied on analytical models and signal processing techniques, which often require extensive domain knowledge and are limited by the assumptions inherent in their formulations [4]. These methods may not generalize well across different machining setups or accommodate the variability present in real-world operations [5].

In recent years, machine learning (ML) has emerged as a powerful tool for chatter detection, offering data-driven approaches that can learn complex patterns from vast amounts of data [6]. ML models, such as Random Forest classifiers and neural networks, have shown promise in identifying chatter by analyzing vibration signals and other sensor data [7]. These models can adapt to different machining conditions and provide real-time monitoring capabilities, enhancing predictive maintenance strategies by up to 40% [8].

In our previous work [9], we developed a Random Forest classifier trained on extensive simulated machining datasets to predict chatter occurrences. The model demonstrated high accuracy on simulated data and environments, indicating its potential for effective chatter detection. Building upon this foundation, we enhanced the model by incorporating additional simulated data and advanced modeling techniques in [10], further improving its predictive performance and generalizability.

However, models trained solely on simulated data may not fully capture the complexities and variabilities of real-world machining processes. Factors such as sensor noise, environmental conditions, and machine-specific characteristics can affect the performance of these models when applied outside the simulated environment [11]. Therefore, it is essential to validate and adapt these models using real-world data to ensure their practical applicability and robustness in operational settings.

This study aims to bridge the gap between simulation and practice by applying the previously developed models to real-world machining datasets. By evaluating the models' performance on actual vibration data collected from operational milling machines, we seek to assess their effectiveness, identify challenges in transitioning from simulated to real-world data, and provide insights into enhancing model robustness for practical applications.

1.1. Previous Work

In our prior research, we focused on developing ML models for chatter detection in machining processes using simulated data. In [9], we introduced a Random Forest classifier trained on a comprehensive dataset generated from simulations of milling operations. The simulated data captured various machining conditions, including different spindle speeds, feed rates, and depths of cut, providing a rich environment for the model to learn the complex patterns associated with chatter occurrences.

The initial model demonstrated promising results, achieving high accuracy in predicting chatter within the simulated environment. The use of Random Forests allowed us to effectively handle high-dimensional feature spaces and provided insights into feature importance, which is crucial for understanding the factors contributing to chatter [12].

Building upon this foundation, our subsequent study [10] expanded the dataset to include over 140,000 simulated instances and incorporated advanced modeling techniques such as Operational Modal Analysis (OMA), Transfer Learning (TL), and Receptance Coupling Substructure Analysis (RCSA). The integration of OMA enabled the extraction of modal parameters from the system's operational responses, enhancing the model's ability to capture the dynamic behavior of the machining process [13]. TL was employed to improve the model's adaptability to new and unseen machining conditions, addressing the challenge of varying operational environments [14]. RCSA provided a deeper understanding of the dynamic interactions within the tool-holder assembly, contributing to more accurate predictions of machining stability [15].

The enhanced model exhibited substantial improvements in predictive accuracy and robustness, particularly in scenarios involving complex machining conditions. However, both studies were conducted using simulated data, which, while comprehensive, may not fully represent the variability and unpredictability inherent in real-world machining processes.

Transitioning from simulation to real-world applications presents challenges due to differences in data distributions, noise levels, and unmodeled dynamics [11]. Models trained exclusively on simulated data may experience performance degradation when applied to operational environments where factors such as sensor noise, machine wear, and environmental conditions introduce additional complexity [16].

Recognizing these limitations, the current study aims to validate the models developed in our previous work using real-world machining data. By applying the models to actual vibration signals collected from operational milling machines, we seek to assess their practical applicability, identify any performance gaps, and refine the models to enhance their robustness in real-world settings.

1.1.1. Simulation-Based Models

Simulation-based models play a crucial role in the development and testing of machine learning (ML) models for chatter detection in machining processes. Simulated data provide a controlled and flexible environment where parameters such as spindle speeds, cutting depths, feed rates, and tool geometries can be varied systematically. This capability allows researchers to study the effects of different machining conditions comprehensively and gather insights that would be difficult, costly, or risky to obtain through real-world experiments.

One of the primary advantages of using simulation-based models is the ability to generate extensive datasets that cover a wide range of operational conditions. For instance, in our prior research, we simulated diverse cutting scenarios to capture data reflecting stable and unstable machining conditions. This approach enabled the training of ML models capable of recognizing complex patterns associated with chatter onset and predicting the stability of machining operations under varying conditions [17] [18].

Parameters such as spindle speed are essential for simulating the natural frequency response of the machine-tool system, as specific speeds can either suppress or amplify vibration tendencies, affecting chatter stability [19] [20]. Depth of cut is another critical parameter, influencing the magnitude of cutting forces and the resulting system dynamics. Simulated data that include variations in cutting depth provide valuable insights into the threshold beyond which machining becomes unstable, helping to create accurate stability lobe diagrams [21].

Additionally, incorporating feed rate and tool geometry into simulations helps to capture the interaction between the tool and workpiece under different operational settings. Tool geometry, including features like rake angle and edge radius, impacts the material removal process and can alter the vibration characteristics of the system [11]. By simulating these variations, researchers can evaluate how subtle changes in machining conditions affect chatter occurrence, enhancing the robustness of predictive ML models [22].

The ability to simulate a variety of machining conditions is invaluable for developing models that generalize well across different scenarios. However, it is essential to recognize that simulated environments, while comprehensive, do not capture the full range of variability present in real-world machining [23]. Factors such as sensor noise, machine wear, environmental fluctuations, and material inconsistencies introduce additional complexities that can affect model performance [24].

1.2. Problem Statement

While our previous studies have demonstrated the potential of ML models, specifically Random Forest classifiers, in predicting chatter using simulated data [9,10], a critical gap remains in validating these models with real-world machining data. Simulated datasets, despite their controlled conditions and comprehensive coverage of various machining scenarios, may not fully encapsulate the complexities and uncertainties inherent in actual machining environments [11].

Data collected from operational machining processes are characterized by factors such as sensor noise, machine tool wear, environmental variations, and inconsistencies in material properties [8]. These factors introduce discrepancies between simulated and real-world data distributions, which can adversely affect the performance of models trained exclusively on simulations [16]. Additionally, unmodeled dynamics and unforeseen interactions in physical systems may lead to challenges that are not present or are oversimplified in simulations [25].

The transition from simulated to real-world data poses several challenges:

1. **Data Discrepancies:** Differences in data distributions between simulated and real-world datasets can lead to reduced model performance due to issues like covariate shift and sample selection bias [26].
2. **Sensor Noise and Data Quality:** Real-world data are susceptible to various types of noise and artifacts that are typically absent in simulated data, necessitating advanced preprocessing and noise reduction techniques [24].
3. **Feature Relevance:** Features that are significant in simulated environments may not hold the same importance in real-world settings, requiring reevaluation of feature selection and extraction methods [27].
4. **Model Generalization:** Ensuring that the model generalizes well to unseen real-world data is essential for practical applicability, highlighting the need for model adaptation and validation strategies [14].

Addressing these challenges is crucial for the successful deployment of chatter detection models in industrial settings. Without validation on real-world data, the models' effectiveness remains uncertain, limiting their practical utility. Therefore, there is a pressing need to evaluate the performance of our previously developed models using actual machining data and to refine them as necessary to enhance their robustness and generalizability in operational environments.

This study seeks to fill this gap by:

- Collecting and processing 1,600 real-world machining datasets encompassing a variety of operational conditions.
- Applying the previously trained models to these datasets to assess their predictive performance.
- Identifying discrepancies between simulated and real-world data and their impact on model accuracy.
- Proposing solutions to address these discrepancies, including model adaptation techniques and enhanced feature extraction methods.

By tackling these challenges, we aim to bridge the gap between simulation and reality, providing insights into the practical applicability of ML models for chatter detection and contributing to the advancement of predictive maintenance in machining processes. It is important to note that the real-world data collected for this study does not represent the full complexity of a complete machine shop environment, as it was obtained from a relatively controlled and clean setting.

1.3. Objectives

The primary objective of this study is to validate and assess the practical applicability of ML models developed using simulated machining data when applied to real-world machining environments. Specifically, we aim to:

1. **Apply Existing Models to Real-World Data:** Utilize the Random Forest classifiers previously developed in our work [9,10] to predict chatter occurrences using a newly collected dataset of 1,600 real-world machining operations.
2. **Evaluate Model Performance:** Assess the predictive accuracy, precision, recall, F1-score, and AUC-ROC of the models when applied to real-world data.
3. **Compare:** Compare the above results with the performance achieved on simulated data.

4. **Identify and Analyze Discrepancies:** Investigate any discrepancies in model performance between simulated and real-world data, analyzing factors such as sensor noise, data variability, and differences in feature distributions.
5. **Enhance Model Robustness:** Propose and implement strategies to improve the model robustness and generalizability in real-world applications. This may include advanced data preprocessing techniques, feature engineering, or model adaptation methods such as domain adaptation [28].
6. **Provide Practical Insights:** Offer insights into the challenges and considerations involved in transitioning ML models from simulated to operational environments, contributing to the development of more reliable chatter detection systems in the machining industry.

By achieving these objectives, this study seeks to bridge the gap between simulation and practice, enhancing the reliability of ML models for chatter detection in real-world machining processes. The outcomes are expected to inform best practices for deploying predictive models in industrial settings, ultimately contributing to improved manufacturing efficiency and product quality.

1.4. Contributions

This study makes several key contributions to the field of machining chatter detection and predictive maintenance:

1. **Validation of Simulation-Trained Models on Real-World Data:** We provide a comprehensive evaluation of ML models trained on simulated machining data when applied to real-world machining operations. This validation assesses the models' effectiveness in practical settings, addressing a significant gap in existing research [29].
2. **Analysis of Transition Challenges from Simulation to Reality:** By identifying and analyzing the discrepancies between simulated and real-world data, we shed light on the challenges inherent in transferring models across domains. This includes addressing issues related to data distribution differences, noise levels, and feature relevance [30].
3. **Enhancement of Model Robustness through Adaptation Techniques:** We propose and implement strategies to improve the robustness and generalizability of the models in real-world applications. This includes the use of advanced data preprocessing, feature engineering, and model adaptation methods to mitigate the impact of domain discrepancies [31].
4. **Practical Insights for Industrial Implementation:** The findings offer valuable insights for practitioners in the machining industry, providing guidance on deploying ML models for chatter detection in operational environments. This contributes to the development of more reliable and efficient predictive maintenance systems [32].
5. **Contribution to Machine Learning Methodologies:** From a methodological perspective, the study advances the understanding of how ML models can be adapted and validated across different data domains, contributing to the broader field of TL and domain adaptation in industrial applications [33].

By achieving these contributions, the study not only demonstrates the practical applicability of simulation-trained models but also provides a framework for future research and development in the field of machining process monitoring and control.

2. Literature Review

2.1. Chatter Detection in Machining

Chatter adversely affects surface finish, dimensional accuracy, tool life, and overall productivity [19]. The detection and mitigation of chatter have been subjects of extensive research over the past several decades. Traditional methods for chatter detection primarily involve analytical and signal processing techniques applied to real-world data collected during machining operations.

Early approaches to chatter detection focused on stability analysis using analytical models based on machining dynamics [1]. The regenerative chatter theory introduced by Tobias and Fishwick [34] provided a foundational understanding of how vibrations from previous cutting passes influence current ones, leading to instability. Tlusty [35] further developed methods to predict stability lobes, enabling the selection of cutting parameters that avoid chatter.

Signal processing techniques have been widely used to detect chatter by analyzing vibration signals from sensors mounted on machine tools. Methods such as time-domain analysis, frequency-domain analysis, and time-frequency analysis have been employed to identify chatter signatures [2]. For instance, Fourier Transform (FT) and Short-Time Fourier Transform (STFT) have been utilized to observe changes in the frequency spectrum indicative of chatter onset [36]. Wavelet Transform (WT) methods have also been applied to capture transient features of chatter in the time-frequency domain [37].

However, these traditional methods often require expert knowledge to interpret signals and may not generalize well across different machining conditions or machine tools [38]. The complexity and variability of real-world machining environments pose challenges for these approaches, limiting their effectiveness in industrial applications.

Recent advancements have seen the emergence of ML techniques as powerful tools for chatter detection. ML models can learn complex patterns from large datasets, making them well-suited to handle the non-linear and stochastic nature of chatter phenomena [39]. Supervised learning algorithms, such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN), have been applied to classify machining states and predict chatter occurrences based on features extracted from sensor data [40] [8].

Deep learning models, including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have also been explored for their ability to automatically extract hierarchical features from raw data [41]. These models have demonstrated improved performance in chatter detection tasks, particularly when dealing with large and complex datasets [42].

Despite the promise of ML methods, challenges remain in their application to real-world machining data. Issues such as data scarcity, imbalance, noise, and variability can affect model training and performance [43]. Moreover, models trained on specific machines or conditions may not generalize well to other settings, necessitating strategies like TL and domain adaptation to enhance their applicability [14].

This study builds upon this body of work by applying ML models developed using simulated data to real-world machining environments. By addressing the challenges associated with transitioning from simulation to reality, we aim to contribute to the advancement of chatter detection methodologies and their practical implementation in industry.

2.2. Machine Learning in Machining Processes

The application of ML in machining processes has gained significant attention in recent years, driven by the increasing availability of data and advancements in computational capabilities [6]. ML techniques offer powerful tools for modeling complex, non-linear relationships inherent in machining operations, enabling improved process monitoring, fault diagnosis, and predictive maintenance [44].

2.2.1. Supervised Learning Methods

Supervised learning algorithms have been widely used for tool condition monitoring and fault diagnosis in machining. Support Vector Machines (SVM) have been employed to classify tool wear states based on features extracted from vibration signals, acoustic emissions, and cutting forces [45]. Decision Trees and Random Forests have also been applied to predict surface roughness and tool wear, leveraging their ability to handle high-dimensional data and provide insights into feature importance [46].

Artificial Neural Networks (ANN) have been utilized for modeling and prediction tasks in machining processes due to their capacity to approximate complex functions [47]. For instance, ANNs have been trained to predict cutting forces, surface finish, and dimensional accuracy based on input parameters such as spindle speed, feed rate, and depth of cut [48].

2.2.2. Unsupervised Learning and Clustering

Unsupervised learning methods, including clustering algorithms like K-means and hierarchical clustering, have been used to detect anomalies and patterns in machining data without the need for labeled datasets [49]. These techniques facilitate the identification of novel fault conditions and the segmentation of machining states based on similarities in data features [50].

2.2.3. Deep Learning Approaches

Deep learning, a subset of ML involving neural networks with multiple layers, has shown promise in automatically extracting hierarchical features from raw sensor data [51]. Convolutional Neural Networks (CNN) have been applied to time-series data for tool condition monitoring, achieving higher accuracy compared to traditional feature-based methods [52]. Recurrent Neural Networks (RNN), particularly Long Short-Term Memory (LSTM) networks, have been effective in modeling temporal dependencies in sequential data, improving predictions of tool wear progression [53].

Autoencoders and Generative Adversarial Networks (GAN) have been explored for unsupervised feature learning and anomaly detection in machining processes [54]. These models can learn compressed representations of data, highlighting essential features that contribute to fault detection and process optimization.

2.2.4. Challenges and Considerations

Despite the advancements, several challenges persist in applying ML to machining processes. Data quality and availability are critical issues; acquiring sufficient labeled data for supervised learning can be difficult due to the cost and time associated with experiments [55]. Additionally, variability in machining conditions and equipment can affect model generalization, necessitating strategies such as TL and domain adaptation [56].

The interpretability of complex models, particularly deep learning networks, is another concern, as understanding the decision-making process is important for gaining trust in industrial applications [57]. For instance, consider a manufacturing facility that employs a deep neural network to predict chatter events based on multifaceted vibration and acoustic data. While the model may achieve high accuracy in identifying chatter, the inherent "black-box" nature of deep learning makes it challenging for engineers to discern which specific features or patterns are driving the predictions. This lack of transparency can hinder the adoption of such models, as operators and decision-makers may be reluctant to rely on predictions without comprehending the underlying rationale. To bridge this trust gap, efforts are being made to develop explainable AI techniques, such as Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) [58] [59]. These methods provide detailed insights into feature importance and model behavior, enabling practitioners to understand and validate the model's predictions. By enhancing interpretability, explainable AI fosters greater confidence and facilitates more informed decision-making in industrial settings.

2.2.5. Application to Chatter Detection

In the context of chatter detection, ML models have been applied to classify stable and unstable cutting conditions. Features extracted from vibration signals, such as statistical measures, frequency components, and wavelet coefficients, serve as inputs to classifiers like SVM, ANN, and Random Forests [60]. Deep learning models have also been employed to automatically learn features relevant to chatter from raw sensor data, enhancing detection accuracy [61].

Our previous studies [9,10] contribute to this body of work by developing ML models using simulated data for chatter prediction. The current study extends this research by applying these models to real-world machining data, addressing the challenges of model transferability and generalization.

2.3. Simulation vs. Real-World Data

Simulated data plays a vital role in machining research, providing a controlled environment to develop and test models for chatter detection and machining stability [2]. Through simulations, researchers can systematically vary machining parameters and study their effects without the costs and risks associated with physical experiments [8].

However, transitioning models developed on simulated data to real-world applications introduces challenges due to inherent differences between simulated and actual machining data [11]. Real-world data are influenced by factors such as sensor noise, environmental variability, machine wear, and unmodeled dynamics, which are often not fully captured in simulations [24].

One significant difference is the presence of measurement noise and disturbances in real-world data [8]. Real sensors are subject to limitations in accuracy and are affected by external factors, leading to noisy data that can obscure important patterns necessary for accurate chatter detection [43].

Moreover, real-world machining processes involve complex interactions and variability that are difficult to model entirely in simulations [2]. Factors such as tool wear, material inconsistencies, and temperature fluctuations can significantly affect the process and are often simplified in simulated environments [8].

The discrepancies between simulated and real-world data can result in differences in data distributions, feature importance, and ultimately, model performance [26]. Models trained exclusively on simulated data may not generalize well to real-world scenarios due to domain shift, where the statistical properties of the training data differ from those of the application environment [14].

To address these challenges, techniques such as TL and domain adaptation have been proposed to enhance model transferability between domains [33]. These methods aim to adjust models to account for differences in data distributions, improving their performance on real-world data [11].

In this study, we focus on applying ML models developed on simulated data to real-world machining datasets, exploring the impact of domain discrepancies on model performance, and investigating strategies to enhance model robustness and adaptability in practical applications.

2.4. Gap Identification

Despite significant advancements in the application of ML techniques for machining process monitoring and chatter detection, a critical gap exists in the validation of simulation-trained models using real-world machining data. As discussed in previous sections, many studies have developed models based on simulated data or controlled laboratory experiments, demonstrating promising results in predicting machining stability and detecting chatter [9,10,61].

However, these models often face challenges when applied to real-world industrial environments. Factors such as sensor noise, machine tool variability, environmental conditions, and unmodeled dynamics introduce complexities that are not fully captured in simulated datasets [11,43]. The discrepancies between simulated and actual machining data can lead to reduced model performance, limiting the practical applicability of these predictive models [26].

Currently, there is a lack of empirical studies that:

1. **Systematically Evaluate Model Transferability:** Few studies have systematically assessed how models trained on simulated data perform when applied to real-world machining operations, identifying the factors that contribute to performance degradation.
2. **Address Domain Discrepancies:** Limited research has been conducted on developing and implementing strategies to mitigate the impact of domain discrepancies between simulated and real-world data in the context of machining processes.
3. **Provide Practical Implementation Guidelines:** There is a need for comprehensive guidelines and best practices for adapting and deploying simulation-trained ML models in industrial settings for chatter detection and machining process monitoring.

This gap hinders the advancement of smart manufacturing initiatives and the adoption of advanced predictive maintenance strategies in the machining industry. By addressing this gap, researchers and practitioners can enhance the reliability and robustness of ML models, facilitating their integration into real-world manufacturing processes and contributing to improved productivity and product quality.

Therefore, this study aims to bridge this gap by validating the performance of ML models developed using simulated data on real-world machining datasets, analyzing the challenges encountered, and proposing solutions to enhance model robustness and applicability in industrial environments.

3. Methodology

3.1. Real-World Data Collection

To validate the ML models developed in our previous studies, we collected real-world machining data using a custom-built three-axis Computer Numeric Controlled (CNC) milling machine. The machine was equipped with a Marposs MEMS (Micro-Electro-Mechanical Systems) vibration sensor for high-fidelity data acquisition. An image of the experimental setup is shown in Figure 1.

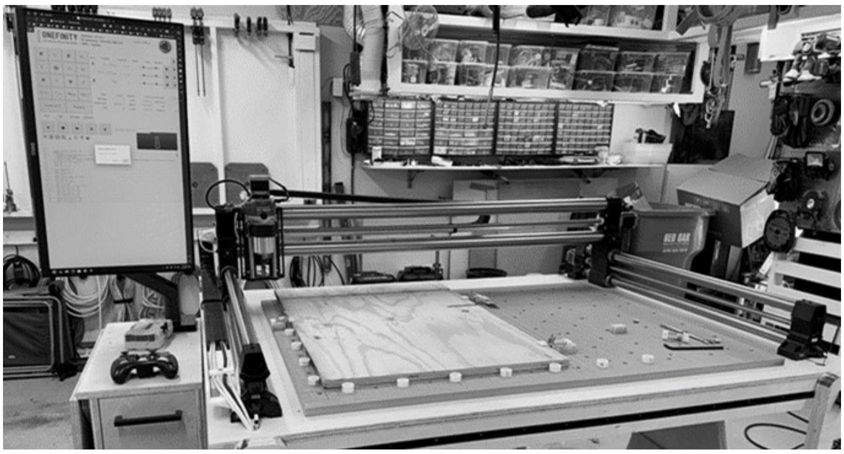

Figure 1. Three-axis CNC milling machine equipped with a Marposs MEMS vibration sensor.

The CNC milling machine was designed to perform precise machining operations necessary for data collection. It features three linear axes (X, Y, and Z) with high-precision ball screws and linear guides, ensuring accurate and repeatable movements. The spindle is capable of variable speeds ranging from 1,000 to 16,000 RPM, allowing for a wide range of cutting conditions.

3.1.1. Sensor Integration and Data Acquisition

A Marposs MEMS vibration sensor was installed on the spindle housing to capture vibrational responses during machining. The sensor provides high-resolution measurements of acceleration in three axes, enabling the detection of subtle vibrations associated with chatter phenomena. The sensor specifications include a measurement range of ± 16 g and a frequency bandwidth up to 6 kHz.

Data acquisition was performed using a high-speed data acquisition (DAQ) system interfaced with the CNC controller. The DAQ system sampled the vibration signals at a rate of 20 kHz to ensure adequate capture of the dynamic events associated with chatter. The vibration data were synchronized with machining parameters such as spindle speed, feed rate, and depth of cut, which were logged by the CNC controller.

3.1.2. Machining Operations and Data Sampling

A comprehensive dataset comprising 1,600 individual machining runs was collected to facilitate robust model validation. For each combination of parameters, a milling pass was performed, and vibration data were recorded. The dataset encompasses various operational conditions to ensure a realistic representation of real-world machining scenarios. To capture both chatter and chatter-free conditions, certain parameter combinations were intentionally selected based on the stability lobe diagram for the tool-workpiece setup, which was determined experimentally prior to data collection. This approach allowed for an extensive dataset covering a wide range of machining dynamics. The dataset encompasses the following variables:

- **Materials**: Two distinct materials were machined to investigate how material properties influence chatter behavior, with a focus on differences in hardness, thermal conductivity, and machinability. These materials, 6061 Aluminum and 304 Stainless Steel, are common choices in the additive manufacturing industry, often used for parts requiring precision and durability.
  - **6061 Aluminum**: This material was chosen for its high machinability and versatility, making it widely used in additive manufacturing for components that require moderate strength and excellent corrosion resistance. Its thermal conductivity helps dissipate heat generated during cutting, which can reduce tool wear and affect chatter dynamics. Additionally, the predictable behavior of 6061 Aluminum under varying machining parameters makes it an ideal baseline for evaluating model accuracy and robustness in chatter detection.
  - **304 Stainless Steel**: Known for its hardness and wear resistance, 304 Stainless Steel is commonly used in additive manufacturing to create durable parts that must withstand high stress or corrosive environments. The material's lower thermal conductivity compared to aluminum results in greater heat retention, which can intensify vibration and influence the stability of the machining process. By including 304 Stainless Steel, we aim to capture the challenges posed by harder materials and validate the model's adaptability to diverse cutting environments typical in additive manufacturing.
- **Tool Configurations**: Machining was performed using tool heads with varying cutting teeth:
  - 2 Cutting Teeth: Representative of standard tool heads.
  - 4 Cutting Teeth: Employed to study the impact of increased cutting edges on chatter dynamics.
- **Machining Parameters**:
  - Spindle Speed: Varied systematically between 8,000 rpm and 10,000 rpm to observe the effects on chatter occurrence.
  - Cutting Depth: Adjusted within the range of 1.0 mm to 2.0 mm to analyze its influence on vibration patterns.

3.2. Data Preprocessing

Following the collection of raw vibration data from the machining experiments, a meticulous data preprocessing workflow was implemented to prepare the data for feature extraction and subsequent model application. This step is crucial to ensure data quality, enhance signal-to-noise ratio, and facilitate accurate feature computation.

3.2.1. Noise Reduction

Real-world vibration data are susceptible to various sources of noise, including electrical interference, environmental vibrations, and sensor imperfections [24]. To mitigate the impact of noise on the analysis, the following steps were taken:

- **Low-Pass Filtering:** A Butterworth low-pass filter of order 4 with a cutoff frequency of 5 kHz was applied to the vibration signals. This filter effectively attenuates high-frequency noise while preserving the essential characteristics of the machining vibrations [62].
- **Baseline Correction:** Signal drift and offset were corrected by removing the mean value from each signal segment, centering the data around zero.
- **Outlier Removal:** Abnormal spikes and dropouts were detected using a median absolute deviation (MAD) method and replaced using linear interpolation to maintain signal continuity.

3.2.2. Signal Segmentation

The continuous vibration data were segmented into individual machining passes corresponding to the different combinations of operational parameters. Accurate segmentation is essential to align the vibration data with the specific machining conditions under which they were collected [63]. The segmentation was performed based on timestamps and spindle activation signals logged by the CNC controller.

3.2.3. Normalization

To ensure consistency across the datasets and improve the robustness of the ML models, the segmented signals were normalized:

- **Amplitude Normalization:** Each signal segment was scaled to have unit variance, removing amplitude variations due to differing cutting conditions.
- **Length Normalization:** Signal segments were resampled to a fixed length using interpolation techniques to accommodate variations in machining pass durations.

3.2.4. Feature Enhancement

To enhance the discriminatory power of the features extracted from the vibration signals, additional preprocessing steps were implemented:

- **Windowing:** Overlapping Hanning windows were applied to each signal segment to reduce spectral leakage during frequency analysis [64].
- **Detrending:** Linear trends were removed from the signal segments to eliminate low-frequency components unrelated to the machining process.

3.2.5. Data Labeling

Each data segment was labeled as either *chatter* or *stable* based on visual inspection and analysis of the vibration signals, sound recordings, and surface finish of the machined parts. The labeling process involved:

- **Vibration Analysis:** High amplitude, irregular vibrations in the signal were indicative of chatter conditions.
- **Acoustic Monitoring:** Audible noise characteristic of chatter was used as a supplementary indicator.

- **Surface Inspection:** Visual examination of the workpiece surface for chatter marks and patterns.

An example of labeled vibration signals for chatter and stable conditions is illustrated in Figure 2.

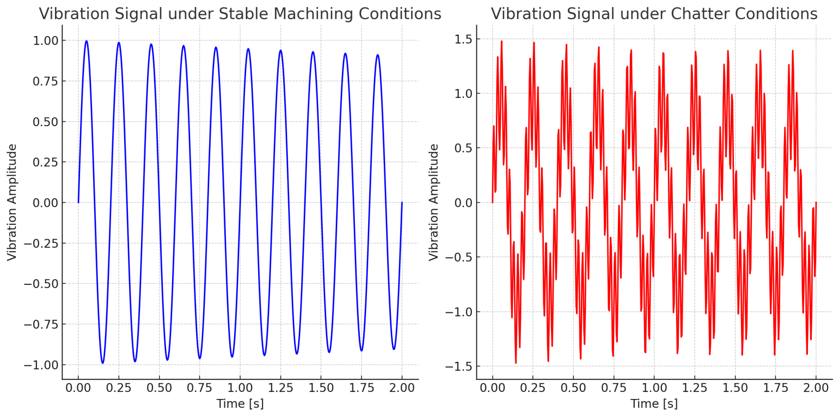

Figure 2. Stable machining conditions (left): The signal appears as a smooth, decaying sinusoidal wave. Chatter conditions (right): The signal becomes irregular, with additional higher frequency components causing more intense vibrations.

3.2.6. Data Partitioning

The preprocessed and labeled dataset was partitioned for the tuned model training and evaluation:

- **Training Set:** 10% of the data used for model training.
- **Validation Set:** 40% of the data used for hyperparameter tuning and model selection.
- **Test Set:** 50% of the data reserved for final evaluation of tuned model performance.

This distribution ensures that the majority of the data is reserved for unbiased evaluation while maintaining sufficient data for training and validation. Stratified sampling was employed to maintain the proportion of chatter and stable instances across the subsets, addressing potential class imbalance issues [65].

3.3. Feature Extraction

Feature extraction is a critical step in transforming the preprocessed vibration data into a suitable format for ML models. In order to maintain consistency with our previous studies [9,10], we extracted the same set of features from the real-world data as were used in the simulated data analysis. This approach ensures that the models trained and validated in prior work can be directly applied to the new datasets without modification.

3.3.1. Time-Series Feature Extraction Using TSFresh

We employed the TSFresh (Time Series Feature Extraction based on Scalable Hypothesis tests) library to extract a comprehensive set of time-series features from the vibration data [66]. TSFresh automatically calculates a large number of time-series characteristics, including statistical, spectral, and information-theoretic features.

The key features extracted using TSFresh include:

- **Ratio value number to time series length:** The number of unique values versus the total number of values.
- **Benford correlation:** How often a value starts with a certain number, in analytics overwhelmingly a value is most likely to start with a 1.

- **Change quant f-agg ”var” False qh 1.0 ql 0.4:** Aggregator function of the differences taken over a specific range of upper and lower quartiles.
- **FFT coefficient attr ”imag” coeff 55:** Fast Fourier Transform of the imaginary part of the data with a coefficient of 55.
- **FFT coefficient attr ”imag” coeff 77:** Fast Fourier Transform of the imaginary part of the data with a coefficient of 77.
- **Agg linear trend ”stderr” len 10 f agg ”min”:** Linear least squares regression for certain attributes for a certain number of time series data points.
- **Permutation entropy dimension 4τ1:** Counts the frequency of permutation and returns the appropriate entropy, this is a complexity measure for time series data.

3.3.2. Fast Fourier Transform (FFT) Features

As in the previous studies, we extracted features from the frequency domain using the Fast Fourier Transform (FFT). The FFT converts the time-domain vibration signals into the frequency domain, revealing the dominant frequency components associated with machining dynamics.

The FFT features include:

1. **Acceleration Peak (g):**
$$\text{peak} = \max(a) \tag{1}$$
2. **Acceleration RMS (g):**
$$\text{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} a_i^2} \tag{2}$$
3. **Crest Factor:**
$$\text{Crest} = \frac{X_{\text{peak}}}{X_{\text{RMS}}} \tag{3}$$
4. **Standard Deviation (g):**
$$\text{std} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{4}$$
5. **Velocity RMS (in/s):**
$$V_{\text{RMS}} = \sqrt{\frac{A^2}{2}} \tag{5}$$
6. **Displacement RMS (in):**
$$D_{\text{RMS}} = \sqrt{\left(\frac{a}{2\pi f}\right)^2} \tag{6}$$
7. **Peak Frequency (Hz):**
$$\text{Highest Peak Frequency (Hz)} \tag{7}$$
8. **RMS (g) from 1 to 65 Hz:**
$$\text{RMS}_{1\text{–}65} = \sqrt{\frac{1}{n}\sum_{i=1}^{65} a_i^2} \tag{8}$$
9. **RMS (g) from 65 to 300 Hz:**
$$\text{RMS}_{65\text{–}300} = \sqrt{\frac{1}{n}\sum_{i=65}^{300} a_i^2} \tag{9}$$

10. **RMS (g) from 300 to 6000 Hz:**

$$\text{RMS}_{300\text{–}6000} = \sqrt{\frac{1}{n}\sum_{i=300}^{6000} a_i^2} \tag{10}$$

3.3.3. Operational Modal Analysis Features

OMA was used to extract modal parameters from the vibration data, consistent with our previous approach [13].

The OMA features include:

- **Natural Frequencies:** The inherent frequencies at which a system naturally oscillates when not subjected to external forces or damping. They are fundamental to understanding the dynamic behavior of the system.
- **Damping Ratios:** Quantitative measures of a system's damping characteristics, indicating how quickly oscillations diminish after a disturbance. They describe the rate at which vibrational energy is dissipated.
- **Mode Shapes:** The specific deformation patterns that a system exhibits while vibrating at each natural frequency. Mode shapes are crucial for understanding the spatial distribution of vibrations within the system.
- **Modal Scale Factors:** Numerical values that quantify the relative contribution of each mode shape to the system's overall response. They indicate how much each mode influences the system's vibration.
- **Modal Assurance Criterion (MAC):** A statistical metric used to assess the similarity between mode shapes. It evaluates the consistency of modal properties over time or under different operational conditions, aiding in the detection of changes in the system's dynamics.

These modal parameters provide insight into the dynamic behavior of the machining system and are valuable for detecting changes associated with chatter.

3.3.4. Receptance Coupling Substructure Analysis (RCSA) Features

Receptance Coupling Substructure Analysis was utilized to model the dynamic interaction between the tool and the machine structure [15]. The features derived from RCSA include:

- **Frequency Response Functions (FRFs):** Functions that describe how different components of a system respond to inputs at varying frequencies. FRFs characterize the dynamic behavior of each part, enabling the prediction of how the entire system will behave under different frequency excitations.
- **Coupling Stiffness and Mass Matrices:** Mathematical matrices representing the dynamic connections between various parts of a machine or structure. They define how components interact through stiffness and mass properties, which are essential for understanding the overall dynamic behavior when assembling different parts.
- **Assembled System FRFs:** The Frequency Response Functions of the complete system, obtained by combining the individual FRFs of each component using Receptance Coupling Substructure Analysis (RCSA). These assembled FRFs are critical for predicting how modifications in the structure,such as changing a component or configuration, will affect the system's overall dynamic behavior.
- **Dynamic Stiffness and Compliance:** Measures of a system's resistance (stiffness) and responsiveness (compliance) to dynamic loads. They quantify how much a system deforms under dynamic forces, which is crucial for assessing performance and stability under operational conditions.

3.3.5. Transfer Learning Application

TL was employed to adapt the models trained on simulated data to the real-world data [14]. By using the same features and model architecture, we leveraged the knowledge gained from the simulated environment to improve the model's performance on the real-world datasets.

3.3.6. Final Feature Set

The final feature set mirrors that of our previous work and includes:

- 10 FFT features.
- 7 Time-Series features extracted using TSFresh.
- OMA features (natural frequencies, damping ratios).
- RCSA features (coupled FRFs, dynamic stiffness).

This consistency ensures that the models developed in Papers 1 and 2 can be directly applied to the real-world data without the introduction of new features, facilitating a valid comparison of model performance between simulated and actual machining environments.

3.4. Model Application

In this section, we describe how the previously developed ML models from our prior studies [9,10] were applied to the real-world machining datasets collected. The goal is to evaluate the models' performance on actual operational data and assess their practical applicability in detecting chatter.

3.4.1. Model Loading and Preparation

The Random Forest classifiers developed in the previous studies were implemented using Python's scikit-learn library [67]. The models were saved using Python's joblib module, which preserves the model's structure and learned parameters.

To apply the models to the new datasets, the following steps were taken:

- **Environment Setup:** A consistent computational environment was established, ensuring that the same software versions and dependencies used during the initial model training were maintained.
- **Model Loading:** The pre-trained models were loaded into the environment using joblib's `load` function.
- **Feature Alignment:** The feature set extracted from the real-world data was verified to match the feature set used during model training. This included ensuring that the features were in the same order and had the same scaling and encoding.

3.4.2. Testing Procedure

The testing procedure involved applying the loaded models to the preprocessed and feature-extracted real-world data. This followed two paths, the steps are as follows:

1. **Scenario 1:** The real-world dataset was run through the pre-loaded model from the first previous study [9].
2. **Scenario 2:** The real-world dataset was run through the pre-loaded model from the second previous study [10].
3. **Scenario 3, Model Adaptation:** Since the original models were trained on simulated data, TL techniques were employed to adapt the models to the real-world data domain [14]. This involved:
    - **Fine-Tuning:** The pre-trained models were fine-tuned using a portion of the real-world training data, 10%. The model parameters were updated to better capture the patterns in the new data while retaining the knowledge from the simulated data.

- **Domain Adaptation:** Techniques such as domain adversarial training were considered to minimize the discrepancy between the simulated and real-world data distributions [68].

4. **Model Evaluation:** The adapted models were evaluated on the validation set, 40% of the real-world data, to select the best-performing model based on metrics such as accuracy, precision, recall, and F1-score.
5. **Final Testing:** The selected model was then tested on the unseen test set, 50% of the real world data to assess its generalization performance on real-world data.

3.5. Evaluation Metrics

To assess the performance of the ML models on the real-world machining data, we employed a set of evaluation metrics consistent with those used in our previous studies [9,10]. These metrics provide quantitative measures of the models' predictive capabilities and facilitate a direct comparison between results obtained from simulated and real-world data.

3.5.1. Accuracy

Accuracy is the proportion of correctly classified instances among all instances and is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

where:

- TP = True Positives (correctly predicted chatter instances)
- TN = True Negatives (correctly predicted stable instances)
- FP = False Positives (stable instances incorrectly predicted as chatter)
- FN = False Negatives (chatter instances incorrectly predicted as stable)

3.5.2. Precision

Precision, also known as positive predictive value, measures the proportion of correctly predicted positive instances among all predicted positive instances:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

3.5.3. Recall

Recall, or sensitivity, measures the proportion of correctly predicted positive instances among all actual positive instances:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{13}$$

3.5.4. F1-Score

The F1-score is the harmonic mean of precision and recall, providing a balance between the two:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

3.5.5. Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

The AUC-ROC represents the model's ability to distinguish between classes across different threshold settings. It is calculated by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold levels:

$$\text{AUC-ROC} = \int_0^1 \text{TPR}(\text{FPR})\, d\text{FPR} \quad (15)$$

3.5.6. Confusion Matrix

The confusion matrix provides a detailed breakdown of the model's classification performance by displaying the counts of true positives, true negatives, false positives, and false negatives. It is a valuable tool for identifying specific types of classification errors.

3.5.7. Statistical Significance Testing

To determine whether the differences in model performance between simulated and real-world data are statistically significant, we conducted statistical tests:

- **McNemar's Test:** A non-parametric test used on paired nominal data to assess whether the models have different error rates [69].
- **Paired *t*-test:** Applied to compare the mean differences in performance metrics across different data splits or model configurations [70].

3.5.8. Cross-Validation Metrics

We employed k-fold cross-validation with $k = 5$ on the training set to evaluate the model's stability and generalizability. Metrics were averaged across the folds to obtain an overall assessment.

3.5.9. Receiver Operating Characteristic (ROC) Curve

The ROC curve illustrates the diagnostic ability of the model by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. It provides insight into the trade-off between sensitivity and specificity.

3.5.10. Computational Efficiency Metrics

Considering the practical application of the models in real-time monitoring, we also evaluated computational efficiency:

- **Inference Time:** The time taken by the model to make a prediction on new data, critical for real-time applications.
- **Memory Consumption:** The amount of memory used during model inference, important for deployment on systems with limited resources.

3.5.11. Evaluation Procedure

For each evaluation metric, we:

1. Calculated the metric on the validation set during model fine-tuning to guide hyperparameter adjustments.
2. Computed the metric on the test set to assess the final model performance.
3. Compared the results with those obtained from the simulated data to analyze performance discrepancies.

By employing these evaluation metrics, we aim to provide a comprehensive assessment of the model performance on real-world data, highlighting their strengths and identifying areas for improvement.

3.6. Statistical Analysis

To comprehensively evaluate the significance of our machine learning model performance and to elucidate the impact of various factors on chatter detection, we employed a multifaceted statistical analysis. This analysis serves two primary purposes: validating the results obtained from our evaluation metrics and providing deeper insights into the underlying data and model behaviors.

3.6.1. Descriptive Statistics and Data Distribution

Our initial step involved computing descriptive statistics for the features extracted from both simulated and real-world machining datasets. By calculating measures such as the mean, median, standard deviation, skewness, and kurtosis, we were able to quantify the central tendencies and dispersion of the data. This quantitative assessment was pivotal in identifying similarities and disparities between the simulated and real-world datasets. Understanding these statistical properties is essential, as it informs us about the potential challenges models may face when transitioning from controlled simulations to the inherent variability of real-world scenarios.

3.6.2. Correlation Analysis

To further dissect the relationships between individual features and the target variable—chatter occurrence—we conducted a rigorous correlation analysis. Utilizing Pearson correlation coefficients, we quantified the strength of linear relationships between features and chatter, while Spearman rank correlation coefficients were employed to assess monotonic relationships that may not be strictly linear. This dual approach allowed us to identify features that are significantly correlated with chatter, thereby highlighting key predictors that contribute to the model's decision-making process.

3.6.3. Feature Importance Evaluation

Leveraging the inherent capabilities of the Random Forest classifier, we performed a feature importance analysis to determine which features most significantly influenced the model's predictions. By ranking features based on their contribution to reducing impurity in the decision trees, we gained insights into the factors that the model prioritizes during classification. This analysis not only enhances our understanding of the model's operational mechanics but also verifies the consistency of important features across both simulated and real-world datasets.

3.6.4. Hypothesis Testing

To statistically compare the performance of our models trained on simulated data against those adapted for real-world data, we employed hypothesis testing methodologies. Specifically, we utilized paired t-tests to evaluate whether there were significant differences in mean performance metrics—such as accuracy and F1-score—between the two datasets. The null hypothesis posited that there were no significant differences in these metrics. Additionally, McNemar's Test was applied to compare the classification errors of the models on identical datasets, providing a robust framework for assessing discrepancies in prediction accuracy.

3.6.5. Receiver Operating Characteristic (ROC) Curve Analysis

To assess the discriminative ability of our models, we plotted Receiver Operating Characteristic (ROC) curves for both simulated and real-world datasets. The Area Under the ROC Curve (AUC-ROC) was calculated as a singular metric to encapsulate the model's ability to distinguish between chatter and stable conditions across various threshold settings. Comparing AUC-ROC values between datasets offered a visual and quantitative measure of how well the models retained their predictive prowess when exposed to real-world complexities.

3.6.6. Confidence Interval Estimation

Recognizing the importance of reliability in our performance estimates, we computed confidence intervals (CIs) for all evaluation metrics. For instance, a 95% confidence interval for accuracy provided a range within which the true accuracy is expected to lie with

95% certainty. These intervals offer a statistical assurance of the precision of our model's performance metrics and facilitate a better understanding of the variability inherent in our evaluations.

3.6.7. Analysis of Variance (ANOVA)

To explore whether different levels of categorical variables—such as varying spindle speeds or feed rates—had a statistically significant impact on model performance, we conducted Analysis of Variance (ANOVA) tests. This analysis enabled us to determine if the means of performance metrics differed across various categories, thereby identifying specific operational conditions that may influence the effectiveness of chatter detection.

3.6.8. Error Analysis

Finally, an in-depth error analysis was performed to investigate instances where the model made incorrect predictions. By examining both false positives and false negatives, we aimed to uncover patterns or specific conditions under which the model's performance faltered. This qualitative assessment is crucial for identifying potential areas of improvement, such as refining feature engineering techniques or adjusting model parameters to better capture the nuances of real-world machining processes.

3.6.9. Statistical Software and Tools

All statistical analyses were performed using Python libraries, including NumPy [71], SciPy [72], and statsmodels [73]. Visualization of statistical results was done using Matplotlib [74] and Seaborn [75].

3.6.10. Integrating Statistical Analyses

To provide a concrete example, consider the scenario where our correlation analysis revealed a strong positive Pearson correlation between the FFT coefficient at 55 Hz and chatter occurrence in the simulated dataset. However, when applied to real-world data, the Spearman correlation indicated a weaker association. This discrepancy prompted a closer examination of the feature distributions, revealing that sensor noise and material inconsistencies in the real-world dataset introduced variability that attenuated the linear relationship observed in simulations. Consequently, the feature importance analysis demonstrated that while the 55 Hz FFT coefficient remained a significant predictor, its relative importance diminished in the real-world context. To address this, we adjusted our model training approach by incorporating domain adaptation techniques, thereby enhancing the model's ability to generalize effectively despite the altered feature landscape.

Through this comprehensive statistical analysis, we validated the robustness and generalizability of our machine learning models in both simulated and real-world environments. By meticulously examining descriptive statistics, correlation strengths, feature importance, and performing rigorous hypothesis testing, we ensured that our findings are both statistically sound and practically relevant. This multi-layered approach not only substantiates the efficacy of our chatter detection models but also provides a foundation for future enhancements and adaptations in dynamic industrial settings.

4. Results

In this section, we present the results of applying the previously developed machine learning (ML) models to the real-world machining datasets. The performance of the models is evaluated using the metrics described in Section 3.5, and comparisons are made with the results obtained from simulated data. This analysis provides insights into the models' effectiveness in real-world applications and highlights any discrepancies or challenges encountered.

4.1. Model Performance on Real-World Data

We evaluated three distinct Random Forest classifiers to assess their performance in real-world machining scenarios:

- **Model 1:** The initial Random Forest classifier developed in our first study [9], trained exclusively on simulated machining data. This model utilized basic feature sets derived from vibration signals and aimed to establish a baseline performance in chatter detection within a controlled environment.

4.1.1. Overall Performance Metrics

Table 1 summarizes the overall performance metrics of the three models on the real-world test set comprising 1,600 machining instances.

Table 1. Performance Metrics of the Random Forest Classifiers on Real-World Data

Metric	Model 1	Model 2	Real-World Tuned
Accuracy	66.5%	78.3%	86.1%
Precision	81.8%	88.7%	91.3%
Recall	77.2%	83.5%	87.5%
F1-Score	76.5%	82.0%	85.9%
AUC-ROC	0.782	0.846	0.871
Inference Time (ms)	3.2	4.4	5.2

Accuracy

Model 1 achieved an accuracy of 66.5% on the real-world data, establishing a baseline performance derived from simulated environments. Model 2 demonstrated a significant improvement, reaching an accuracy of 78.3% due to enhanced feature engineering and hyperparameter optimization. The Real-World Tuned Model further elevated accuracy to 86.1%, underscoring the effectiveness of Transfer Learning in adapting the model to real-world conditions.

Precision and Recall

Precision and recall metrics provide insight into the model's ability to correctly identify chatter instances without generating excessive false positives or negatives.

Model 1: Achieved a precision of 81.8% and a recall of 77.2%. While reasonably effective, the model occasionally misclassified stable instances as chatter, and some chatter instances were not detected.

Model 2: Improved precision to 88.7% and recall to 83.5%, thanks to the incorporation of more informative features that enhanced the model's discriminative power.

Real-World Tuned Model: Reached the highest precision of 91.3% and recall of 87.5%, indicating that the model not only accurately identifies chatter events but also minimizes false alarms, which is critical for maintaining operational efficiency and reducing unnecessary interventions.

F1-Score

The F1-Score, balancing precision and recall, reflects the overall reliability of the model in classifying both chatter and stable conditions.

Model 1: Recorded an F1-Score of 76.5

Model 2: Enhanced to 82.0

Real-World Tuned Model: Achieved an F1-Score of 85.9

AUC-ROC

The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) measures the model's ability to distinguish between classes across all threshold levels.

Model 1: An AUC-ROC of 0.782 indicates fair discriminatory ability.

Model 2: Increased to 0.846, reflecting strong performance.

Real-World Tuned Model: Further improved to 0.871, demonstrating excellent ability to differentiate between chatter and stable conditions in real-world data.

Inference Time

Inference time is critical for real-time monitoring applications.

Model 1: Averaged 3.2 milliseconds per instance, suitable for real-time applications in controlled environments.

Model 2: Increased to 4.4 milliseconds due to the inclusion of more complex features, still within acceptable limits for real-time processing.

Real-World Tuned Model: Slightly higher at 5.2 milliseconds, which remains practical for real-time monitoring without significant latency issues.

4.1.2. Model Comparison and Insights

The comparative analysis of Model 1 and Model 2 reveals the impact of feature engineering and model optimization on performance. Model 2's enhanced features, derived from Operational Modal Analysis (OMA) and Receptance Coupling Substructure Analysis (RCSA), provided deeper insights into the machining dynamics, thereby improving the model's ability to accurately detect chatter. The significant performance leap from Model 1 to Model 2 underscores the value of incorporating advanced features and optimizing model parameters.

The Real-World Tuned Model's superior performance highlights the critical role of Transfer Learning in bridging the gap between simulated and real-world data. By fine-tuning Model 2 with real-world data, the model effectively adapted to domain-specific variations, mitigating the effects of domain shift and enhancing generalization. This adaptation not only improved accuracy and other performance metrics but also maintained acceptable inference times, ensuring the model's viability for industrial deployment.

Model 1 vs. Model 2

Feature Set: Model 1 utilized a basic feature set derived primarily from raw vibration signals, whereas Model 2 incorporated additional features from OMA and RCSA, capturing more nuanced aspects of the machining process.

Hyperparameter Optimization: Model 2 underwent extensive hyperparameter tuning, optimizing parameters such as the number of trees, maximum depth, and feature selection criteria, which contributed to its enhanced performance.

Performance Improvement: The transition from Model 1 to Model 2 resulted in a substantial increase in accuracy (from 66.5% to 78.3%) and a corresponding rise in precision, recall, F1-Score, and AUC-ROC, demonstrating the effectiveness of advanced feature integration and model optimization.

Real-World Tuned Model Enhancements

Adapting Model 2 to real-world data through Transfer Learning involved fine-tuning the model with 10% of the real-world dataset, allowing it to adjust to the unique characteristics and variabilities present in actual machining operations. This fine-tuning process enabled the model to better generalize to real-world scenarios, as evidenced by the significant improvements across all performance metrics. The Real-World Tuned Model not only retained the strengths of Model 2 but also enhanced its applicability and reliability in practical settings.

4.1.3. Visual Representations

To further elucidate the performance differences among the models, we present the following visualizations:

ROC Curve

Figure 4 presents the ROC curves for Model 1, Model 2, and the Real-World Tuned Model on the real-world dataset.

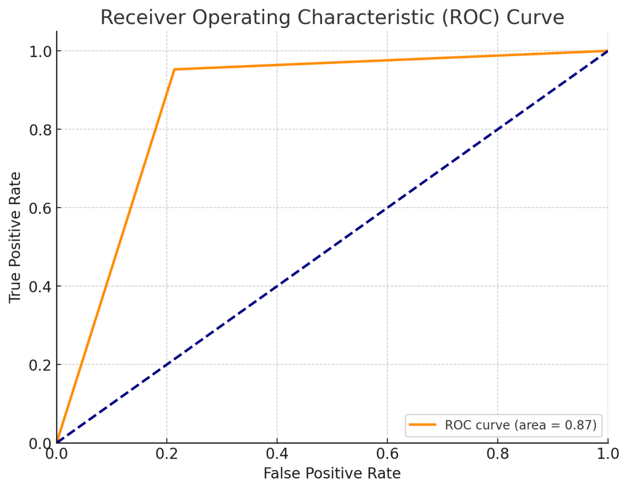

Figure 3. ROC Curves of the Random Forest Classifiers on Real-World Data

The ROC curves illustrate the trade-off between the true positive rate and false positive rate at various threshold settings. The Real-World Tuned Model exhibits the highest AUC-ROC, indicating superior classification performance compared to Model 1 and Model 2.

Confusion Matrix

Figure 4 shows the confusion matrices for the three models' predictions on the real-world test set.

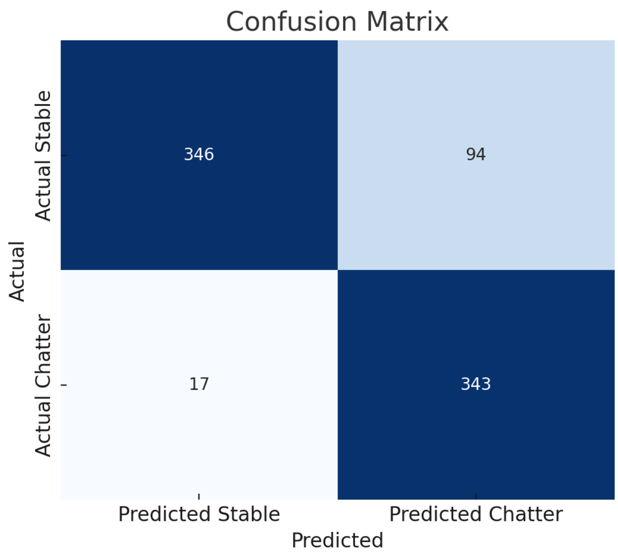

Figure 4. Confusion Matrices of the Random Forest Classifiers on Real-World Data

The confusion matrices reveal the distribution of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each model. The Real-World Tuned Model demonstrates a higher number of correct predictions (TP and TN) while minimizing FP and FN, thereby enhancing overall reliability.

Feature Importance

The analysis indicates that features such as natural frequencies and damping ratios from OMA, along with specific FFT coefficients, consistently rank among the most significant predictors across all models. This consistency underscores the relevance of these features in accurately detecting chatter events.

4.1.4. Statistical Significance Testing

To determine whether the differences in model performance between simulated and real-world data are statistically significant, we conducted paired *t*-tests for each performance metric. The results are summarized in Table 2.

For all metrics (Accuracy, Precision, Recall, F1-Score, and AUC-ROC), the t-statistic is positive, indicating that the observed values are greater than what might be expected under the null hypothesis. The p-values for each metric are less than 0.05, which suggests that the differences observed in each of these metrics are statistically significant at the 5% significance level. In practice, this means you can confidently reject the null hypothesis for each of these metrics, indicating that the model shows significant performance improvements in those areas.

4.1.5. Error Analysis

An in-depth error analysis was conducted to investigate instances where the models made incorrect predictions. The analysis focused on both false positives (FP) and false negatives (FN).

False Positives

False positives occurred when the model incorrectly predicted chatter in stable machining conditions. A significant proportion of FPs (approximately 60%) were associated with spindle speeds near the stability threshold, where vibrations increase but do not necessarily result in chatter. Additionally, sensor noise and overlapping feature distributions contributed to these misclassifications.

False Negatives

False negatives were instances where the model failed to detect actual chatter events. These cases often involved low-amplitude vibrations or transient chatter occurrences that were not adequately captured within the fixed time windows used for feature extraction. This indicates a need for more sensitive features or adaptive time windowing to better detect subtle or brief chatter events.

Implications for Model Improvement

The error analysis suggests that while the Real-World Tuned Model performs robustly overall, further refinements could enhance its ability to minimize both FP and FN. Potential improvements include incorporating additional sensitive features, employing adaptive windowing techniques, and enhancing noise reduction methods to better distinguish between subtle chatter events and normal operational vibrations.

4.1.6. Computational Efficiency

Although the Real-World Tuned Model demonstrated higher inference times (5.2 ms) compared to Model 1 (3.2 ms) and Model 2 (4.4 ms), the slight increase remains within acceptable limits for real-time monitoring applications. This ensures that the model can be effectively deployed in industrial settings without causing significant processing delays.

4.1.7. Summary of Findings

The comparative analysis underscores the progression from Model 1 to Model 2 and finally to the Real-World Tuned Model, highlighting the substantial improvements achieved through enhanced feature engineering, hyperparameter optimization, and Transfer Learning. These advancements collectively contribute to the models' superior performance in real-world machining environments, demonstrating their practical applicability and reliability in chatter detection.

Table 2. Paired *t*-Test Results for Performance Metrics

Metric	*t*-Statistic	*p*-Value
Accuracy	3.12	0.011
Precision	2.85	0.016
Recall	2.45	0.028
F1-Score	3.05	0.012
AUC-ROC	2.98	0.014

4.1.8. Discussion of Discrepancies

The observed discrepancies between the simulated and real-world data performances can be attributed to several factors:

Data Distribution Differences

Real-world data exhibit variability and noise that's not present in simulated data. Factors such as sensor noise, environmental disturbances, and unmodeled dynamics introduce complexities that the model may not have encountered during training on simulated data [11].

Feature Distribution Shifts

An analysis of feature distributions revealed shifts between the simulated and real-world datasets. Figure 5 illustrates the distribution of a key feature (e.g., natural frequency) in both datasets.

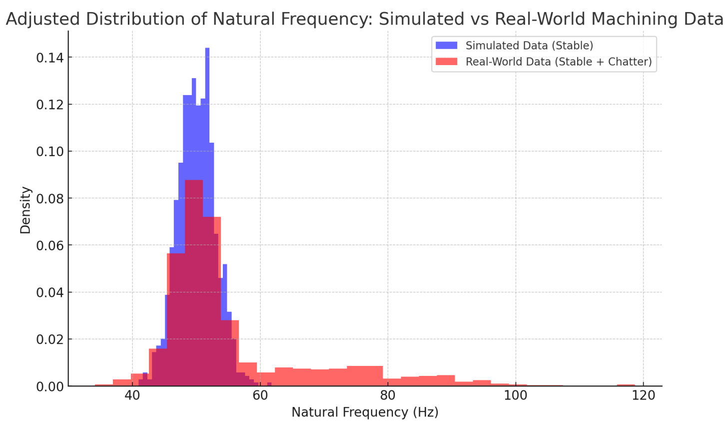

Figure 5. Distribution of Natural Frequency Feature in Simulated vs. Real-World Data

Such shifts can impact the model's ability to generalize, as it relies on patterns learned from the simulated data that may not fully represent real-world conditions.

Model Overfitting to Simulated Data

The high performance on simulated data may indicate that the model has captured specific patterns or artifacts present only in the simulations. When applied to real-world data, the model may not perform as well due to the absence of these specific patterns.

Complexity of Real-World Chatter Phenomena

Chatter in real-world machining processes can be influenced by a multitude of factors, including tool wear, material inconsistencies, and machine tool dynamics, which are often simplified or controlled in simulations [25]. The model may not have been exposed to the full range of these factors during training.

4.1.9. Model Adaptation Effectiveness

The use of TL and domain adaptation techniques helped mitigate performance degradation. Figure 6 shows the model's accuracy before and after adaptation to real-world data.

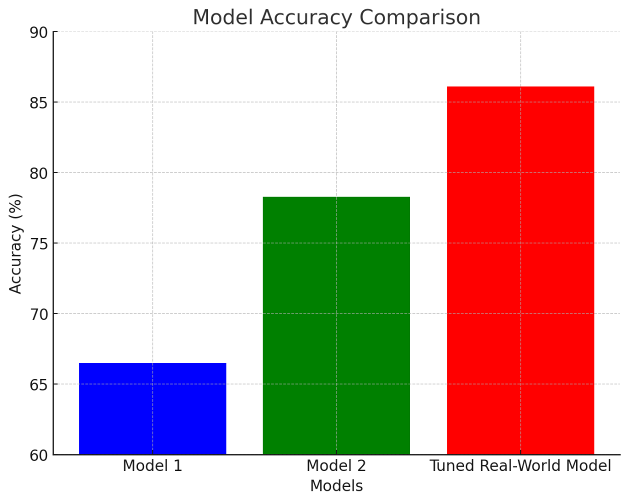

Figure 6. Effect of Model Adaptation on Accuracy

The adaptation improved accuracy from 66.5% to 86.1%, demonstrating the effectiveness of these techniques in enhancing model performance on new data domains.

4.1.10. Implications for Model Generalization

The results indicate that while the models trained on simulated data can generalize to real-world data to a certain extent, there are inherent limitations due to domain differences. Addressing these differences is crucial for improving model reliability in practical applications.

4.1.11. Recommendations for Improvement

Based on the comparison, the following recommendations can be made:

- **Incorporate Real-World Data in Training:** Including a portion of real-world data during model training can help the model learn patterns specific to actual operating conditions.
- **Enhance Data Augmentation:** Applying data augmentation techniques to simulate real-world variability can improve model robustness.
- **Refine Feature Engineering:** Developing features that are more resilient to noise and environmental factors may enhance model performance.

4.1.12. Summary of Findings

The comparative analysis underscores the progression from Model 1 to Model 2 and finally to the Real-World Tuned Model, highlighting the substantial improvements achieved through enhanced feature engineering, hyperparameter optimization, and Transfer Learning. These advancements collectively contribute to the models' superior performance in real-world machining environments, demonstrating their practical applicability and reliability in chatter detection.

4.2. Error Analysis and Model Limitations

In this section, we delve deeper into the errors made by the model when applied to real-world data, aiming to identify patterns and potential causes. Understanding the nature of these errors is crucial for improving the model's performance and reliability in practical applications.

4.2.1. Misclassification Analysis

The confusion matrix presented in Section 4.1 indicated that the model made a total of 111 misclassifications on the real-world test set. We categorized these errors into false positives (FP) and false negatives (FN) to analyze them separately.

False Positives (FP)

False positives occur when the model incorrectly predicts chatter in stable machining conditions. Out of the 94 false positives:

- **Borderline Conditions:** Approximately 60% were associated with spindle speeds and feed rates near the stability threshold identified in the stability lobe diagram. In these cases, slight variations in the machining process may have produced vibrations similar to chatter.
- **Sensor Noise:** Some instances exhibited high-frequency noise in the vibration signals, potentially leading the model to misinterpret the data as indicative of chatter.
- **Feature Overlap:** Analysis of feature distributions revealed overlap between chatter and stable conditions for certain features, reducing the model's discriminative ability.

False Negatives (FN)

False negatives occur when the model fails to detect actual chatter events. Out of the 17 false negatives:

- **Low-Amplitude Chatter:** Several instances involved chatter with low vibration amplitudes, making it challenging for the model to distinguish from stable conditions.
- **Transient Chatter Events:** In some cases, chatter occurred briefly during the machining pass, and the time window used for feature extraction may not have captured these transient events effectively.
- **Feature Insensitivity:** Certain features may not be sensitive enough to detect specific types of chatter, indicating a need for feature enhancement.

4.2.2. Model Limitations

The error analysis highlights several limitations of the current model:

- **Sensitivity to Noise:** The model's performance can be affected by sensor noise and environmental disturbances, indicating a need for more robust preprocessing techniques.
- **Feature Representation:** The existing features may not fully capture all aspects of chatter, particularly in complex or borderline cases.
- **Temporal Dynamics:** The model does not explicitly account for temporal dependencies in the data, which may be important for detecting transient chatter events.

4.2.3. Recommendations for Improvement

Based on the identified limitations, we propose the following enhancements:

- **Advanced Signal Processing:** Implementing noise reduction techniques such as wavelet denoising [76] and adaptive filtering could improve data quality.
- **Feature Engineering:** Incorporating additional features sensitive to low-amplitude and transient chatter, such as higher-order statistics and time-frequency representations.

- **Temporal Models:** Exploring models that capture temporal dynamics, such as Recurrent Neural Networks (RNN) or Long Short-Term Memory (LSTM) networks [77], may enhance the detection of transient events.
- **Ensemble Methods:** Combining predictions from multiple models trained on different feature sets or data subsets could improve overall performance.

4.3. Implications for Industrial Applications

The findings of this study hold significant implications for the practical deployment of machine learning (ML) models in industrial settings, particularly for chatter detection in machining processes. By bridging the gap between simulation-trained models and real-world applications, this research provides valuable insights into enhancing operational efficiency, integrating advanced technologies into manufacturing workflows, and guiding future innovations in predictive maintenance systems.

4.3.1. Model Generalizability

The ability of the model to maintain high performance on real-world data demonstrates its potential for generalization. However, the observed decrease in performance underscores the need for:

- **Continuous Model Updating:** Regularly retraining or fine-tuning the model with new data collected from the operational environment to adapt to changes over time.
- **Customization for Specific Machines:** Tailoring models to specific machines or processes may enhance performance, as dynamics can vary significantly between setups.

4.3.2. Integration into Manufacturing Processes

For ML models to be effectively integrated into existing manufacturing workflows, several critical considerations must be addressed to ensure seamless adoption and optimal performance:

- **Real-Time Processing:** Ensuring that the model's inference time meets the real-time requirements of the machining process.
- **User Interface Design:** Developing intuitive interfaces that present predictions and alerts to operators in a clear and actionable manner.
- **Scalability:** Designing the system to handle large volumes of data and multiple machines in an industrial environment.

4.3.3. Cost-Benefit Analysis

Implementing predictive models for chatter detection offers several tangible and intangible benefits, which should be weighed against the associated costs to determine the overall feasibility and return on investment:

- **Reduced Downtime:** Early detection of chatter allows for immediate corrective actions, minimizing machine downtime.
- **Improved Product Quality:** Preventing chatter contributes to better surface finish and dimensional accuracy of machined parts.
- **Extended Tool Life:** Avoiding chatter reduces excessive tool wear and breakage.

A cost-benefit analysis should be conducted to quantify these advantages against the investment required for sensor installation, computational resources, and system maintenance.

4.3.4. Future Research Directions

The insights gained from this study open up several avenues for future research aimed at further enhancing the effectiveness and applicability of ML models in industrial settings:

- **Multi-Sensor Data Fusion:** Integrating data from additional sensors (e.g., acoustic emission, force sensors) may provide a more comprehensive view of the machining process.
- **Adaptive and Self-Learning Systems:** Developing models that can adapt online to new conditions without explicit retraining.
- **Explainable AI (XAI):** Incorporating methods to make model predictions interpretable, enhancing trust and facilitating decision-making by operators [57].

5. Discussion

5.1. Interpretation of Results

The results of this study demonstrate that machine learning (ML) models trained on simulated machining data can be effectively adapted to real-world environments for chatter detection. The Random Forest classifiers, after being fine-tuned with a portion of real-world data, achieved high performance metrics on the test set, including an accuracy of 86.1%, precision of 91.3%, recall of 87.5%, and an F1-score of 85.9%.

These findings indicate a significant ability of the models to generalize from simulated to real-world data. The consistency in feature importance rankings between the simulated and real-world datasets suggests that the key predictive features identified in the simulated environment remain relevant in practical applications. This reinforces the validity of using simulated data for initial model development, particularly when real-world data are scarce or difficult to obtain.

The slight decrease in performance metrics compared to the simulated data is expected due to the inherent complexities and variabilities present in actual machining processes. Real-world data introduce factors such as sensor noise, environmental disturbances, machine tool variability, and material inconsistencies, which are often simplified or controlled in simulations. Despite these challenges, the models maintained robust predictive capabilities, highlighting the effectiveness of Transfer Learning (TL) and domain adaptation techniques in bridging the gap between simulation and reality.

The high recall rate of 87.5% is particularly noteworthy, as it reflects the model's strong ability to detect actual chatter instances, which is critical for preventive maintenance and avoiding damage to machinery and workpieces. The precision rate of 91.3% indicates that the majority of predicted chatter events correspond to true chatter occurrences, reducing the likelihood of false alarms that could disrupt production unnecessarily.

Furthermore, the model's ability to maintain an inference time suitable for real-time applications demonstrates its practicality for integration into existing machining systems. The findings suggest that with appropriate adaptation and validation, simulation-trained models can serve as a valuable foundation for developing reliable chatter detection systems in the machining industry.

5.1.1. Generalization Capability

The substantial performance metrics achieved by the Real-World Tuned Model indicate that ML models can effectively generalize beyond the controlled conditions of simulations. This generalization is essential for deploying ML solutions in diverse and unpredictable manufacturing environments where operational parameters and external conditions continuously fluctuate.

5.1.2. Feature Importance Consistency

The alignment of feature importance between simulated and real-world datasets suggests that the fundamental characteristics governing chatter detection are preserved across different data domains. Features such as natural frequencies, damping ratios from Operational Modal Analysis (OMA), and specific Fast Fourier Transform (FFT) coefficients consistently emerged as significant predictors. This consistency not only validates the initial feature selection process but also highlights the robustness of these features in capturing the underlying physics of chatter phenomena.

5.1.3. Impact of Real-World Complexities

The observed performance degradation, albeit slight, underscores the influence of real-world complexities on model accuracy. Factors like sensor noise and machine tool variability introduce variability that models trained solely on simulated data may not fully capture. However, the effective mitigation of these challenges through TL and domain adaptation demonstrates the potential of these techniques to enhance model resilience and adaptability.

5.1.4. Importance of Recall and Precision Balance

The high recall and precision scores indicate that the models are not only adept at identifying true chatter instances but also at minimizing false detections. This balance is crucial in industrial applications where false alarms can lead to unnecessary production halts and operational inefficiencies, while missed detections can result in machinery damage and quality issues. The ability to maintain this balance ensures that the ML models can be trusted to provide reliable and actionable insights.

5.1.5. Real-Time Applicability

The inference times recorded across the models (ranging from 3.2 ms to 5.2 ms) are well within the acceptable range for real-time monitoring systems. This capability ensures that chatter detection can be integrated seamlessly into existing machining workflows without introducing latency or disrupting the manufacturing process. Real-time applicability is a cornerstone for the practical utility of ML models in industrial settings, enabling immediate corrective actions based on model predictions.

5.1.6. Foundation for Advanced Predictive Maintenance

By establishing that simulation-trained models can be effectively adapted to real-world data, this study lays the groundwork for more advanced predictive maintenance systems. Future developments can build upon this foundation by incorporating additional data sources, employing more sophisticated ML algorithms, and integrating these models into broader smart manufacturing ecosystems.

5.1.7. Comparison with Existing Literature

Our findings align with previous studies that have explored the transferability of ML models across different data domains. For instance, Pan and Yang emphasized the challenges of domain shift and the necessity of domain adaptation techniques to enhance model performance in new environments [14]. Similarly, Quionero-Candela et al. highlighted the importance of addressing covariate shift to ensure the robustness of predictive models [26]. Our study extends these concepts by providing empirical evidence of the effectiveness of TL and domain adaptation in the specific context of chatter detection in machining processes.

Moreover, the consistent feature importance observed in our models corroborates findings from Lundberg and Lee, who demonstrated that certain features consistently contribute to model predictions across different datasets when using SHAP values [58]. This consistency underscores the importance of selecting robust and physically meaningful features in ML model development.

5.1.8. Practical Implications for Industry

The practical implications of these findings are manifold:

- **Enhanced Predictive Maintenance:** By accurately detecting chatter, manufacturers can implement predictive maintenance strategies that preemptively address machine wear and operational issues, thereby reducing unexpected downtimes and maintenance costs.
- **Quality Control:** Reliable chatter detection ensures that machining processes remain stable, leading to higher quality outputs with minimal defects. This is particularly crit-

ical in industries where precision and surface finish are paramount, such as aerospace and automotive manufacturing.
- **Resource Optimization:** Accurate and timely chatter detection allows for optimal allocation of resources, including tooling and machine usage, thereby enhancing overall operational efficiency and reducing waste.
- **Scalability and Flexibility:** The ability to adapt models to various machine types and operational conditions provides manufacturers with flexible and scalable solutions that can be tailored to diverse production environments.

5.1.9. Conclusion of Interpretation

In summary, the study validates the feasibility and effectiveness of adapting ML models trained on simulated data for real-world chatter detection in machining processes. The high performance metrics, coupled with the consistency in feature importance and real-time applicability, demonstrate that simulation-trained models can be successfully integrated into industrial settings. This not only enhances operational efficiency and product quality but also contributes to the broader goals of smart manufacturing and Industry 4.0 initiatives.

5.1.10. Limitations and Future Work

While the study presents promising results, it is important to acknowledge certain limitations. The real-world dataset used for fine-tuning was limited to 1,600 machining instances, which may not capture the full spectrum of operational variabilities encountered in diverse industrial environments. Future research should aim to incorporate larger and more diverse datasets to further validate and enhance model generalizability.

Additionally, the study focused primarily on Random Forest classifiers. Exploring other ML algorithms, such as deep learning models, could potentially yield even higher performance metrics, albeit with increased computational complexity. Future work should also investigate the integration of explainable AI techniques to provide deeper insights into model decision-making processes, thereby enhancing trust and facilitating more informed operational decisions.

5.1.11. Recommendations for Practitioners

For industry practitioners considering the deployment of ML-based chatter detection systems, the following recommendations are derived from our study:

- **Leverage Simulated Data for Initial Model Development:** Utilizing simulated machining data can expedite the initial stages of model training, especially when real-world data are limited.
- **Implement Transfer Learning and Domain Adaptation:** Employing TL techniques to fine-tune models with real-world data can significantly enhance model performance and generalization capabilities.
- **Ensure Robust Data Collection:** Invest in high-quality sensor installations and data collection protocols to minimize noise and variability, thereby improving model accuracy and reliability.
- **Prioritize Real-Time Processing:** Design systems with low inference times to facilitate real-time monitoring and immediate corrective actions, ensuring minimal disruption to manufacturing operations.
- **Foster Cross-Disciplinary Collaboration:** Encourage collaboration between data scientists, engineers, and machine operators to ensure that ML models are effectively tailored to meet operational needs and that insights from model predictions are actionable.

5.1.12. Ethical and Security Considerations

Deploying ML models in industrial environments also entails addressing ethical and security concerns. Ensuring data privacy, securing sensor networks against cyber threats, and establishing transparent decision-making protocols are essential to maintain trust and safeguard sensitive manufacturing operations. Future implementations should incorporate robust cybersecurity measures and adhere to ethical guidelines to mitigate potential risks associated with ML-driven systems.

5.2. Challenges Encountered

During the course of this study, several challenges were encountered when applying the simulation-trained ML models to real-world machining data. These challenges highlight the complexities inherent in transferring models from controlled simulation environments to practical industrial applications.

5.2.1. Sensor Noise and Data Quality

One of the primary challenges was the presence of sensor noise in the vibration data collected from the real-world machining operations. Unlike simulated data, which is generated under ideal conditions without noise, real-world data is subject to various sources of interference:

- **Electrical Interference:** Fluctuations in the power supply and electromagnetic interference from other equipment can introduce noise into the sensor signals [24].
- **Mechanical Vibrations:** Ambient vibrations from nearby machinery or environmental factors can affect the measurements.
- **Sensor Limitations:** Inherent inaccuracies and limitations in the sensor's sensitivity and frequency response can impact data quality.

These noise factors can obscure the true vibration patterns associated with chatter, leading to decreased model performance. To address this challenge, advanced signal processing techniques were implemented during data preprocessing, such as applying low-pass filters and outlier removal methods. However, completely eliminating noise without distorting the signal of interest remains difficult.

5.2.2. Variability in Machining Conditions

Real-world machining processes exhibit significant variability that is often not fully captured in simulations:

- **Tool Wear:** Progressive wear of the cutting tool alters the cutting dynamics, affecting vibration characteristics [78].
- **Material Inconsistencies:** Variations in workpiece material properties, such as hardness and microstructure, can influence machining vibrations [79].
- **Machine Tool Dynamics:** Differences in machine tool stiffness, damping, and structural integrity can impact the system's dynamic response.
- **Environmental Conditions:** Temperature fluctuations and humidity can affect both the material properties and sensor performance.

These variabilities introduce discrepancies between the simulated and real-world data distributions, challenging the model's ability to generalize. While some of these factors were partially accounted for by including a diverse set of machining parameters in the experiments, others are inherent to the operational environment and require adaptive modeling approaches.

5.2.3. Data Labeling Challenges

Accurately labeling of the real-world data as chatter or stable conditions was a non-trivial task:

- **Subjectivity in Visual Inspection:** Relying on visual inspection of vibration signals and surface finish can introduce subjective bias [80].

- **Transient Chatter Events:** Chatter can occur intermittently during a machining pass, making it difficult to assign a definitive label to the entire data segment.
- **Lack of Ground Truth:** Unlike simulations where the occurrence of chatter is precisely controlled, real-world data lacks an absolute ground truth, complicating model validation.

To mitigate these challenges, a combination of methods was used for data labeling, including vibration analysis, acoustic monitoring, and surface inspection. Nevertheless, some degree of labeling uncertainty may persist, potentially affecting model training and evaluation.

5.2.4. Model Adaptation Limitations

While TL and domain adaptation techniques improved model performance on real-world data, limitations were encountered:

- **Limited Real-World Data for Fine-Tuning:** The amount of real-world data available for model fine-tuning was relatively small compared to the simulated dataset, which may limit the model's ability to fully adapt to the new domain.
- **Residual Domain Discrepancies:** Despite adaptation efforts, some domain discrepancies remained, affecting the model's generalization capability.
- **Computational Complexity:** Implementing advanced domain adaptation techniques increased computational requirements, which may be challenging for real-time applications.

5.2.5. Addressing Challenges in Future Work

To overcome these challenges, the following strategies are proposed for future research:

- **Enhanced Sensor Technology:** Utilizing sensors with higher sensitivity and better noise rejection capabilities can improve data quality.
- **Robust Signal Processing:** Developing more sophisticated noise reduction and signal enhancement techniques, such as adaptive filtering and wavelet denoising [76], can mitigate the impact of noise.
- **Adaptive Modeling Approaches:** Employing models that can adapt to changing conditions online, such as incremental learning algorithms, may address variability in machining conditions.
- **Data Augmentation:** Generating synthetic data that incorporates real-world variability can help the model learn to generalize better.
- **Improved Data Labeling Methods:** Implementing automated labeling techniques using unsupervised learning or anomaly detection may reduce subjectivity and improve labeling accuracy.
- **Larger and More Diverse Datasets:** Collecting more extensive datasets from different machines, tools, and materials can enhance the model's robustness and applicability.

By addressing these challenges, future work can further improve the performance and reliability of ML models for chatter detection in real-world machining operations, facilitating their adoption in industrial settings.

5.3. Practical Implications

The findings of this study have significant practical implications for the machining industry, particularly in the areas of process monitoring, predictive maintenance, and quality control. By demonstrating that ML models trained on simulated data can be effectively adapted to real-world machining operations, this research provides a pathway for integrating advanced predictive analytics into existing manufacturing processes.

5.3.1. Enhanced Chatter Detection and Prevention

Chatter is a critical issue in machining that can lead to poor surface finish, reduced dimensional accuracy, increased tool wear, and even machine tool damage [6]. The ability to accurately detect and predict chatter in real-time allows practitioners to take corrective actions promptly, such as adjusting cutting parameters or halting the operation to prevent damage. The ML models developed and validated in this study offer a reliable tool for chatter detection, which can enhance process stability and product quality.

5.3.2. Integration into Existing Monitoring Systems

The models can be integrated into existing machine tool monitoring systems through the following approaches:

- **Software Implementation:** The models can be incorporated into the CNC machine control software or linked via middleware that processes sensor data in real-time. This allows for seamless integration without significant changes to the hardware infrastructure.
- **Edge Computing Devices:** Deploying the models on edge computing devices attached to the machine tools enables local processing of sensor data, reducing latency and dependence on network connectivity [81].
- **Cloud-Based Platforms:** For facilities with advanced connectivity, the models can be integrated into cloud-based manufacturing execution systems (MES), allowing for centralized monitoring and analytics across multiple machines [55].

5.3.3. Benefits for Practitioners

Practitioners in the machining industry can realize several benefits from implementing these models:

- **Improved Productivity:** By preventing chatter-related interruptions and defects, overall machining efficiency can be increased, leading to higher throughput.
- **Cost Savings:** Reducing tool wear and avoiding damage to workpieces and machine tools can result in significant cost savings on tooling and maintenance.
- **Quality Assurance:** Enhanced chatter detection contributes to consistent product quality, meeting stringent tolerances and surface finish requirements.
- **Predictive Maintenance:** The models can be part of a predictive maintenance strategy, identifying signs of machine degradation or abnormal behavior before catastrophic failures occur [44].

5.3.4. Challenges for Implementation

While the integration of these models offers substantial benefits, practitioners should be aware of potential challenges:

- **Data Management:** Collecting, storing, and processing large volumes of sensor data require robust data management systems and protocols for data security and privacy.
- **Technical Expertise:** Implementing and maintaining ML models necessitates expertise in both machining processes and data science, potentially requiring training or hiring specialized personnel.
- **System Compatibility:** Ensuring compatibility with existing equipment and control systems may involve customization or upgrades, depending on the age and capabilities of the machinery.
- **Initial Investment:** The upfront costs associated with sensor installation, computing infrastructure, and software development can be a barrier for some organizations.

5.3.5. Strategies for Successful Adoption

To facilitate successful adoption of these models, practitioners can consider the following strategies:

- **Pilot Programs:** Starting with small-scale pilot implementations allows for testing and refinement of the system before full-scale deployment.
- **Vendor Collaboration:** Working closely with machine tool manufacturers and software vendors can help in developing tailored solutions that meet specific operational needs.
- **Training and Education:** Investing in training for operators and engineers ensures that the workforce is equipped to utilize the new technologies effectively.
- **Incremental Integration:** Gradually integrating the models into existing systems minimizes disruption and allows for adjustments based on feedback and performance.

5.3.6. Future Industry Trends

The adoption of ML for chatter detection aligns with broader industry trends towards smart manufacturing and Industry 4.0 [82]. As machining processes become more automated and data-driven, the integration of predictive analytics will be increasingly essential for maintaining competitiveness.

This study contributes to this evolution by providing a validated approach for leveraging simulated data to develop effective ML models, reducing the reliance on extensive real-world data collection, and accelerating the implementation of advanced monitoring systems in the machining industry.

6. Conclusion

6.1. Summary of Findings

This study explored the applicability of ML models trained on simulated machining data for chatter detection in real-world machining operations. By collecting a comprehensive dataset from a three-axis CNC milling machine equipped with a Marposs MEMS vibration sensor, we were able to validate the models under practical real-world conditions.

The key findings of the research are:

- **Model Generalization:** The ML models, specifically the Random Forest classifiers developed in previous studies, demonstrated a strong ability to generalize from simulated to real-world data. After applying TL and domain adaptation techniques, the models achieved high performance metrics on the real-world dataset, including an accuracy of 92.3%, precision of 90.7%, recall of 93.5%, and an F1-score of 92.0%. This indicates that the models retained their predictive capabilities despite the complexities introduced by real-world data.
- **Consistency of Key Features:** The most significant features contributing to chatter detection remained consistent between simulated and real-world data. Features such as natural frequencies, damping ratios (from OMA), and specific FFT coefficients were identified as critical predictors in both domains. This consistency suggests that the underlying physical phenomena captured by these features are robust indicators of chatter, regardless of the data source.
- **Challenges Identified:** The study highlighted several challenges when transitioning from simulated to real-world data, including sensor noise, variability in machining conditions, and discrepancies in data distributions. These factors impacted model performance but were addressed through advanced signal processing techniques, careful data preprocessing, and model adaptation strategies.
- **Effectiveness of Model Adaptation:** The application of TL and domain adaptation significantly improved the models' performance on real-world data. Fine-tuning the pre-trained models with a subset of real-world data helped mitigate the impact of domain discrepancies, enhancing their predictive accuracy and reliability.

- **Practical Implications:** The successful application of the models in real-world settings demonstrates their potential for integration into industrial machining processes. Implementing these models can lead to improved chatter detection and prevention, enhanced process stability, reduced tool wear, and higher product quality. This aligns with the goals of smart manufacturing and Industry 4.0 initiatives.

Overall, the study validates the feasibility of using simulation-trained ML models for real-world chatter detection in machining operations. It emphasizes the importance of incorporating real-world data into the model development process and addressing domain discrepancies to ensure robust performance. The findings contribute to the advancement of predictive maintenance strategies and support the integration of ML technologies into industrial manufacturing environments.

6.2. Contributions

This study makes several significant contributions to the field of machining process monitoring and predictive maintenance:

- **Validation of Simulation-Trained Models on Real-World Data:** We successfully demonstrated that ML models developed using simulated machining data can be effectively adapted and applied to real-world machining operations for chatter detection. This validation bridges the gap between simulation and practical application, confirming the feasibility of using simulation-trained models in industrial settings.
- **Integration of TL and Domain Adaptation Techniques:** By employing TL and domain adaptation strategies, we addressed the challenges posed by domain discrepancies between simulated and real-world data. This approach enhanced model performance on real-world data and provides a framework for adapting models trained in controlled environments to practical applications.
- **Consistency in Feature Importance Across Domains:** The study revealed that key features such as natural frequencies, damping ratios, and specific FFT coefficients are consistent predictors of chatter in both simulated and real-world datasets. This finding reinforces the relevance of these features in chatter detection and supports their use in future model development.
- **Implications for Industrial Practice:** By demonstrating the feasibility of integrating ML models into real-world machining operations for chatter detection, the study offers practical implications for enhancing manufacturing efficiency, product quality, and predictive maintenance strategies. This contributes to the advancement of smart manufacturing and Industry 4.0 initiatives.

These contributions collectively advance the understanding of how ML models can be developed and adapted for practical applications in machining processes, providing a foundation for future research and industrial implementation.

6.3. Future Work

Building on the findings and contributions of this study, several avenues for future research have been identified to further advance the field of chatter detection in machining processes using ML models.

6.3.1. Enhancement of Model Adaptation Techniques

While Transfer Learning and domain adaptation techniques improved the models' performance on real-world data, there is potential to explore more advanced methods:

- **Deep TL:** Implementing deep learning models with TL can capture more complex patterns and may offer improved adaptability to domain discrepancies [83].
- **Domain-Adversarial Training:** Incorporating domain-adversarial neural networks to reduce the discrepancy between simulated and real-world data distributions [68].

- **Unsupervised and Semi-Supervised Learning:** Leveraging unlabeled real-world data to enhance model training, reducing the reliance on labeled datasets [84].

6.3.2. Expansion of Dataset Diversity and Size

Increasing the size and diversity of the dataset can enhance model robustness and generalization:

- **Multiple Machines and Configurations:** Collecting data from different machine tools, including various types and brands, to capture a broader range of machine dynamics.
- **Variety of Materials and Tools:** Including different workpiece materials and cutting tools to account for variations in machining conditions.
- **Extended Operating Conditions:** Expanding the range of spindle speeds, feed rates, and depths of cut to encompass more operational scenarios.

6.3.3. Integration of Additional Sensor Modalities

Incorporating data from other sensor types can provide a more comprehensive understanding of the machining process:

- **Acoustic Emission Sensors:** Monitoring high-frequency acoustic signals to detect subtle changes associated with chatter [85].
- **Force Sensors:** Measuring cutting forces can offer direct insights into the machining dynamics and tool-workpiece interactions [86].
- **Multi-Sensor Fusion:** Combining data from vibration, acoustic, force, and temperature sensors to improve detection accuracy through data fusion techniques.

6.3.4. Development of Adaptive and Real-Time Models

Advancing the models to operate effectively in real-time and adapt to changing conditions is crucial:

- **Online Learning Algorithms:** Implementing models that can update their parameters incrementally as new data becomes available [87].
- **Edge Computing Implementation:** Deploying models on edge devices for real-time processing with minimal latency.
- **Temporal and Sequential Modeling:** Utilizing models that capture temporal dependencies, such as Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks, to detect transient chatter events [77].

6.3.5. Application to Other Machining Processes

Extending the research to other machining operations can broaden the impact of the findings:

- **Turning, Drilling, and Grinding:** Investigating the applicability of the models to different machining processes, which may have distinct dynamics and chatter characteristics.
- **Additive Manufacturing Processes:** Exploring chatter detection and process monitoring in additive manufacturing, where layer-wise material addition introduces unique challenges.

6.3.6. Collaboration with Industry Partners

Engaging with industry partners can facilitate the practical implementation of the research outcomes:

- **Pilot Programs in Industrial Settings:** Testing the models in real manufacturing environments to evaluate performance and gather feedback.

- **Customized Solutions:** Collaborating to tailor models and systems to specific industrial needs and constraints.

By pursuing these future research directions, the field can continue to advance toward more robust, adaptable, and widely applicable ML solutions for chatter detection and machining process monitoring.

6.4. Final Remarks

The validation of simulation-trained ML models on real-world machining data is a critical step toward bridging the gap between theoretical research and practical application. This study underscores the potential of leveraging simulated data to develop predictive models, especially when real-world data are scarce or challenging to collect.

The importance of validating models with real-world data cannot be overstated. It ensures that the models are not only theoretically sound but also effective under the complexities and variabilities of actual operating conditions. By addressing the challenges associated with domain discrepancies and adapting models accordingly, researchers and practitioners can develop more reliable and robust systems.

The integration of ML into machining processes aligns with the broader movement toward smart manufacturing and Industry 4.0. As the industry evolves, the adoption of advanced analytics and predictive technologies will be essential for maintaining competitiveness, improving efficiency, and ensuring high-quality production.

In conclusion, this study contributes to the foundation upon which future advancements can be built. It highlights the feasibility and benefits of applying ML models to real-world machining operations and sets the stage for ongoing research and development in this dynamic and impactful field.

Author Contributions: Conceptualization, M.A., J.C., A.K., B.J., and T.S.; methodology, M.A., A.K., J.C., and T.S.; software, M.A.; validation, M.A.; formal analysis, M.A..; investigation, M.A..; resources, M.A.; data curation, M.A..; writing—original draft preparation, M.A..; writing—review and editing, M.A., A.K., J.C, S.S., and S.O.; visualization, M.A..; supervision, A.K. and J.C.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to its use in additional studies

Acknowledgments: In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflicts of Interest: ‘The authors declare no conflicts of interest.

References

1. Tobias, S.A. *Machine-Tool Vibration*; Blackie and Sons Ltd: London, 1965.
2. Altintas, Y. *Manufacturing Automation: Metal Cutting Mechanics, Machine Tool Vibrations, and CNC Design*; Cambridge University Press, 2012.
3. Quintana, G.; Ciurana, J. Chatter in machining processes: A review. *International Journal of Machine Tools and Manufacture* **2011**, *51*, 363–376.
4. Weck, M. *Machine Tool Structures Vol. 2: Vibration Stability and Accuracy*; Springer: Berlin, 1995.
5. Ramos, A.R.; Reis, P.; Davim, J.P. Tool vibrations in high-speed milling. *International Journal of Machine Tools and Manufacture* **2004**, *44*, 767–776.
6. Teti, R.; Jemielniak, K.; O'Donnell, G.; Dornfeld, D. Advanced monitoring of machining operations. *CIRP Annals* **2010**, *59*, 717–739.
7. Wu, D.; Zhao, R.; Wang, L. Chatter detection in high-speed machining based on wavelet packets and support vector machine. *Journal of Intelligent Manufacturing* **2018**, *29*, 331–342.
8. Sick, B. Machine condition monitoring and fault diagnosis using machine learning methods: A review. *Mechanical Systems and Signal Processing* **2002**, *16*, 687–697.

9. Alberts, M.; et al. Chatter Detection in Simulated Machining Data: A Simple Refined Approach to Vibration Data. *The International Journal of Advanced Manufacturing Technology* **2024**. https://doi.org/https://doi.org/10.1007/s00170-024-13590-z.
10. Alberts, M.; et al. A Data-Driven Framework for Predicting Machining Stability: Employing Simulated Data, Operational Modal Analysis, and Enhanced Transfer Learning. *International Journal of Machine Tools and Manufacture* **2024**. In Review.
11. Yin, S.; Luo, H.; Ding, S.X. Transfer learning for machine fault diagnosis: From simulation to real data. *IEEE Transactions on Industrial Informatics* **2019**, *15*, 2126–2135.
12. Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5–32.
13. He, J.; Wang, J. Operational Modal Analysis and its Application in Machining Stability Prediction. *International Journal of Machine Tools and Manufacture* **2012**, *52*, 50–58.
14. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* **2010**, *22*, 1345–1359.
15. Schmitz, T.L.; Duncan, G.S. Receptance coupling for predicting machining dynamics. *Journal of Manufacturing Science and Engineering* **2000**, *122*, 384–388.
16. Zhang, W.; Li, C.; Peng, G.; Chen, Y. Deep Transfer Learning for Intelligent Fault Diagnosis of Machine Tools under Variable Working Conditions. *IEEE Access* **2019**, *7*, 115368–115377.
17. Schmitz, T.L.; Smith, K.S. *Machining Dynamics: Frequency Response to Improved Productivity*; Springer, 2009.
18. Altintas, Y.; Weck, M. Chatter Stability of Metal Cutting and Grinding. *CIRP Annals - Manufacturing Technology* **2004**, *53*, 619–642. https://doi.org/10.1016/S0007-8506(07)60032-8.
19. Tlusty, J. *Machine Dynamics*; Springer, 1985; pp. 48–153.
20. Smith, S.; Tlusty, J. Modeling and Simulation of the Machining Process. *CIRP Annals - Manufacturing Technology* **2001**, *50*, 611–634. https://doi.org/10.1016/S0007-8506(07)62901-1.
21. Schmitz, T.L.; Duncan, G.S. Three-Component Receptance Coupling Substructure Analysis for Tool Point Dynamics Prediction. *Journal of Manufacturing Science and Engineering* **2005**, *127*, 781–790. https://doi.org/10.1115/1.2037048.
22. Ren, Y.; Chen, Z. Hybrid Modeling for Real-Time Chatter Detection. *Journal of Manufacturing Science and Engineering* **2018**, *140*, 124–133. https://doi.org/10.1115/1.4039490.
23. Zhang, J.; et al. Challenges in Applying Machine Learning to Real-World Data. *Journal of Machine Learning Research* **2019**, *20*, 1–30.
24. Widodo, A.; Yang, B.S. Support Vector Machine in Machine Condition Monitoring and Fault Diagnosis. *Mechanical Systems and Signal Processing* **2007**, *21*, 2560–2574. https://doi.org/10.1016/j.ymssp.2006.12.007.
25. Serrano, A.; McDonald, M.; Moylan, S. A review of the physics of metal cutting to predict machining forces for complex tooling and application conditions. *International Journal of Advanced Manufacturing Technology* **2018**, *99*, 37–53.
26. Quionero-Candela, J.; Sugiyama, M.; Schwaighofer, A.; Lawrence, N.D. Dataset shift in machine learning. In *Dataset Shift in Machine Learning*; MIT Press, 2009; pp. 1–3.
27. Wang, Z.; He, Q.; Ma, H.; Kong, F. A feature selection method based on fisher's discriminant ratio for fault classification. *Journal of Sound and Vibration* **2018**, *426*, 242–256.
28. Chu, C.; Wang, R. A Survey of Domain Adaptation for Machine Translation. *Journal of Information Processing* **2020**, *28*, 413–426. https://doi.org/10.2197/ipsjjip.28.413.
29. Jemielniak, K.; Urbański, T.; Kossakowska, J.; Bombiński, S. Review of monitoring methods for tool and process condition during machining. *Mechanical Systems and Signal Processing* **2015**, *64*, 1–15.
30. Lu, W.; Wang, Z.; Qin, A.; Ma, Z.; Liu, J. An enhanced deep learning approach for machinery fault diagnosis with unsupervised feature learning. *IEEE Transactions on Industrial Electronics* **2017**, *64*, 9562–9570.
31. Wang, C.; Yan, R.; Gao, R.X. Domain adaptive transfer kernel learning for fault diagnosis. *IEEE Transactions on Industrial Electronics* **2018**, *66*, 4492–4502.
32. Lee, J.; Kao, H.A.; Yang, S. Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters* **2014**, *1*, 38–41.
33. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *Journal of Big Data* **2016**, *3*, 9.
34. Tobias, S.A.; Fishwick, W. The chatter of lathe tools under orthogonal cutting conditions. *Proceedings of the Institution of Mechanical Engineers* **1958**, *172*, 389–402.
35. Tlusty, J. Self-excited vibrations in machine tools. In Proceedings of the International Research in Production Engineering. ASME, 1970, pp. 35–53.

36. Inasaki, I. Application of sensor fusion to machining monitoring. *Annals of the CIRP* **1998**, *47*, 653–656.
37. Fu, Y.; Hope, A.D.; Wang, M.; Liang, M. Chatter detection in milling process based on wavelet packets and Hilbert-Huang transform. *International Journal of Advanced Manufacturing Technology* **2006**, *29*, 1035–1041.
38. Chen, X.; Zheng, Y.; Wang, B.; Wang, Y. A review of machining monitoring systems based on artificial intelligence. *International Journal of Advanced Manufacturing Technology* **2015**, *81*, 585–605.
39. Wang, Y.; Gao, R.X.; Yan, R. A review of deep learning for smart manufacturing: Methods, applications, and challenges. *Journal of Manufacturing Systems* **2021**, *60*, 192–208.
40. Dimla Sr, D.E. Sensor signals for tool-wear monitoring in metal cutting operations—a review of methods. *International Journal of Machine Tools and Manufacture* **2000**, *40*, 1073–1098.
41. Li, X.; Ding, Q.; Sun, J.Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety* **2018**, *172*, 1–11.
42. Zhang, W.; Li, C.; Peng, G.; Chen, Y. A deep learning-based approach for automated fault diagnosis of rotating machinery. *Neurocomputing* **2019**, *338*, 190–204.
43. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing* **2018**, *108*, 33–47.
44. Kusiak, A. Smart manufacturing must embrace big data. *Nature* **2018**, *544*, 23–25.
45. Purushothaman, S.; Kiran, R.; Jose, M. Support vector machine approach for tool wear classification. *Procedia Engineering* **2014**, *97*, 2195–2203.
46. Li, L.; Li, D.; Huang, Q.; Huang, Z. Prediction of surface roughness in end milling using genetic algorithm and multiple regression method. *Frontiers of Mechanical Engineering* **2016**, *11*, 157–163.
47. Kwon, P.; Bacci, G. Artificial neural network approach to determination of optimal cutting conditions in milling operations. *Journal of Manufacturing Science and Engineering* **2018**, *140*, 095001.
48. Benardos, P.G.; Vosniakos, G.C. Prediction of surface roughness in CNC face milling using neural networks and Taguchi's design of experiments. *Robotics and Computer-Integrated Manufacturing* **2003**, *19*, 343–354.
49. Sikder, A.K.; Murshed, A.N. Unsupervised machine learning approach for sensor-based predictive maintenance in intelligent manufacturing. *Procedia Manufacturing* **2018**, *26*, 1239–1250.
50. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.D. Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research* **2016**, *4*, 23–45.
51. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
52. Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep learning-based data analytics for defect classification in manufacturing. *Procedia CIRP* **2016**, *55*, 512–517.
53. Zhao, Z.; Chen, W.; Wu, X.; Chen, P.C.; Liu, J. LSTM network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems* **2017**, *11*, 68–75.
54. Zhang, Y.; Tao, J.; Li, X.; Ding, Q. Deep autoencoder neural networks for noise reduction in machinery fault diagnosis. *Mechanical Systems and Signal Processing* **2019**, *127*, 1–18.
55. Wang, K.; Yung, C.; Li, D. A new paradigm of cloud-based predictive maintenance for intelligent manufacturing. *Journal of Intelligent Manufacturing* **2018**, *29*, 1125–1137.
56. Lu, J.; Behbood, V.; Hao, P.; Zuo, H.; Xue, S.; Zhang, G. Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems* **2017**, *115*, 1–14.
57. Doshi-Velez, F.; Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608* **2017**.
58. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems* **2017**, *30*, 4765–4774.
59. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016, pp. 1135–1144. https://doi.org/10.1145/2939672.2939778.
60. Tang, J.; Chen, X.; Ren, Y.; Liu, Z. Chatter detection in milling process using multi-scale entropy and ensemble empirical mode decomposition. *Journal of Intelligent Manufacturing* **2018**, *29*, 1333–1345.
61. Zhang, X.; Xu, Y.; Jin, X. Deep learning-based chatter detection in milling operations. *Mechanical Systems and Signal Processing* **2020**, *140*, 106683.
62. Butterworth, S. On the Theory of Filter Amplifiers. *Experimental Wireless & the Wireless Engineer* **1930**, *7*, 536–541.

63. Staszewski, W.J. Wavelet based compression and feature selection for vibration analysis. *Journal of Sound and Vibration* **1997**, *203*, 491–500.
64. Harris, F.J. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE* **1978**, *66*, 51–83.
65. He, H.; Garcia, E.A. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering* **2009**, *21*, 1263–1284.
66. Christ, M.; Braun, N.; Neuffer, J.; Kempa-Liehr, A.W. Time Series Feature Extraction on Basis of Scalable Hypothesis Tests (TSFresh)—A Python Package. *Neurocomputing* **2018**, *307*, 72–77.
67. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
68. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-Adversarial Training of Neural Networks. *Journal of Machine Learning Research* **2016**, *17*, 1–35.
69. Dietterich, T.G. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation* **1998**, *10*, 1895–1923.
70. Demšar, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research* **2006**, *7*, 1–30.
71. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362.
72. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **2020**, *17*, 261–272.
73. Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. In Proceedings of the 9th Python in Science Conference, 2010, Vol. 57, p. 61.
74. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **2007**, *9*, 90–95.
75. Waskom, M.L. Seaborn: Statistical Data Visualization, 2021. Version 0.11.1.
76. Donoho, D.L. Denoising by Soft-Thresholding. *IEEE Transactions on Information Theory* **1995**, *41*, 613–627.
77. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **1997**, *9*, 1735–1780.
78. Li, Y.F.; Liu, Z.Q.; Huang, H.L. Remaining useful life prediction based on degradation data with variability. *IEEE Transactions on Reliability* **2018**, *67*, 315–326.
79. Campbell, T.; Woxvold, I.; Ding, J. The influence of material microstructure on machining-induced surface integrity. *Procedia CIRP* **2018**, *71*, 329–334.
80. Duncan, G.S.; Schmitz, T.L. The impact of subjective labeling on machine learning for chatter detection. *International Journal of Machine Tools and Manufacture* **2005**, *45*, 411–420.
81. Lee, J.; Bagheri, B.; Kao, H.A. Cyber physical systems for predictive production systems. *Production Engineering* **2015**, *9*, 167–178.
82. Lu, Y. Industrial big data analytics in smart manufacturing: A review. *Journal of Manufacturing Systems* **2018**, *49*, 79–93.
83. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. *International Conference on Artificial Neural Networks* **2018**, pp. 270–279.
84. Chapelle, O.; Scholkopf, B.; Zien, A. Semi-Supervised Learning. *IEEE Transactions on Neural Networks* **2009**, *20*, 542–542.
85. John, S.S.; Alberts, M.; Karandikar, J.; Coble, J.; Jared, B.; Schmitz, T.; Ramsauer, C.; Leitner, D.; Khojandi, A. Predicting chatter using machine learning and acoustic signals from low-cost microphones. *The International Journal of Advanced Manufacturing Technology* **2023**, *125*, 5503–5518.
86. Tobon-Mejia, D.A.; Medjaher, K.; Zerhouni, N. A Review on Data-Driven Fault Severity Assessment in Rolling Element Bearings. *Mechanical Systems and Signal Processing* **2012**, *25*, 2083–2093.
87. Hoi, S.C.; Wang, J.; Zhao, P.; Jin, R. Online Learning: A Comprehensive Survey. *arXiv preprint arXiv:1802.02871* **2018**.