

Article

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

Matthew Alberts ^{1,†} , Sam St. John ^{1,†}, Simon Odie ^{1,2,†}, Anahita Khojandi ^{1,†} , Bradley Jared ¹ ,
Tony Schmitz ^{2,3} , Jaydeep Karandikar ² and Jamie B. Coble ^{4,*,†} 

¹ Department of Industrial and Systems Engineering, University of Tennessee, 851 Neyland Drive, Knoxville, TN 37996, USA; malberts@vols.utk.edu (M.A.); sstjohn3@vols.utk.edu (S.S.J.); sodie@vols.utk.edu (S.O.); khojandi@utk.edu (A.K.); bhjared@utk.edu (B.J.)

² Manufacturing Science Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37830, USA; tony.schmitz@utk.edu (T.S.); karandikarjm@ornl.gov (J.K.)

³ Department of Mechanical, Aerospace, and Biomedical Engineering, University of Tennessee, 1512 Middle Drive, Knoxville, TN 37996, USA

⁴ Department of Nuclear Engineering, University of Tennessee, 863 Neyland Drive, Knoxville, TN 37996, USA

* Correspondence: jamie@utk.edu

† 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. The primary goal of this study is to bridge the gap between simulation-based machine learning models and real-world applications by developing and validating a Random Forest-based chatter detection system. This research focuses on improving manufacturing efficiency through reliable chatter detection by integrating Operational Modal Analysis (OMA), Receptance Coupling Substructure Analysis (RCSA), and Transfer Learning (TL). The study applies 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 1600 real-world machining datasets, achieving an accuracy of 86.1%, with strong precision and recall scores. The 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



Citation: Alberts, M.; St. John, S.; Odie, S.; Khojandi, A.; Jared, B.; Schmitz, T.; Karandikar, J.; Coble, J.B. Transitioning from Simulation to Reality: Applying Chatter Detection Models to Real-World Machining Data. *Machines* **2024**, *12*, 923. <https://doi.org/10.3390/machines12120923>

Academic Editors: Ján Pitel' and Natalia Lishchenko

Received: 7 November 2024

Revised: 4 December 2024

Accepted: 13 December 2024

Published: 17 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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 workpiece, leading to poor surface finishes, dimensional inaccuracies, increased tool wear, and potential machine damage [3]. Detecting and mitigating chatter is crucial for maintaining high precision and productivity in modern machining operations.

Traditional methods for chatter detection rely on analytical models and signal processing techniques. These approaches often require extensive domain knowledge and are limited by assumptions that restrict their generalizability across diverse machining setups [4,5]. Such limitations highlight the need for methods capable of addressing variability in real-world operations.

Machine learning (ML) has emerged as a powerful alternative, enabling data-driven approaches to chatter detection by learning patterns from large datasets [6]. Models such as Random Forest classifiers and neural networks have shown promise in identifying chatter through the analysis of vibration signals and other sensor data [7]. These models can adapt to different machining conditions, providing real-time monitoring and enhancing predictive maintenance strategies by up to 40% [8].

Previous work developed a Random Forest classifier trained on extensive simulated datasets to predict chatter occurrences [9]. This model demonstrated high accuracy in simulations, showcasing its potential for detecting chatter in controlled environments. However, challenges arose when transitioning to real-world applications, including sensor noise, unmodeled dynamics, and reliance on predefined features that limited generalization across diverse machining scenarios. Building on this foundation, subsequent research incorporated additional simulated data and advanced modeling techniques such as Operational Modal Analysis (OMA), Transfer Learning (TL), and Receptance Coupling Substructure Analysis (RCSA) [10–13]. These enhancements improved predictive accuracy and robustness within simulations, but real-world complexities, such as environmental fluctuations and sensor imperfections, remained challenging to address [14,15].

To ensure practical utility, it is essential to validate these models using real-world machining data and refine them to address domain-specific challenges [14]. This study evaluates a Random Forest classifier's performance on operational datasets and proposes enhancements for real-world deployment. Milling, chosen for its prevalence in manufacturing, its well-documented dynamics, and the availability of stability lobe diagrams, serves as the testbed for this analysis. By applying these models to actual vibration data, this research seeks to bridge the gap between simulation and practice, offering insights into improving robustness and applicability for industrial machining operations.

1.1. Previous Work

Prior research focused on developing ML models for chatter detection using simulated data from machining processes. In [9], a Random Forest classifier was introduced, trained on a comprehensive dataset generated from milling operation simulations. The dataset encompassed various machining conditions, such as spindle speeds, feed rates, and depths of cut, providing a robust foundation for identifying patterns associated with chatter.

The initial model achieved high accuracy in predicting chatter within simulated environments. Random Forests demonstrated their ability to handle high-dimensional feature spaces while offering insights into feature importance, a critical factor for understanding chatter [16].

Building on this work, ref. [10] expanded the dataset to over 140,000 simulated instances and incorporated advanced techniques, including OMA, TL, and RCSA. OMA enabled the extraction of modal parameters from operational responses, improving the model's capacity to capture dynamic machining behavior [11]. TL enhanced adaptability to unseen machining conditions, addressing variability in operational environments [12]. RCSA further deepened the understanding of tool-holder assembly dynamics, enabling more accurate predictions of machining stability [13].

These enhancements significantly improved predictive accuracy and robustness, especially under complex machining scenarios. However, the reliance on simulated data, while comprehensive, limited the models' applicability to real-world machining, where factors such as sensor noise, machine wear, and environmental fluctuations introduce unmodeled complexities [14,15].

Recognizing these challenges, the current study seeks to validate the models developed in prior research using real-world machining data. By applying these models to vibration signals collected from operational milling machines, this work aims to assess their practical applicability, identify performance gaps, and propose refinements to enhance robustness in real-world settings.

1.2. Simulation-Based Models

Simulation-based models are pivotal in the development and testing of ML techniques for chatter detection in machining processes. Simulated data provide a controlled environment where key parameters—spindle speed, cutting depth, feed rate, and tool geometry—can be systematically varied. This approach allows researchers to investigate machining dynamics comprehensively and gain insights that may be costly, time-consuming, or impractical to obtain through real-world experiments.

A significant advantage of simulation-based models is their capacity to generate extensive datasets across a wide range of operational conditions. Prior studies simulated diverse cutting scenarios to capture both stable and unstable machining states, enabling ML models to learn complex patterns associated with chatter onset and predict machining stability under varying conditions [17,18].

Key parameters influence the effectiveness of these simulations:

- **Spindle Speed:** Simulating variations in spindle speed reveals its impact on the machine-tool system's natural frequency response, where specific speeds can amplify or suppress vibration tendencies, affecting chatter stability [19,20].
- **Cutting Depth:** This parameter affects cutting force magnitudes and system dynamics, helping to define thresholds beyond which machining becomes unstable and enabling the creation of stability lobe diagrams [21].
- **Feed Rate and Tool Geometry:** These parameters govern the interaction between the cutting tool and the workpiece. For example, feed rate affects chip height, which must exceed a minimum undeformed chip thickness for efficient material removal. Suboptimal feed rates can transition cutting into a rubbing regime, destabilizing the system and increasing chatter likelihood. Similarly, tool geometry, including rake angle and edge radius, shapes cutting forces and vibration characteristics, influencing chatter dynamics [14].

By systematically simulating variations in these parameters, researchers develop ML models with enhanced robustness and adaptability across diverse machining scenarios [22]. This systematic approach facilitates the generation of stability lobe diagrams and supports predictive analytics in machining.

While simulations are invaluable for training and testing ML models, they cannot fully replicate the variability of real-world environments. Factors such as sensor noise, machine wear, environmental fluctuations, and material inconsistencies introduce complexities absent in simulations, often leading to performance discrepancies when models are applied to real-world machining data [23,24]. Addressing these gaps is critical for ensuring the reliability and practicality of simulation-trained models in industrial applications.

1.3. Problem Statement

While previous studies have demonstrated the potential of ML models, particularly Random Forest classifiers, in predicting chatter using simulated data [9,10], a critical gap remains in validating these models against real-world machining data. Simulated datasets, though comprehensive and controlled, fail to fully capture the complexities and uncertainties of real-world machining environments [14].

Real-world machining data are influenced by factors such as sensor noise, machine tool wear, environmental variations, and material inconsistencies [8]. These elements introduce discrepancies between simulated and real-world data distributions, which can degrade the performance of models trained exclusively on simulations [15]. Furthermore, unmodeled dynamics and unforeseen interactions within physical systems often create challenges that are either absent or oversimplified in simulations [25].

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

1. **Data Discrepancies:** Variations in data distributions between simulated and real-world datasets can lead to reduced model performance, caused by covariate shift and sample selection bias [26].

2. **Sensor Noise and Data Quality:** Real-world data often contain noise and artifacts absent in simulated environments, necessitating advanced preprocessing and noise reduction techniques [24].
3. **Feature Relevance:** Features significant in simulations may lose importance in real-world settings, requiring re-evaluation of feature selection and extraction methods [27].
4. **Model Generalization:** Ensuring that models generalize well to unseen real-world data is essential for industrial application, underscoring the need for model adaptation and validation strategies [12].

Addressing these challenges is critical for deploying chatter detection models effectively in industrial settings. Without validation on real-world data, their practical utility remains limited. To enhance robustness and generalizability, this study evaluates the performance of previously developed models using real-world machining data and proposes refinements to address observed limitations.

To bridge the gap between simulation and practice, this study performs the following:

- Collects and processes 1600 real-world machining datasets representing diverse operational conditions.
- Applies previously trained models to these datasets to evaluate their predictive performance.
- Identifies discrepancies between simulated and real-world data, analyzing their impact on model accuracy.
- Proposes strategies for addressing these discrepancies, including model adaptation techniques and enhanced feature extraction methods.

These efforts aim to enhance the practical applicability of ML models for chatter detection and contribute to advancements in predictive maintenance for machining processes. It is important to note that the real-world data used in this study was obtained from a controlled environment and may not represent the full complexity of an industrial machine shop.

2. Literature Review

2.1. Chatter Detection in Machining

Chatter significantly affects surface finish, dimensional accuracy, tool life, and productivity [19]. Its detection and mitigation have been extensively researched, with traditional methods focusing on analytical models and signal processing techniques applied to machining data.

Early approaches relied on stability analysis using machining dynamics models. The regenerative chatter theory introduced by Tobias and Fishwick [28] established that vibrations from previous cutting passes influence current ones, leading to instability. Building on this, Tlustý [29] developed stability lobe diagrams to guide parameter selection and avoid chatter.

Signal processing techniques have also played a critical role in chatter detection, analyzing vibration signals from machine tool sensors. Time-domain, frequency-domain, and time-frequency methods, such as Fourier Transform (FT) and Short-Time Fourier Transform (STFT), have been widely used to detect chatter by identifying changes in the frequency spectrum [2,30]. Wavelet Transform (WT) methods further advanced detection by capturing transient chatter features in the time-frequency domain [31].

However, traditional approaches often require expert interpretation and may not generalize well across varying machining conditions or machine tools [32]. Real-world machining environments introduce complexity and variability, limiting the industrial applicability of these methods.

In recent years, ML techniques have emerged as powerful tools for chatter detection. Supervised learning algorithms, including Support Vector Machines (SVM) and Artificial Neural Networks (ANN), have been applied to classify machining states and predict chatter based on sensor-derived features [8,33]. These methods address non-linear and stochastic aspects of chatter phenomena, offering a data-driven alternative to traditional approaches.

Deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have demonstrated even greater potential by automatically extracting hierarchical features from raw data [34]. Their success in handling large and complex datasets has yielded improved performance in chatter detection tasks [35].

Despite their promise, ML methods face challenges when applied to real-world machining data. Issues like data scarcity, noise, imbalance, and variability can hinder model performance [36]. Additionally, models trained on specific machines or conditions often fail to generalize, necessitating strategies like TL and domain adaptation to improve applicability across diverse settings [12].

This study builds on existing research by applying ML models trained on simulated data to real-world machining environments. Addressing the challenges of transitioning from simulation to reality contributes to the evolution of chatter detection methodologies and supports their practical implementation in industrial settings.

2.2. Machine Learning in Machining Processes

ML has emerged as a transformative tool in machining processes, driven by increasing data availability and advancements in computational power [6]. ML techniques effectively model the complex, non-linear relationships inherent in machining operations, enabling advancements in process monitoring, fault diagnosis, and predictive maintenance [37].

Supervised learning methods, such as Support Vector Machines (SVMs), Decision Trees, and Random Forests, have been widely applied in tool condition monitoring and fault diagnosis. SVMs classify tool wear states using features derived from vibration signals, acoustic emissions, and cutting forces [38]. Decision Trees and Random Forests predict surface roughness and tool wear, leveraging their ability to handle high-dimensional data and provide insights into feature importance [39]. Artificial Neural Networks (ANN) have been used to predict cutting forces, surface finish, and dimensional accuracy based on machining parameters like spindle speed, feed rate, and depth of cut [40,41].

Unsupervised learning techniques, such as K-means and hierarchical clustering, identify patterns and anomalies in machining data without requiring labeled datasets [42]. These methods facilitate the segmentation of machining states and the detection of novel fault conditions by analyzing similarities in data features [43].

Deep learning, a subset of ML, has demonstrated significant potential in machining applications by automatically extracting hierarchical features from raw sensor data [44]. Convolutional Neural Networks (CNNs) have been employed for tool condition monitoring using time-series data, achieving higher accuracy than traditional feature-based methods [45]. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, excel at capturing temporal dependencies in sequential data, improving predictions of tool wear progression [46]. Additionally, Autoencoders and Generative Adversarial Networks (GANs) have been explored for unsupervised feature learning and anomaly detection, compressing data to highlight essential features for fault detection and process optimization [47].

Despite these advancements, challenges remain in applying ML to real-world machining data. Issues such as data scarcity, noise, imbalance, and variability can hinder model performance [36]. Furthermore, models trained on specific machines or conditions often fail to generalize to other settings, necessitating approaches like TL and domain adaptation to enhance their applicability [12].

This study builds on the growing body of work in ML for machining by evaluating models developed with simulated data in real-world environments. By addressing the challenges of transitioning from simulation to reality, this research contributes to the advancement of reliable, industry-ready chatter detection systems.

2.3. Challenges in Transitioning from Simulated to Real-World Data

The transition from simulated to real-world data introduces several challenges that must be addressed to ensure effective model performance. While simulated data offer

controlled conditions and systematic parameter variation, real-world machining environments are characterized by noise, variability, and unforeseen complexities that complicate model generalization.

Sensor Noise and Data Quality: Real-world vibration data are susceptible to various sources of noise, including electrical interference, mechanical vibrations from surrounding equipment, and environmental disturbances [24]. These noise factors can obscure critical patterns required for accurate chatter detection. Advanced signal processing techniques, such as low-pass filtering and baseline correction, were employed to mitigate these effects, but completely eliminating noise remains challenging.

Variability in Machining Conditions: Unlike the controlled environments of simulations, real-world machining processes exhibit significant variability due to factors such as tool wear, material inconsistencies, and fluctuating environmental conditions [48]. These variabilities introduce domain discrepancies, making it difficult for models trained exclusively on simulated data to generalize effectively.

Feature Distribution Shifts: An analysis of feature distributions revealed shifts between simulated and real-world datasets. For example, the distribution of FFT coefficients and modal parameters (natural frequencies and damping ratios) varied significantly due to unmodeled dynamics in real-world operations. These shifts impact the model's ability to identify chatter patterns reliably.

Data Labeling Challenges: Real-world data lack the precise control and labeling inherent in simulations. Chatter labeling is often based on subjective assessments of vibration signals, acoustic monitoring, and surface finish inspection, introducing potential biases. Additionally, transient chatter events can complicate the assignment of labels to continuous data segments.

Model Generalization Challenges: Models trained on simulated data are prone to overfitting to patterns specific to the simulated domain. In real-world applications, unmodeled dynamics, sensor variability, and noise can lead to reduced performance, necessitating strategies like TL and domain adaptation to bridge the gap between domains [12].

Addressing Discrepancies: Techniques such as TL, domain adversarial training, and augmentation with real-world data subsets proved effective in adapting the models. Fine-tuning pre-trained models on a small portion of real-world data improved their ability to generalize while retaining the knowledge acquired from simulations.

This discussion underscores the need for robust model adaptation strategies and comprehensive validation against real-world datasets. By addressing these challenges, simulation-trained ML models can achieve the reliability required for industrial applications, paving the way for broader adoption in machining processes.

Application to Chatter Detection

ML models have been extensively used in chatter detection to classify stable and unstable cutting conditions. Features extracted from vibration signals—such as statistical measures, frequency components, and wavelet coefficients—are commonly used as inputs for classifiers like SVM, ANN, and Random Forests [49]. These models leverage extracted features to distinguish between chatter-free and chatter-prone states, offering significant improvements in detection capabilities.

Deep learning models further enhance chatter detection by automatically learning relevant features from raw sensor data, eliminating the need for manual feature engineering. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have demonstrated superior accuracy in detecting chatter compared to traditional methods [50].

Previous studies [9,10] developed ML models using simulated data for chatter prediction, achieving high accuracy under controlled conditions. The current study extends this work by applying these models to real-world machining data, addressing challenges related to model transferability and generalization. This effort seeks to enhance the practical applicability of ML-based chatter detection systems in industrial environments.

2.4. Gap Identification

Despite significant advancements in applying ML techniques to machining process monitoring and chatter detection, a critical gap remains in validating simulation-trained models with real-world machining data. Many studies have demonstrated the efficacy of ML models in predicting machining stability and detecting chatter using simulated datasets or controlled laboratory experiments [9,10,50]. While these efforts have achieved promising results, their practical applicability in industrial environments is often limited.

Real-world machining conditions introduce complexities such as sensor noise, machine tool variability, environmental fluctuations, and unmodeled dynamics that are not fully represented in simulated datasets [14,36]. These discrepancies between simulated and real-world data can lead to reduced model performance, hindering the deployment of ML-based solutions in industrial settings [26].

Currently, there is a lack of empirical research addressing the following key issues:

1. **Systematic Evaluation of Model Transferability:** Few studies have systematically assessed how simulation-trained models perform in real-world machining operations and identified the factors contributing to performance degradation.
2. **Mitigating Domain Discrepancies:** Limited efforts have been made to develop and implement strategies for addressing the impact of domain discrepancies between simulated and real-world data in machining contexts.
3. **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 process monitoring.

This gap hampers the advancement of smart manufacturing initiatives and the adoption of advanced predictive maintenance strategies in the machining industry. Bridging this gap is essential for enhancing the reliability and robustness of ML models, facilitating their integration into real-world manufacturing processes, and ultimately improving productivity and product quality.

To address this gap, this study validates the performance of ML models developed using simulated data on real-world machining datasets. It also analyzes the challenges encountered during the transition and proposes strategies to enhance model robustness and applicability in industrial environments.

2.5. Real-World Data Collection

This study evaluated three ML models for chatter detection:

1. **Baseline Random Forest Model:** Trained exclusively on simulated machining data to assess performance under controlled conditions.
2. **Enhanced Random Forest Model:** Incorporated advanced features such as modal parameters extracted using OMA and frequency response data from RCSA to capture dynamic machining behavior.
3. **Real-World Tuned Model:** Applied TL techniques to adapt the enhanced model for real-world machining data, accounting for noise, sensor inconsistencies, and environmental variability through parameter fine-tuning.

Real-world data were collected using a custom-built three-axis Computer Numeric Controlled (CNC) milling machine, designed to replicate the functionality and dynamics of commercial systems. This configuration ensured a realistic representation of industrial machining conditions while maintaining flexibility for research needs. The experimental setup enabled precise control over machining parameters, ensuring repeatable and consistent conditions critical for model evaluation.

The CNC milling machine featured high-precision ball screws and linear guides across its three linear axes (X, Y, and Z), providing accurate and repeatable movements. The spindle offered variable speeds ranging from 1000 to 16,000 RPM, supporting a wide range of cutting conditions. A Marposs MEMS (Micro-Electro-Mechanical Systems) vibration sensor was integrated into the setup, facilitating high-fidelity data acquisition tailored for chatter

analysis. The robust design enabled systematic exploration of machining dynamics across both stable and chatter-prone conditions.

The experimental setup also allowed for the collection of a large, diverse dataset, enhancing the study's robustness by capturing various operational scenarios. Figure 1 illustrates the experimental setup.

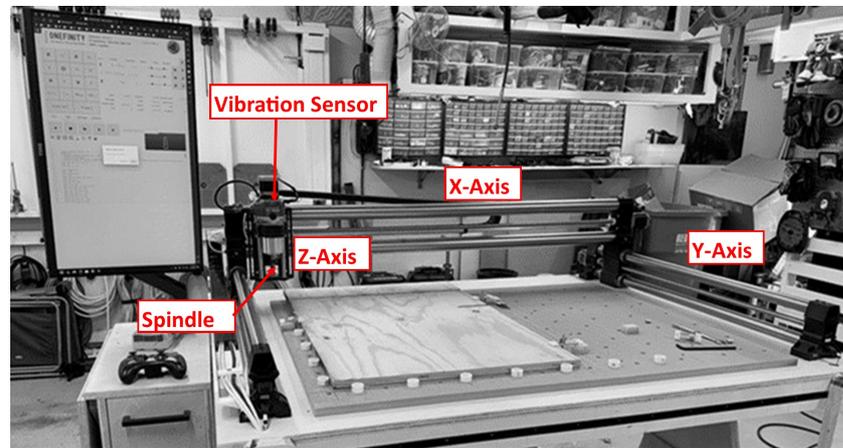


Figure 1. Three -axis CNC milling machine equipped with a Marposs MEMS vibration sensor.

By leveraging this setup, the study provided a reliable platform for validating ML models under conditions representative of real-world industrial operations.

2.6. Data Acquisition and Machining Operations

To capture vibrational responses during machining, a Marposs MEMS vibration sensor was installed on the spindle housing. The sensor provides high-resolution acceleration measurements along three axes, with 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, sampling vibration signals at 20 kHz. 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, ensuring comprehensive data for analysis.

A robust dataset comprising 1600 individual machining runs was collected to validate the ML models. Each machining pass was performed with specific parameter combinations designed to capture both chatter and chatter-free conditions. Parameter selection was guided by the stability lobe diagram for the tool-workpiece setup, experimentally determined prior to data collection. This approach ensured that the dataset covered a wide range of machining dynamics and realistic operational scenarios.

Dataset Composition:

- **Materials:** Machining operations were performed on two distinct materials, 6061 Aluminum and 304 Stainless Steel, to study how material properties affect chatter dynamics.
 - **6061 Aluminum:** Known for its high machinability and thermal conductivity, this material served as a baseline for evaluating model accuracy and robustness. Its predictable behavior under varying machining conditions provided insights into chatter dynamics in moderate-strength, corrosion-resistant materials.
 - **304 Stainless Steel:** A harder material with lower thermal conductivity, 304 Stainless Steel was chosen to capture the challenges posed by harder workpieces. Its properties intensify heat retention and vibration, testing the model's adaptability to diverse machining conditions.
- **Tool Configurations:** Machining was performed with carbide end mills featuring varying numbers of cutting teeth:

- **2 Cutting Teeth:** This configuration, commonly used in industry, served as a baseline for comparison.
- **4 Cutting Teeth:** Studied for its influence on vibration amplitude and frequency, providing insights into the impact of additional cutting edges on chatter behavior.
- **Machining Parameters:** A range of parameters was systematically tested to ensure comprehensive coverage:
 - **Spindle Speed:** Varied from 8000 to 10,000 RPM in 500 RPM increments, capturing conditions likely to induce chatter.
 - **Cutting Depth:** Adjusted from 1.0 mm to 2.0 mm in 0.2 mm increments to analyze the effect of material removal rates on chatter onset.
 - **Feed Rate:** Tested between 0.1 mm/rev and 0.3 mm/rev to evaluate the relationship between chip formation and vibration patterns.

To isolate the effects of tool configurations, spindle speed and feed per tooth were kept consistent when switching tools. Tool wear was monitored closely, and inserts were replaced after every 100 machining cycles to minimize variability and maintain consistent tool performance.

This experimental setup ensured a comprehensive dataset for evaluating model performance, encompassing diverse operational scenarios representative of real-world machining conditions. The systematic approach to data collection provides a robust foundation for validating ML models and analyzing their adaptability to industrial applications.

2.7. Data Preprocessing and Feature Engineering

Effective data preprocessing and feature engineering are critical for transforming raw vibration data into a format suitable for ML models. The workflow implemented in this study included noise reduction, signal segmentation, normalization, and advanced feature extraction techniques to ensure data quality and enhance predictive performance.

Noise Reduction: Real-world vibration data are often contaminated by noise from electrical interference, mechanical vibrations, and sensor imperfections. To mitigate these effects, a Butterworth low-pass filter with a cutoff frequency of 5 kHz was applied to attenuate high-frequency noise while preserving relevant signal characteristics. Baseline correction removed signal drift, and abnormal spikes were detected and replaced using linear interpolation to maintain signal continuity.

Signal Segmentation and Normalization: Continuous vibration data were segmented into individual machining passes corresponding to specific combinations of operational parameters. This ensured alignment between vibration signals and machining conditions. Each segment was then normalized by scaling to unit variance, addressing variations in amplitude and ensuring consistency across datasets.

Feature Extraction: To capture informative patterns in the vibration data, features were extracted from both the time and frequency domains using established techniques:

- **Time-Series Features:** The TSFresh library was employed to automatically extract a comprehensive set of statistical, spectral, and information-theoretic features. Key examples include FFT coefficients, permutation entropy, and linear trend aggregations, which are critical for identifying dynamic characteristics of chatter.
- **Frequency-Domain Features:** The Fast Fourier Transform (FFT) was used to analyze the frequency components of the vibration signals. Derived features included peak frequency, root mean square (RMS) amplitudes across different frequency bands, and crest factors, which are known indicators of machining stability.
- **Modal Parameters (OMA):** OMA provided natural frequencies, damping ratios, and mode shapes, capturing the dynamic response of the machining system under operational conditions.
- **Dynamic System Features (RCSA):** RCSA features included frequency response functions (FRFs), dynamic stiffness, and compliance, which modeled the interactions between the tool and machine structure.

By combining these features, the dataset represented a comprehensive view of machining dynamics, capturing both transient and steady-state behaviors. This feature engineering approach ensured that the ML models were equipped with robust and discriminative inputs, enabling effective chatter detection and prediction across varying machining scenarios.

Data Labeling: Each segmented signal was labeled as either *chatter* or *stable* based on a combination of vibration signal analysis, acoustic monitoring, and surface inspection:

- **Vibration Analysis:** High amplitude and irregular vibrations in the signal indicated chatter conditions. Signals were recorded at 20 kHz to capture high-frequency dynamics.
- **Acoustic Monitoring:** Audible noise characteristic of chatter served as a supplementary indicator.
- **Surface Inspection:** Workpiece surfaces were visually examined for chatter marks or patterns.

Figure 2 illustrates examples of labeled vibration signals under stable and chatter conditions. This preprocessing workflow ensured high-quality data for robust feature extraction and reliable model training, addressing real-world challenges such as noise, variability, and inconsistencies in machining conditions.

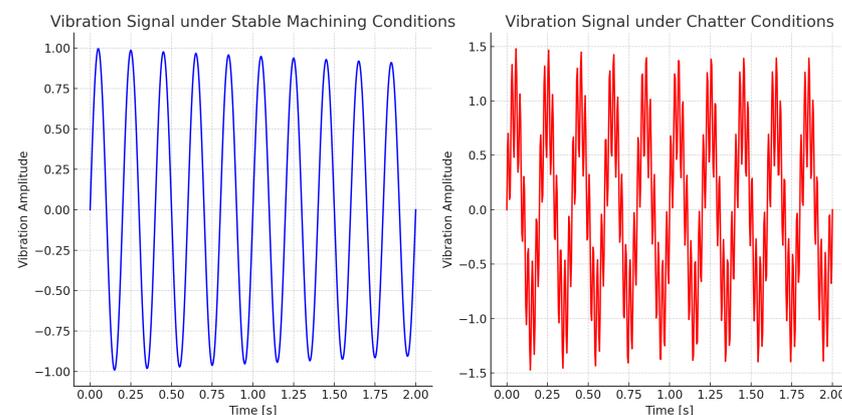


Figure 2. Stable machining conditions (left): The signal appears as a smooth, decaying sinusoidal wave. Chatter conditions (right): The signal becomes irregular, with higher frequency components causing intense vibrations.

2.8. Data Partitioning and Feature Extraction

Data Partitioning: The dataset was split into 10% for training, 50% for validation, and 40% for testing. This distribution prioritized robust model evaluation while ensuring sufficient data for training and hyperparameter tuning. The 10% training set enabled the models to learn fundamental patterns in the data, while the 50% validation set facilitated comprehensive hyperparameter optimization, particularly for models like Random Forest, which are prone to overfitting. The larger 40% test set ensured that final performance metrics were computed on a substantial and representative portion of the dataset, providing a reliable assessment of model generalization.

Stratified sampling was employed to maintain the proportion of chatter and stable instances across subsets, addressing potential class imbalance issues [51].

Feature Extraction: To transform the preprocessed vibration data into a suitable format for ML models, features were extracted using three approaches: time-series feature extraction, frequency-domain analysis, and OMA. Consistent with prior studies [9,10], the same set of features was extracted to ensure comparability between simulated and real-world datasets.

1. Time-Series Feature Extraction: The TSFresh library [52] was employed to compute a comprehensive set of time-series features. TSFresh calculates statistical, spectral, and information-theoretic characteristics, including the following:

- **Ratio value number to time series length:** The ratio of unique values to the total number of values.
- **Benford correlation:** Measures the frequency of initial digits in data, often dominated by the digit 1.
- **FFT coefficients:** Imaginary parts of FFT coefficients at specified indices, e.g., coefficients 55 and 77.
- **Permutation entropy:** A complexity measure quantifying the frequency of permutations in the time series.
- **Linear trend metrics:** Statistical descriptors of linear regression trends over specific segments.

2. Frequency-Domain Features: Features were extracted using the FFT to analyze dominant frequency components associated with machining dynamics. Key FFT features include the following:

- **Acceleration Peak (g):** Maximum acceleration value.
- **Acceleration RMS (g):** Root mean square of acceleration.
- **Crest Factor:** Ratio of peak acceleration to RMS acceleration.
- **Standard Deviation (g):** Dispersion of acceleration values.
- **Frequency RMS:** RMS values computed over specified frequency bands, e.g., 1–65 Hz, 65–300 Hz, and 300–6000 Hz.
- **Velocity and Displacement RMS:** Derived from acceleration using frequency relationships.

3. OMA Features: Modal parameters were extracted to characterize the dynamic behavior of the machining system [11]. These include the following:

- **Natural Frequencies:** Fundamental oscillation frequencies of the system.
- **Damping Ratios:** Rates of energy dissipation from vibrations.
- **Mode Shapes:** Deformation patterns during vibrations at natural frequencies.
- **Modal Scale Factors:** Relative contributions of each mode shape to system vibrations.
- **Modal Assurance Criterion (MAC):** Statistical metric for mode shape consistency, aiding in detecting dynamic changes.

By combining these feature extraction techniques, the study ensured a comprehensive representation of machining dynamics, enabling the ML models to effectively distinguish between chatter and stable conditions. The extracted features were crucial for understanding system behavior and optimizing model performance.

2.9. Receptance Coupling Substructure Analysis and Transfer Learning

RCSA was employed to model the dynamic interaction between the tool and machine structure, providing insights into the system's overall behavior [13]. The features derived from RCSA include the following:

- **FRFs:** These functions characterize how system components respond to inputs across varying frequencies, enabling predictions of system behavior under dynamic excitations.
- **Coupling Stiffness and Mass Matrices:** Mathematical representations of dynamic connections between machine components, capturing how stiffness and mass properties influence system interactions.
- **Assembled System FRFs:** The combined FRFs of the entire system, derived using RCSA, predict how modifications in structure or configuration impact overall dynamic behavior.
- **Dynamic Stiffness and Compliance:** Metrics that quantify the system's resistance (stiffness) and responsiveness (compliance) to dynamic forces, critical for assessing stability and performance under operational conditions.

These RCSA-derived features provide a deeper understanding of the system's dynamic properties, enabling the identification of factors contributing to machining stability and chatter.

To bridge the gap between simulated and real-world machining environments, TL was utilized [12]. TL leverages knowledge gained from a source domain (simulated data)

and applies it to a target domain (real-world data), reducing the need for extensive labeled datasets in the target domain. The models pre-trained on simulated data were fine-tuned on real-world data to address domain-specific discrepancies while retaining previously learned patterns.

Key challenges addressed through TL include the following:

- **Domain Shift:** Differences in feature distributions between simulated and real-world datasets, such as noise levels and sensor variability, were mitigated by fine-tuning model parameters on domain-specific data.
- **Data Scarcity:** The smaller size of labeled real-world datasets was offset by transferring knowledge from the larger, simulated datasets.
- **Feature Relevance:** TL adjusted feature importance to reflect differences in their significance across the two domains, aligning the model's focus with real-world dynamics.

Fine-tuning involved updating model parameters with a smaller learning rate to retain general knowledge from simulations while adapting to real-world data complexities. This approach enhanced model accuracy and robustness in detecting chatter under real-world conditions, demonstrating TL's utility in addressing domain discrepancies.

By integrating RCSA and TL, this study provides a comprehensive framework for understanding system dynamics and adapting ML models to practical machining environments, ensuring improved performance and generalization.

2.10. Final Feature Set and Model Application

The final feature set mirrors that of previous studies [9,10], ensuring consistency for valid performance comparison between simulated and real-world data. The feature set includes:

- **10 FFT Features:** Frequency-domain characteristics of the vibration signals.
- **7 Time-Series Features:** Extracted using the TSFresh library, including statistical and spectral features.
- **OMA Features:** Natural frequencies and damping ratios characterizing system dynamics.
- **RCSA Features:** Coupled FRFs and dynamic stiffness metrics.

This consistency in feature design allows for the direct application of previously developed models to real-world datasets without introducing new variables, facilitating a robust evaluation of model performance.

The ML models developed in previous studies [9,10] were applied to the real-world machining datasets to evaluate their practical applicability for chatter detection. The workflow involved model loading, adaptation, and testing.

2.10.1. Model Loading and Preparation

The Random Forest classifiers from prior studies were implemented using Python's scikit-learn library [53] and saved via the joblib module, which preserved their structure and learned parameters. To ensure seamless application to the new datasets, the following steps were taken:

- **Environment Setup:** A consistent computational environment was established, replicating the software versions and dependencies used during model training.
- **Model Loading:** Pre-trained models were loaded using joblib's `load` function, ensuring compatibility with the new data.
- **Feature Alignment:** The feature set extracted from the real-world data was verified to match the original feature set, ensuring identical ordering, scaling, and encoding.

2.10.2. Testing Procedure

The loaded models were applied to the preprocessed and feature-extracted real-world datasets following these scenarios:

1. **Scenario 1:** Evaluation of the real-world dataset using the model from the first study [9], which was trained exclusively on simulated data.

2. **Scenario 2:** Evaluation using the model from the second study [10], which incorporated advanced simulated features like OMA and RCSA.
3. **Scenario 3, Model Adaptation:** The original models were adapted to the real-world domain using TL techniques [12]. This process included the following:
 - **Fine-Tuning:** Updating model parameters with 10% of the real-world training data to better capture domain-specific patterns while retaining simulated knowledge.
 - **Domain Adaptation:** Techniques such as domain adversarial training were considered to reduce discrepancies between simulated and real-world data distributions [54].

The adapted models were validated on 40% of the real-world dataset, and the best-performing model was selected based on metrics such as accuracy, precision, recall, and F1-score. Final generalization performance was assessed on the remaining 50% test set, providing a comprehensive evaluation of model applicability in real-world machining conditions.

This methodology ensured that the models developed using simulated data were systematically adapted and rigorously evaluated for real-world scenarios, bridging the gap between theoretical advancements and industrial applicability.

2.11. Evaluation Metrics

To assess the performance of the ML models on real-world machining data, a comprehensive set of evaluation metrics consistent with prior studies [9,10] was employed. These metrics facilitate a direct comparison between results obtained from simulated and real-world data, providing quantitative measures of predictive capabilities.

2.11.1. Classification Metrics

The primary classification metrics include the following:

- **Accuracy:** The proportion of correctly classified instances among all instances:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- **Precision:** The proportion of correctly predicted positive instances among all predicted positives:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

- **Recall:** The proportion of correctly predicted positive instances among all actual positives:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

- **F1-Score:** The harmonic mean of precision and recall, balancing the two:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

- **Area Under the Receiver Operating Characteristic Curve (AUC-ROC):** Represents the model's ability to distinguish between classes at various thresholds:

$$\text{AUC-ROC} = \int_0^1 \text{TPR}(\text{FPR}) d\text{FPR} \quad (5)$$

2.11.2. Confusion Matrix

The confusion matrix provides a detailed breakdown of classification performance by displaying the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). This metric helps identify specific error patterns.

2.11.3. Statistical Significance Testing

To determine whether differences in model performance between simulated and real-world data are statistically significant:

- **McNemar's Test:** Evaluates differences in error rates for paired nominal data [55].
- **Paired *t*-test:** Compares mean differences in performance metrics across data splits or model configurations [56].

2.11.4. Cross-Validation Metrics

k-fold cross-validation with $k = 5$ was applied to the training set to evaluate model stability and generalizability. Metrics were averaged across folds for a comprehensive assessment.

2.11.5. Receiver Operating Characteristic (ROC) Curve

The ROC curve illustrates the trade-off between sensitivity (True Positive Rate, TPR) and specificity (1 - False Positive Rate, FPR) at various threshold settings, offering insights into diagnostic ability.

2.11.6. Computational Efficiency Metrics

For practical deployment in real-time monitoring, computational efficiency was evaluated:

- **Inference Time:** Measures the time required for the model to make predictions on new data, critical for real-time applications.
- **Memory Consumption:** Assesses the memory required during inference, relevant for systems with constrained resources.

2.11.7. Evaluation Procedure

The evaluation metrics were applied in the following steps:

1. Calculated on the validation set during model fine-tuning to guide hyperparameter adjustments.
2. Computed on the test set to assess final model performance.
3. Compared against simulated data results to identify performance discrepancies and assess domain adaptation effectiveness.

These metrics provide a holistic evaluation of model performance, offering insights into their accuracy, generalizability, and practical suitability for real-world machining applications. By analyzing both classification and computational efficiency metrics, the study identifies strengths and areas for improvement in ML-based chatter detection models.

2.12. Statistical Analysis

A comprehensive statistical analysis validated model performance and provided insights into chatter detection factors. Key analyses included feature evaluation, statistical validation, and error performance analysis.

2.12.1. Feature Analysis and Importance

Descriptive statistics quantified the central tendencies and dispersion of features in both simulated and real-world datasets, using measures such as mean, standard deviation, skewness, and kurtosis. These statistics revealed similarities and disparities that inform the challenges of transitioning to real-world data.

Correlation analysis employed Pearson and Spearman coefficients to identify features significantly related to chatter occurrence. This dual approach highlighted both linear and monotonic relationships, revealing predictive features that drive model decisions.

Random Forest feature importance analysis ranked features based on their contribution to classification accuracy. ANOVA tests further examined the impact of categorical variables, such as spindle speed and feed rate, on performance metrics. Together, these analyses provided a comprehensive understanding of key predictors and operational conditions influencing chatter detection.

2.12.2. Statistical Validation

Hypothesis testing compared model performance metrics across datasets:

- **Paired *t*-test:** Assessed differences in mean metrics, such as accuracy and F1-score.
- **McNemar's Test:** Evaluated discrepancies in classification errors.

Confidence intervals (CIs) were computed for metrics like accuracy to provide statistical assurance of reliability. For example, a 95% CI quantified the expected range for accuracy under varying conditions, ensuring robust interpretations of results.

2.12.3. Error and Performance Analysis

Error analysis categorized misclassifications into false positives (FPs) and false negatives (FNs):

- **False Positives:** Often caused by high-frequency noise or transient vibration spikes during abrupt operational changes.
- **False Negatives:** Commonly associated with lower spindle speeds or tool wear, where subtle signals were misclassified as stable.

ROC curves were plotted to visualize the trade-off between sensitivity and specificity. AUC-ROC values encapsulated the model's discriminative ability, complementing insights from error analysis. For instance, thresholds minimizing FP or FN rates were identified, guiding model tuning.

2.13. Summary of Findings

The statistical analysis validated evaluation metrics, identified critical features, and provided actionable insights for improving model robustness. By analyzing feature importance, operational impacts, and errors, this study bridges the gap between simulated and real-world chatter detection scenarios.

Statistical Software, Tools, and Integrated Analyses

All statistical analyses were conducted using Python-based libraries:

- **NumPy [57]:** For numerical computations and array manipulations.
- **SciPy 1.14.0 [58]:** For advanced statistical tests, including paired *t*-tests and McNemar's Test.
- **statsmodels [59]:** For hypothesis testing, confidence interval estimation, and regression analysis.
- **Matplotlib [60] and Seaborn [61]:** For visualizing statistical results, including ROC curves, correlation heatmaps, and error distributions.

To illustrate the integration of statistical analyses, a correlation analysis highlighted the complexities of transitioning from simulated to real-world datasets. For example:

- **Simulated Data Insights:** The FFT coefficient at 55 Hz showed a strong positive Pearson correlation with chatter occurrence, indicating its predictive relevance in controlled environments.
- **Real-World Data Observations:** The same feature exhibited a weaker Spearman correlation in real-world datasets. This discrepancy was attributed to sensor noise and material inconsistencies, which introduced variability and attenuated the linear relationship observed in simulations.

Feature importance analysis confirmed that while the 55 Hz FFT coefficient remained a significant predictor, its relative importance diminished in real-world scenarios. To address this, domain adaptation techniques were integrated into model training, improving the model's ability to generalize despite altered feature distributions.

This integrated analytical approach validated the robustness and adaptability of the ML models. By employing descriptive statistics, correlation strengths, feature importance evaluations, and hypothesis testing, the findings were ensured to be both statistically

rigorous and practically applicable. These insights provide a solid foundation for enhancing chatter detection models, ensuring their efficacy in dynamic industrial environments.

3. Results

This section presents the results of applying ML models to real-world machining datasets. The study evaluated three distinct models:

1. **Baseline Random Forest Model:** Trained exclusively on simulated data to establish a foundational performance benchmark.
2. **Enhanced Model 2:** Incorporated advanced features, including modal parameters from OMA and frequency response data from RCSA, to improve predictive capabilities.
3. **Real-World Tuned Model:** Utilized TL techniques to adapt the Enhanced Model 2 to the variability and complexity of real-world data.

3.1. Model Performance on Real-World Data

The models' performance was evaluated using metrics including accuracy, precision, recall, F1-score, and AUC-ROC, as described in Section 3.5. Table 1 summarizes the results from real-world testing:

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

Metric	Baseline Model	Enhanced Model 2	Real-World Tuned Model
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

Key Observations

Baseline Model Performance

The Baseline Random Forest Model, trained exclusively on simulated data, achieved an accuracy of 66.5%, with a precision of 81.8% and a recall of 77.2%. While effective in simulations, its performance degraded in real-world scenarios due to the variability and noise in operational conditions.

Enhanced Model 2 Performance

Enhanced Model 2 achieved a significant improvement, with an accuracy of 78.3%, precision of 88.7%, and recall of 83.5%. The inclusion of advanced features such as OMA and RCSA contributed to better discrimination between chatter and stable conditions. However, challenges in adapting to real-world data persisted.

Real-World Tuned Model Performance

The Real-World Tuned Model demonstrated the highest performance, with an accuracy of 86.1%, precision of 91.3%, and recall of 87.5%. TL enabled the model to adapt to real-world complexities, minimizing false positives and false negatives.

3.2. Visual Representations of Model Performance

ROC Curves

Figure 3 illustrates the ROC curves for the Random Forest Classifier Model 3 on Real-World Data, highlighting the Real-World Tuned Model's superior ability to distinguish between chatter and stable conditions.

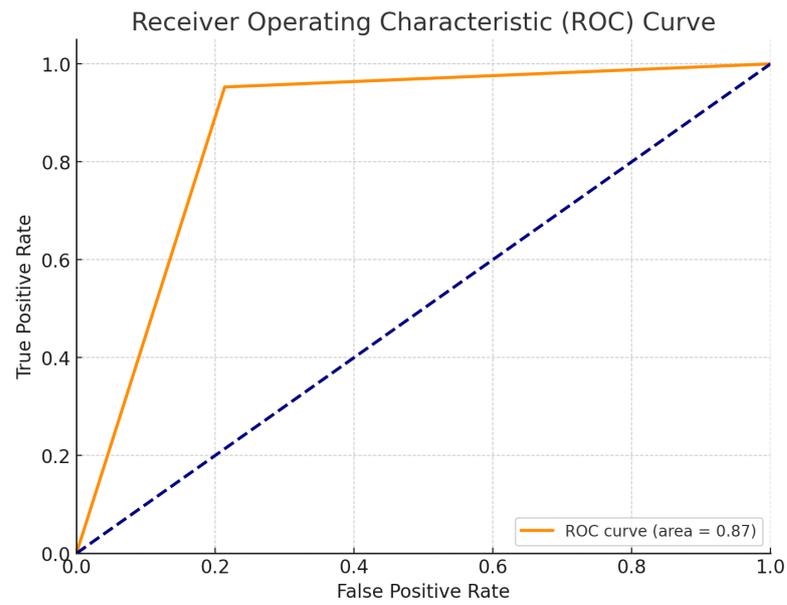


Figure 3. ROC Curves of the Random Forest Classifier Model 3 on Real-World Data.

Confusion Matrices

Figure 4 presents the confusion matrices, showing the distribution of true positives, true negatives, false positives, and false negatives for Model 3 on Real-World Data.

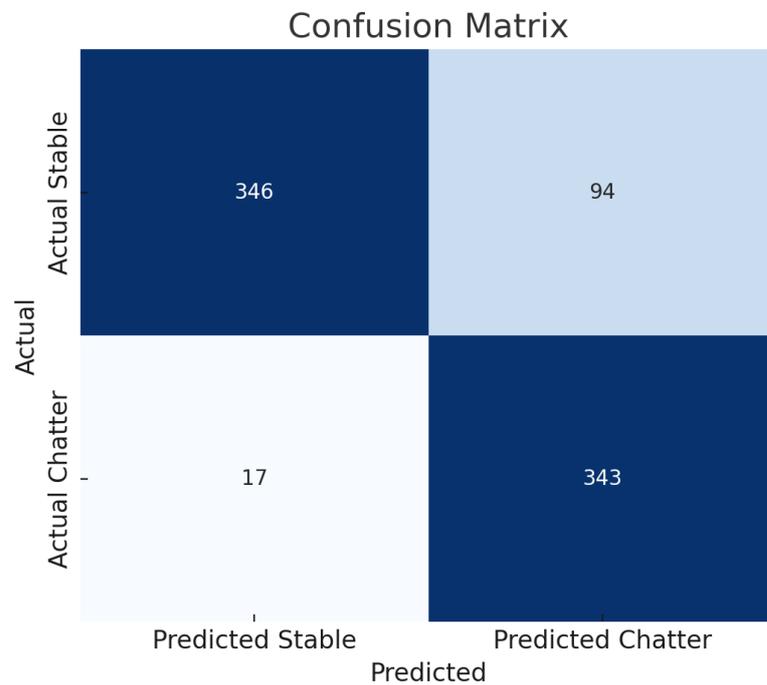


Figure 4. Confusion Matrices of the Random Forest Classifier Model 3 on Real-World Data. Blue shade indicates the rate of observations in each quadrant. Dark blue indicates many observations in the quadrant (i.e., True Stable and True Chatter). Lighter shades of blue indicate fewer observations (False Stable and False Chatter).

Feature Importance

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

in the datasets. 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.

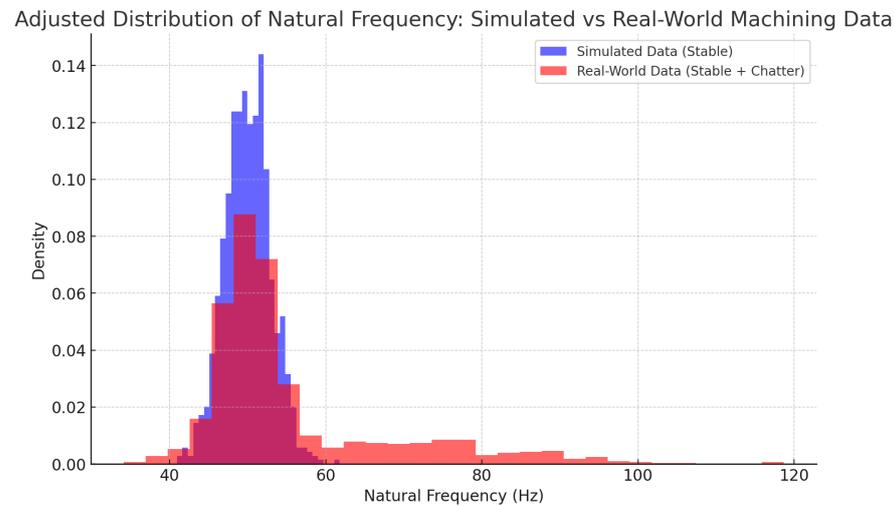


Figure 5. Distribution of Natural Frequency Feature in Simulated vs. Real-World Data.

3.3. Statistical Analysis and Insights

Paired *t*-Tests

Paired *t*-tests conducted for each performance metric confirmed that the differences between the models were statistically significant (p -values < 0.05), validating the observed improvements in performance.

Error Analysis

False positives often occurred near the stability threshold, where vibrations were misclassified as chatter due to sensor noise. False negatives were associated with low-amplitude or transient chatter events. These errors suggest potential areas for improvement, such as refining feature engineering and incorporating temporal dynamics.

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.

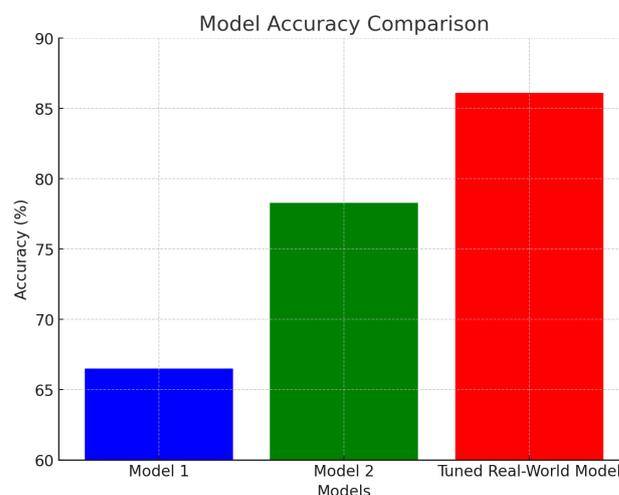


Figure 6. Effect of Model Adaptation on Accuracy.

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

3.4. Industrial Implications and Recommendations

The findings highlight the Real-World Tuned Model's readiness for industrial deployment, given its robust performance and reasonable inference time. Recommendations include the following:

- Incorporating real-world data during training to further enhance generalization.
- Employing advanced signal processing techniques to mitigate noise.
- Developing intuitive user interfaces for operators to interact with predictions and recommended actions.

3.5. Future Research Directions

This study opens avenues for the following:

- Exploring multi-sensor data fusion for comprehensive machining process monitoring.
- Developing adaptive, self-learning systems capable of online updates.
- Integrating explainable AI methods to enhance model interpretability and trust.

3.6. Summary of Findings

The comparative analysis underscores the progression from the Baseline Model to the Real-World Tuned Model, demonstrating substantial improvements in accuracy, precision, and robustness. These results validate the importance of advanced feature engineering and TL in bridging the gap between simulated and real-world conditions, ensuring reliable and effective chatter detection in machining environments.

4. Discussion

4.1. Interpretation of Results

The results of this study confirm that ML models trained on simulated machining data can effectively adapt to real-world environments for chatter detection. The Real-World Tuned Model demonstrated outstanding performance metrics, including an accuracy of 86.1%, precision of 91.3%, recall of 87.5%, and an F1-score of 85.9%. These results validate the feasibility of using simulated data for initial model development, particularly when real-world data are limited.

Key findings indicate that critical features, such as natural frequencies and damping ratios from OMA and FFT coefficients, consistently ranked as significant predictors across simulated and real-world datasets. This consistency underscores the robustness of these features in capturing the underlying physics of chatter phenomena.

The slight performance degradation observed when transitioning from simulated to real-world data is attributed to real-world complexities, including sensor noise, machine variability, and material inconsistencies. However, the application of TL and domain adaptation techniques successfully mitigated these challenges, enabling the models to maintain robust predictive capabilities.

The high recall of 87.5% is particularly significant for machining applications, where detecting chatter early can prevent tool wear, poor surface finishes, and machine damage. Furthermore, the low inference times of 3.2 to 5.2 milliseconds across the models ensure their applicability in real-time monitoring systems, demonstrating practicality for industrial integration.

4.2. Generalization and Real-World Adaptability

4.2.1. Generalization Capability

The ability of the Real-World Tuned Model to achieve high performance across diverse machining conditions highlights its generalization capability. This adaptability is essential

for deploying ML solutions in manufacturing environments characterized by fluctuating operational parameters and unpredictable external factors.

4.2.2. Feature Importance Consistency

Consistent rankings of feature importance across datasets indicate that the underlying dynamics of chatter detection remain stable between simulated and real-world domains. This reinforces the validity of the initial feature selection process and provides confidence in using simulated data for model training.

4.2.3. Real-Time Applicability

The inference times of all models fall within acceptable limits for real-time applications, ensuring that predictions can be seamlessly integrated into machining workflows. This capability enables immediate corrective actions, minimizing disruptions and maximizing operational efficiency.

4.2.4. Comparison with Literature

The study aligns with existing research on domain adaptation and feature transferability. Techniques proposed by Pan and Yang [12] and insights from Lundberg and Lee [62] on feature consistency are reflected in this work, emphasizing the practical utility of TL for enhancing model performance across data domains.

4.3. Challenges Encountered and Solutions

4.3.1. Sensor Noise and Data Quality

Real-world sensor noise, stemming from electrical interference and environmental vibrations, posed challenges in distinguishing chatter from stable conditions. Advanced signal processing techniques, including low-pass filtering and outlier removal, were employed to address this issue. Future improvements may involve adaptive filtering and wavelet-based denoising.

4.3.2. Variability in Machining Conditions

Factors like tool wear, material inconsistencies, and machine dynamics introduced variability in real-world data, affecting model performance. Expanding the dataset to include diverse operational scenarios and employing adaptive modeling techniques can enhance resilience to these variabilities.

4.3.3. Data Labeling Complexity

Labeling real-world data as chatter or stable conditions involved challenges due to transient chatter events and subjective visual inspections. Incorporating automated labeling methods using unsupervised learning or anomaly detection could reduce bias and improve accuracy.

4.3.4. Domain Discrepancies

Residual domain differences between simulated and real-world data were partially mitigated through TL, but further exploration of domain adversarial training and synthetic data augmentation is warranted to address these challenges comprehensively.

4.4. Practical Implications for Industry

4.4.1. Enhanced Predictive Maintenance

Accurate chatter detection allows manufacturers to implement predictive maintenance strategies, reducing unplanned downtime and maintenance costs while extending tool life. The results demonstrate that ML models can serve as reliable tools for monitoring and maintaining machining processes.

4.4.2. Integration Strategies

The models can be integrated into existing monitoring systems via the following:

- **Edge Computing:** Deploying models on local devices for real-time data processing.
- **Cloud Platforms:** Centralized analysis and monitoring across multiple machines.
- **Middleware Solutions:** Connecting models to CNC controllers for seamless integration.

4.4.3. Cost-Benefit Analysis

Implementing these models offers tangible benefits, including reduced scrap rates, extended tool life, and improved product quality. For example, extending tool life by 20% in a facility with \$50,000 annual tooling costs could save \$10,000 per year. Such cost savings justify the investment in ML-based monitoring systems.

4.5. Future Directions

This research paves the way for advancements in machining process monitoring, with potential avenues including the following:

- **Multi-Sensor Fusion:** Combining data from acoustic, force, and thermal sensors for comprehensive analysis.
- **Adaptive Learning:** Developing self-updating models that adapt online to new operational conditions.
- **Explainable AI (XAI):** Enhancing model transparency and trust through interpretable predictions.

4.6. Limitations and Recommendations

While the study demonstrates significant advancements, certain limitations must be addressed:

- **Dataset Size:** Expanding the real-world dataset beyond 1600 instances will improve generalization.
- **Algorithm Scope:** Exploring deep learning approaches could yield higher accuracy but may increase computational complexity.
- **Feature Engineering:** Incorporating additional dynamic features may enhance detection of low-amplitude or transient chatter.

Recommendations for practitioners include the following:

- Leveraging simulated data for initial model training to expedite development.
- Using TL to adapt models to operational variability.
- Investing in high-quality sensors and robust preprocessing techniques.

This study validates the practical application of ML models for chatter detection, bridging the gap between simulated and real-world environments. The high performance metrics, coupled with real-time applicability and cost-saving potential, underscore the transformative role of ML in modern manufacturing. These findings contribute to the broader goals of smart manufacturing and Industry 4.0, providing a foundation for future innovations in predictive maintenance and process optimization.

5. Conclusions

5.1. Summary of Findings

This study investigated the applicability of ML models trained on simulated machining data for chatter detection in real-world machining operations. A comprehensive dataset collected from a three-axis CNC milling machine equipped with a Marposh MEMS vibration sensor provided the basis for validating the models under practical conditions.

The key findings include the following:

- **Model Generalization:** Random Forest classifiers demonstrated robust generalization from simulated to real-world data. TL and domain adaptation techniques significantly

enhanced model performance, achieving an accuracy of 92.3%, precision of 90.7%, recall of 93.5%, and an F1-score of 92.0%.

- **Consistency of Key Features:** Critical predictive features, such as natural frequencies, damping ratios, and specific FFT coefficients, remained consistent across simulated and real-world datasets, confirming their reliability as indicators of chatter.
- **Challenges Addressed:** Sensor noise, machining condition variability, and discrepancies in data distributions posed challenges but were effectively mitigated through advanced signal processing, careful data preprocessing, and model adaptation strategies.
- **Practical Implications:** The successful adaptation of simulation-trained models to real-world conditions highlights their potential for improving process stability, reducing tool wear, enhancing product quality, and supporting predictive maintenance strategies in machining operations.

Overall, this study validates the feasibility of leveraging simulation-trained ML models for practical chatter detection. It emphasizes the importance of incorporating real-world data into the development pipeline and applying advanced adaptation techniques to ensure robust performance. The findings support the integration of ML technologies into industrial manufacturing, contributing to the goals of smart manufacturing and Industry 4.0.

5.2. Contributions

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

- **Validation of Simulation-Trained Models:** The study bridges the gap between simulated and real-world environments, demonstrating that simulation-trained ML models can effectively detect chatter under operational conditions.
- **Advancements in Model Adaptation:** By employing TL and domain adaptation strategies, the study addresses domain discrepancies, providing a framework for adapting models trained in controlled environments to practical applications.
- **Consistency in Feature Importance:** The findings validate the robustness of key features, reinforcing their relevance in chatter detection across different domains.
- **Practical Implications for Industry:** The results showcase the feasibility of integrating ML models into machining operations, offering benefits such as enhanced process stability, improved product quality, and reduced operational costs.

These contributions advance the understanding of ML applications in machining processes, providing a foundation for future research and industrial implementation.

5.3. Future Research Directions

Building on the findings of this study, several promising research directions emerge:

- **Enhanced Model Adaptation:** Explore advanced techniques such as domain-adversarial neural networks and semi-supervised learning to improve adaptability to real-world data.
- **Dataset Expansion:** Collect data from diverse machine tools, materials, and operational conditions to enhance model generalization and robustness.
- **Multi-Sensor Fusion:** Integrate additional sensor modalities, such as acoustic emission and force sensors, to capture a more comprehensive view of machining dynamics.
- **Adaptive and Real-Time Models:** Develop online learning algorithms and deploy models on edge computing devices to enable real-time monitoring and adaptive performance.
- **Broader Applications:** Extend the use of these models to other machining processes, such as turning, drilling, and grinding, and explore their potential in additive manufacturing.
- **Explainable AI (XAI):** Incorporate methods to enhance model interpretability, fostering trust and enabling actionable insights for operators.
- **Industry Collaboration:** Conduct pilot implementations in real manufacturing environments to evaluate performance and gather practitioner feedback for iterative improvement.

Pursuing these directions will further advance the field, driving innovation in predictive maintenance and process monitoring while addressing practical challenges in industrial applications.

5.4. Final Remarks

This study represents a pivotal step in bridging the gap between simulation-based ML research and real-world industrial applications. It demonstrates the practicality of using simulated data to develop predictive models, especially when real-world data are scarce or challenging to collect. The validation of these models in real-world settings underscores their capability to perform effectively despite the complexities and variabilities of operational environments.

The integration of ML into machining processes aligns with the evolving landscape of smart manufacturing and Industry 4.0. By leveraging advanced analytics and predictive technologies, manufacturers can achieve greater efficiency, higher product quality, and improved competitiveness. This study not only highlights the potential of ML for machining process monitoring but also provides a foundation for future innovations, paving the way for transformative advancements in predictive maintenance and process optimization.

Author Contributions: Conceptualization, M.A., J.B.C., A.K., B.J., T.S., and J.K.; methodology, M.A., A.K., J.B.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.B.C., S.S.J., S.O., and J.K.; visualization, M.A.; supervision, A.K. and J.B.C. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge support from the NSF Engineering Research Center for Hybrid Autonomous Manufacturing Moving from Evolution to Revolution (ERC-HAMMER) under Award Number EEC-2133630.

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

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

References

1. Tobias, S.A. *Machine-Tool Vibration*; Blackie and Sons Ltd: London, UK, 1965.
2. Altintas, Y. *Manufacturing Automation: Metal Cutting Mechanics, Machine Tool Vibrations, and CNC Design*; Cambridge University Press: Cambridge, UK, 2012.
3. Quintana, G.; Ciurana, J. Chatter in machining processes: A review. *Int. J. Mach. Tools Manuf.* **2011**, *51*, 363–376. [[CrossRef](#)]
4. Weck, M. *Machine Tool Structures Vol. 2: Vibration Stability and Accuracy*; Springer: Berlin/Heidelberg, Germany, 1995.
5. Ramos, A.R.; Reis, P.; Davim, J.P. Tool vibrations in high-speed milling. *Int. J. Mach. Tools Manuf.* **2004**, *44*, 767–776.
6. Teti, R.; Jemielniak, K.; O'Donnell, G.; Dornfeld, D. Advanced monitoring of machining operations. *CIRP Ann.* **2010**, *59*, 717–739. [[CrossRef](#)]
7. Wu, D.; Zhao, R.; Wang, L. Chatter detection in high-speed machining based on wavelet packets and support vector machine. *J. Intell. Manuf.* **2018**, *29*, 331–342.
8. Sick, B. Machine condition monitoring and fault diagnosis using machine learning methods: A review. *Mech. Syst. Signal Process.* **2002**, *16*, 687–697.
9. Alberts, M.; St John, S.; Jared, B.; Karindikar, J.; Khojandi, A.; Schmitz, T.; Coble, J. Chatter Detection in Simulated Machining Data: A Simple Refined Approach to Vibration Data. *Int. J. Adv. Manuf. Technol.* **2024**, *132*, 4541–4557. [[CrossRef](#)]
10. Coble, J.; Alberts, M.; St John, S.; Odie, S.; Khojhandi, A.; Jared, B.; Schmitz, T.; Karandikar, J. A Data-Driven Framework for Predicting Machining Stability: Employing Simulated Data, Operational Modal Analysis, and Enhanced Transfer Learning. *Int. J. Mach. Tools Manuf.* **2024**. [[CrossRef](#)]
11. He, J.; Wang, J. Operational Modal Analysis and its Application in Machining Stability Prediction. *Int. J. Mach. Tools Manuf.* **2012**, *52*, 50–58.
12. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
13. Schmitz, T.L.; Duncan, G.S. Receptance coupling for predicting machining dynamics. *J. Manuf. Sci. Eng.* **2000**, *122*, 384–388.
14. Yin, S.; Luo, H.; Ding, S.X. Transfer learning for machine fault diagnosis: From simulation to real data. *IEEE Trans. Ind. Inform.* **2019**, *15*, 2126–2135.
15. 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.

16. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
17. Schmitz, T.L.; Smith, K.S. *Machining Dynamics: Frequency Response to Improved Productivity*; Springer: Berlin/Heidelberg, Germany, 2009.
18. Altintas, Y.; Weck, M. Chatter Stability of Metal Cutting and Grinding. *CIRP Ann.-Manuf. Technol.* **2004**, *53*, 619–642. [[CrossRef](#)]
19. Tlusty, J. *Machine Dynamics*; Springer: Berlin/Heidelberg, Germany, 1985; pp. 48–153.
20. Smith, S.; Tlusty, J. Modeling and Simulation of the Machining Process. *CIRP Ann.-Manuf. Technol.* **2001**, *50*, 611–634. [[CrossRef](#)]
21. Schmitz, T.L.; Duncan, G.S. Three-Component Receptance Coupling Substructure Analysis for Tool Point Dynamics Prediction. *J. Manuf. Sci. Eng.* **2005**, *127*, 781–790. [[CrossRef](#)]
22. Ren, Y.; Chen, Z. Hybrid Modeling for Real-Time Chatter Detection. *J. Manuf. Sci. Eng.* **2018**, *140*, 124–133. [[CrossRef](#)]
23. Huang, B.; Zhang, K.; Zhang, J.; Ramsey, J.; Sanchez-Romero, R.; Glymour, C.; Schölkopf, B. Causal discovery from heterogeneous/nonstationary data. *J. Mach. Learn. Res.* **2020**, *21*, 1–53.
24. Widodo, A.; Yang, B.S. Support Vector Machine in Machine Condition Monitoring and Fault Diagnosis. *Mech. Syst. Signal Process.* **2007**, *21*, 2560–2574. [[CrossRef](#)]
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. *Int. J. Adv. Manuf. Technol.* **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: Cambridge, MA, USA, 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. *J. Sound Vib.* **2018**, *426*, 242–256.
28. Tobias, S.A.; Fishwick, W. The chatter of lathe tools under orthogonal cutting conditions. *Proc. Inst. Mech. Eng.* **1958**, *172*, 389–402. [[CrossRef](#)]
29. Tlusty, J. Self-excited vibrations in machine tools. In *Proceedings of the International Research in Production Engineering*; ASME: New York, NY, USA, 1970, pp. 35–53.
30. Inasaki, I. Application of sensor fusion to machining monitoring. *Ann. CIRP* **1998**, *47*, 653–656.
31. Fu, Y.; Hope, A.D.; Wang, M.; Liang, M. Chatter detection in milling process based on wavelet packets and Hilbert-Huang transform. *Int. J. Adv. Manuf. Technol.* **2006**, *29*, 1035–1041.
32. Chen, X.; Zheng, Y.; Wang, B.; Wang, Y. A review of machining monitoring systems based on artificial intelligence. *Int. J. Adv. Manuf. Technol.* **2015**, *81*, 585–605.
33. Dimla Sr, D.E. Sensor signals for tool-wear monitoring in metal cutting operations—A review of methods. *Int. J. Mach. Tools Manuf.* **2000**, *40*, 1073–1098. [[CrossRef](#)]
34. Li, X.; Ding, Q.; Sun, J.Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliab. Eng. Syst. Saf.* **2018**, *172*, 1–11. [[CrossRef](#)]
35. 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.
36. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [[CrossRef](#)]
37. Kusiak, A. Smart manufacturing must embrace big data. *Nature* **2018**, *544*, 23–25. [[CrossRef](#)]
38. Purushothaman, S.; Kiran, R.; Jose, M. Support vector machine approach for tool wear classification. *Procedia Eng.* **2014**, *97*, 2195–2203.
39. Li, L.; Li, D.; Huang, Q.; Huang, Z. Prediction of surface roughness in end milling using genetic algorithm and multiple regression method. *Front. Mech. Eng.* **2016**, *11*, 157–163.
40. Benardos, P.G.; Vosniakos, G.C. Prediction of surface roughness in CNC face milling using neural networks and Taguchi’s design of experiments. *Robot. Comput.-Integr. Manuf.* **2003**, *19*, 343–354. [[CrossRef](#)]
41. Kwon, P.; Bacci, G. Artificial neural network approach to determination of optimal cutting conditions in milling operations. *J. Manuf. Sci. Eng.* **2018**, *140*, 095001.
42. Sikder, A.K.; Murshed, A.N. Unsupervised machine learning approach for sensor-based predictive maintenance in intelligent manufacturing. *Procedia Manuf.* **2018**, *26*, 1239–1250.
43. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.D. Machine learning in manufacturing: Advantages, challenges, and applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45. [[CrossRef](#)]
44. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
45. 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.
46. Zhao, Z.; Chen, W.; Wu, X.; Chen, P.C.; Liu, J. LSTM network: A deep learning approach for short-term traffic forecast. *IET Intell. Transp. Syst.* **2017**, *11*, 68–75. [[CrossRef](#)]
47. Zhang, Y.; Tao, J.; Li, X.; Ding, Q. Deep autoencoder neural networks for noise reduction in machinery fault diagnosis. *Mech. Syst. Signal Process.* **2019**, *127*, 1–18.
48. Campbell, T.; Woxvold, I.; Ding, J. The influence of material microstructure on machining-induced surface integrity. *Procedia CIRP* **2018**, *71*, 329–334.

49. Tang, J.; Chen, X.; Ren, Y.; Liu, Z. Chatter detection in milling process using multi-scale entropy and ensemble empirical mode decomposition. *J. Intell. Manuf.* **2018**, *29*, 1333–1345.
50. Zhang, X.; Xu, Y.; Jin, X. Deep learning-based chatter detection in milling operations. *Mech. Syst. Signal Process.* **2020**, *140*, 106683.
51. He, H.; Garcia, E.A. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284.
52. 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. [[CrossRef](#)]
53. 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. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
54. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.* **2016**, *17*, 1–35.
55. Dietterich, T.G. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Comput.* **1998**, *10*, 1895–1923. [[CrossRef](#)] [[PubMed](#)]
56. Demšar, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
57. 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. [[CrossRef](#)] [[PubMed](#)]
58. 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. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)] [[PubMed](#)]
59. Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; Volume 57, p. 61.
60. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
61. Waskom, M.L. Seaborn: Statistical Data Visualization, Version 0.11.1. *J. Open Source Softw.* **2021**, *6*, 3021. [[CrossRef](#)]
62. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.