Otto von Guericke University
Institute for Intelligent Cooperating Systems
Autonomous Multisensor Systems (AMS)

Using LSTM for Transfer Learning to Predict the Energy Consumption of CNC Machines

December 3, 2024

Ajay Raja Ram Alageshan, Bhavya Baburaj Chovatta Valappil,
Dinesh Babu Gopinath Hemavathi, Sai Nithin Reddy Ankireddy, Yashika Balaji

ams.ovgu.de

– *Using LSTM for Transfer Learning to Predict the Energy Consumption of CNC Machines*–

Ajay Raja Ram Alageshan, Bhavya Baburaj Chovatta Valappil, Dinesh Babu Gopinath Hemavathi, Sai Nithin Reddy Ankireddy, Yashika Balaji

December 3, 2024

Abstract

CNC milling machines consume significant energy, presenting operational costs and efficiency challenges in industrial automation. Accurate spindle current prediction is critical for optimizing energy use and reducing expenses. However, traditional predictive models rely heavily on large labeled datasets, making them resource-intensive. This project proposes a transfer learning approach using the Long Short-Term Memory (LSTM) networks to predict spindle current across materials, methods, or types. The model achieves high accuracy with reduced data and computational requirements by training on spindle current data from one material, type, or method and fine-tuning on another. LSTM networks are well-suited for this time-series task, while transfer learning enables efficient adaptation with minimal new data. Results demonstrate reasonable cross-material accuracy, indicating that this approach is scalable and cost-effective for CNC machine predictive maintenance. This work underscores the benefits of transfer learning for energy efficiency in industrial automation applications.

Contents

1 Introduction

Precision-driven predictive models have become increasingly vital in recent years as industrial process automation has enhanced machine efficiency and operational cost-effectiveness. In this context, continuous monitoring is necessary to maintain optimal performance and energy efficiency of CNC (Computer Numerical Control) milling machines, which are crucial to modern manufacturing. Spindle current, one of the most important operational parameters, is an essential indication of machine performance that facilitates predictive maintenance and better operational understanding. Nevertheless, there are numerous challenges in developing reliable prediction models for these, the most significant of which is the absence of machine-specific, labeled data across a range of operating conditions. The accuracy of traditional machine learning models depends on massive, labeled datasets, necessitating significant time and resources in data collecting.

Despite the potential benefits, traditional machine learning models for spindle current prediction face challenges in real-world applications. These models rely on large volumes of labeled data specific to each material and machining condition, which can be costly and time-consuming to collect. Additionally, machine learning models trained on one set of operational data often struggle to generalize to new materials or different machining environments, limiting their scalability. This data dependency poses a barrier to deploying predictive maintenance solutions across diverse industrial setups.

To overcome these restrictions, transfer learning has become a viable method, enabling knowledge transfer from one domain to another with minimal data requirements. The key advantage of transfer learning is its ability to improve model performance in target domains with minimal information by utilizing knowledge from large, pre-existing data sets. Long Short-Term Memory (LSTM) networks are particularly effective in time-series forecasting tasks due to their ability to capture long-term dependencies in sequential data, making them well-suited for spindle current prediction.

Research in various fields, such as IoT sensor fusion for industrial applications and time-series forecasting in environmental settings demonstrates how transfer learning can be used to overcome data scarcity, lower computational requirements, and speed up model deployment[11, 5]. For instance, Obara and Nakamura (2022)[4] demonstrated the effectiveness of LSTM-based transfer learning in oceanographic forecasting, while Dong et al. (2023)[8] applied a transfer learning framework for predicting the Remaining Useful Life (RUL) of rolling bearings under varying operational conditions. However, the application of transfer learning to CNC spindle current prediction, especially in cross-material scenarios, remains underexplored.

This report introduces a novel transfer learning approach using LSTM networks to predict spindle currents across different materials, specifically steel and aluminum. By training an LSTM model on one material and fine-tuning it on another, we demonstrate that it is possible to achieve high prediction accuracy with reduced data and computational requirements.

The paper's primary contributions are as follows

- **Optimized Predictive Modeling with Transfer Learning:** To overcome the draw backs of traditional machine learning models, we introduce a unique method for predictive maintenance in CNC milling machines that uses transfer learning to improve spindle current prediction accuracy
- **Reduction of Data and Computational Costs:** We preserve model robustness while reducing the need for significant data collecting and computation by transferring learned features from one material's spindle data to another
- **Improved Model Sturdiness in a Variety of Operational Situations:** A versatile usage across many CNC machining settings is made possible by fine-tuning the LSTM model with little data on a variety of materials.

The remainder of this report is structured as follows: Section 2 reviews background knowledge, highlighting the limitations of current approaches. Section 3 presents our methodology, including the LSTM model architecture and transfer learning strategy. Section 4 discusses the experimental setup where we perform different experiments to analyze the performance. In Section 5 we analyze the results. Finally, Sections 6 and 7, conclude with discussions on challenges, limitations, and directions for future research.

2 Background

LSTM networks and transfer learning have been increasingly popular in a variety of domains for time-series forecasting and predictive maintenance. Because typical machine learning models require a lot of data, transfer learning has become a viable option for situations when there is a shortage of labeled data. Its efficacy in industrial applications, especially in situations needing flexibility across various operational conditions, has been demonstrated by recent studies.

By utilizing information from comparable tasks, transfer learning—a key strategy for data-scarce scenarios—has proven successful in a variety of applications. In applications, where labeled data is frequently insufficient, this approach is especially useful. Recent research emphasizes the versatility of transfer learning and its potential to improve prediction accuracy, lessen overfitting hazards, and use less data.

In situations when labeled data is limited, transfer learning has emerged as a flexible method for handling data problems. This approach has proven effective in several fields, including wind power prediction, human activity detection, and financial forecasting. Weber et al(2023)[10] emphasize the potential of transfer learning to generalize across tasks with less labeled data, lowering overfitting and improving model efficiency in thorough systematic mapping analysis. The study highlights the need for more flexible transfer learning models by pointing up enduring issues including negative transfer and source selection issues despite these advantages.

Dong et al. (2023)[8] investigated how transfer learning can be used to forecast rolling bearings' remaining usable life (RUL), where a lack of data can make predictions less accurate. Using Bi-LSTM in conjunction with attention processes, they created a transfer learning model that can adjust to various deterioration circumstances. Their results emphasize the capacity of transfer learning to maintain accuracy throughout various operating states, demonstrating its applicability in predictive maintenance tasks. Moreover, recent research has emphasized transfer learning's adaptability in predictive maintenance for CNC machines. By fine-tuning pre-trained LSTM models on a smaller set of task-specific data, it becomes feasible to predict spindle currents in CNC milling operations with reduced data and computational requirements. This approach aligns with findings from Chen et al. (2021)[11], where the TrAdaBoost model showed the benefit of applying instance-based transfer learning for complex industrial tasks. Such methods enhance model robustness in environments where operational conditions may vary but extensive labeled data is costly or impractical to obtain.

LSTM networks with transfer learning were used by Obara and Nakamura (2022)[4] to anticipate significant wave heights off the coast of Tohoku, Japan. By combining LSTM with Transfer Learning, was able to make better predictions with less data. This outcome demonstrates that transfer learning is not only adaptable enough for oceanographic forecasting, but also that it is helpful for applications involving monitoring, where large labeled datasets are often unavailable.

Despite these advancements, transfer learning in CNC predictive maintenance has received limited exploration, particularly in cross-material adaptation. This project aims to bridge this gap by using transfer learning to improve spindle current prediction across various operational conditions without requiring extensive labeled data for each material. Unlike previous work that primarily focuses on adapting within single material or machinery setups, this research explores a cross-material, cross-condition predictive maintenance solution, potentially extending the application of transfer learning in industrial automation.

Taken together, these studies underline the significant potential of transfer learning for industrial applications, particularly in data-scarce environments with high variability. The combination of transfer learning and LSTM networks offers a robust approach to predictive modeling in CNC milling, where rapid, accurate adaptation to new data is essential. This project contributes to the field by extending transfer learning's applicability across different operational conditions, demonstrating how cross-material, cross-condition generalization can enhance predictive maintenance, reduce operational costs, and optimize energy efficiency in CNC machines.

3 Method

Using LSTM for Transfer Learning to predict the energy consumption of CNC machines under various operational conditions is the aim of this project. In this chapter, we define the methodology behind this aim. In Figure 1, a basic overview of the components and workflow is shown.

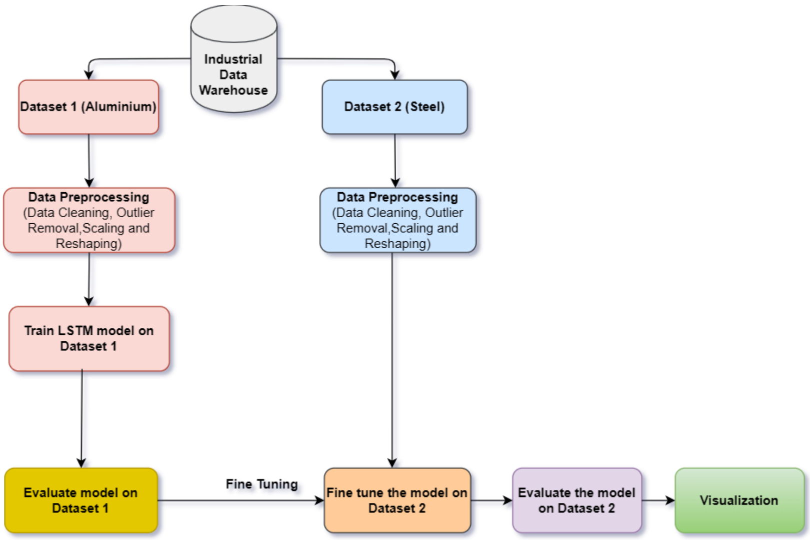

Figure 1: Methodology Overview: LSTM-Based Transfer Learning Framework

3.1 Research Approach

The primary research topic for this project was to identify how precise the energy consumption is in CNC milling machines. We found that transfer learning is a potential method to use existing data from one domain to enhance predictions for another, especially considering the unpredictability of machining settings and the high expense of gathering sizable labeled datasets.

Our strategy was founded on the following ideas:

- Transferring learned patterns from one dataset to another can reduce the demand for large labeled datasets.
- Spindle current measurements are time-series, so handle them with LSTM networks, which are good at understanding temporal dependencies in sequential data.

3.2 Data Collection and Understanding

We have collected the data from a CNC milling machine used to produce different objects (here referred to as Bautiel 1, Bautiel 2) using steel and aluminum under varying operating conditions(with and without air-cut). Various component designs are made of steel and aluminum and have certain geometric and machining needs. The spindle current (curr_sp) is affected by the variable tool paths (positions pos_x, pos_y, pos_z), spindle position (pos_sp), velocities (v_x, v_y, v_z), accelerations (a_x, a_y, a_z), and forces around different axes that are needed for various items during CNC milling. Based on the combination of item type (Bautiel 1 and Bautiel 2), material type (steel and aluminum), and the presence or absence of air cuts, data was gathered across eight scenarios to understand how these aspects affect machine performance thoroughly.

For the application of transfer learning, this organized dataset comprising various object types, materials, and machining circumstances provides a strong basis. The dataset captures a variety of operating situations by incorporating both air-cut and non-air-cut settings for various materials and objects, enabling prediction algorithms to generalize across a range of scenarios. In CNC milling optimization tasks, this improves model performance and adaptability by facilitating the efficient transfer of acquired insights from one machining scenario to another.

Acceleration, position, and velocity recorded along the x, y, and z axes are important characteristics that show the machine's operational states in multiple spatial dimensions. A key measure of machine load and energy consumption, the spindle current is represented by the project's target variable, curr_sp. Machine state variables are measured once and time-stamped in each row.

3.3 Data Preprocessing

The raw dataset included numerous features representing different facets of the CNC machine's functioning. These included simulated values for acceleration, position, and velocity along multiple axes, as well as current values along the x, y, and z axes. Data preprocessing concentrated on outlier control, and data preparation to refine this dataset and maximize spindle current prediction accuracy.

- **Dataset Overview and Initial Cleaning:** The datasets, consisted of nearly 68000-12000 rows, each representing a single timestamped measurement of machine state variables. A preliminary cleaning step was applied, concentrating on removing extraneous features. Simulated values (such as acceleration, velocity, and position data) and specific current directions (x, y, and z) were removed to retain only the variables essential for predicting spindle current. This step resulted in a focused dataset that emphasizes spindle-specific measurements.
- **Data Scaling:** After outlier removal(discussed in Section 4.), all remaining features were standardized using a Standard Scaler[3]. This transformed each feature to have a mean of

0 and a standard deviation of 1, which was necessary to ensure that all variables were on a comparable scale. Standardization helped the model to interpret features consistently, enhancing model convergence during training.

- **Reshaping for Model Input:** To enable the LSTM model to capture temporal dependencies, the dataset was reshaped to accommodate a time step of 10. This transformation created a three-dimensional structure suited for sequential modeling, where each input sample consisted of data from 10 consecutive time points. This temporal structuring allowed the LSTM model to learn time-dependent patterns in spindle current and other key variables, which are crucial for accurately modeling complex, sequential processes.

3.4 Model Selection and Design

The predictive model was constructed using an LSTM network, chosen for its capacity to capture long-term dependencies in sequential data, which is essential in time-series predictions like spindle current. The architecture was designed to balance prediction accuracy and model generalizability, which is particularly important for transfer learning applications. The below section explains the structure of the LSTM network we have used in detail.

1. **First LSTM Layer**: The initial LSTM layer comprises 50 units with the ReLU (Rectified Linear Unit) activation function. ReLU helps accelerate convergence by mitigating the vanishing gradient problem often encountered in deep networks, especially when learning long-term dependencies in sequential data. This layer is set with return_sequences=True, allowing it to output a sequence rather than a single value, which helps maintain temporal dependencies as information passes through subsequent layers.

2. **Dropout Regularization Layer**: Following the first LSTM layer, a dropout layer with a rate of 20% was included to prevent overfitting by randomly deactivating 20% of the neurons during each training iteration. This approach encourages the network to generalize better by learning more distributed representations of the data, critical for transfer learning where the model must adapt to new but related tasks.

3. **Second LSTM Layer**: The second LSTM layer contains 25 units and uses the ReLU activation function. By setting return_sequences=False, this layer outputs the final state of the sequence, distilling the temporal patterns into a fixed-length output. Reducing the number of units in this layer further helps manage model complexity while preserving sufficient capacity to learn relevant features.

4. **Dense Output Layer**: The final layer is a fully connected dense layer with a single neuron for regression, outputting a continuous value representing the predicted spindle current. This configuration is optimal for regression tasks, where a specific numeric prediction is required.

The model is trained with the Adam optimizer, chosen for its ability to adapt learning rates during training, enhancing optimization and model convergence. This architecture was chosen to make use of the model's capacity to learn complex data with its ability to generalize the new data. The dropout layer helps prevent overfitting by randomly deactivating a portion of neurons during training. Mean Squared Error (MSE) is used as the loss function, making it well-suited for regression tasks along with Root Mean Squared Error (RMSE). Additionally, Mean Absolute Error (MAE) serves as a metric to provide an interpretable measure of prediction accuracy.

3.5 Transfer Learning Approach

To improve prediction accuracy across different machining scenarios, a two-stage transfer learning approach was employed. The process involved pretraining the LSTM model on a dataset from one scenario and fine-tuning it on data from another, allowing the model to transfer learned features and adapt with minimal new data. In this project, there are 4 different types of datasets for each material type(steel and aluminum), we have already discussed the dataset earlier. The initial training phase focused on, for example, Bautiel 1 without air cut for steel, allowing the model to learn the underlying patterns. Subsequently, we fine-tuned the model using Bautiel 1 without air cut for aluminum, enabling it to adapt and generalize its knowledge to improve predictions(Similarly for others in all combinations).

1. **Initial Training**: In this phase, the model was trained on the source dataset, using data from one material, to capture general temporal patterns in the spindle current. This initial training provides a foundational model that understands broad temporal features, enhancing generalization capabilities across different materials. The model was trained for up to 100 epochs, with early stopping to prevent overfitting if validation loss did not improve over 10 epochs. Techniques such as early stopping and learning rate reduction were applied to optimize the model and prevent overfitting.
   - **Early stopping**: To prevent overfitting of the data, training was halted if the validation didn't improve for 10 consecutive epochs.
   - **Learning rate Reduction**: Adjusted learning rate dynamically for efficient training.
   - **Model checkpoint**: The best-performing model based on the validation loss results during the training process.
2. **Transfer Learning Phase**: Transfer learning allows adaptation from the source material to the target material with minimal additional data by fine-tuning the pre-trained model. This method reduces the computational cost and data requirements, an essential benefit in data-limited industrial contexts.The fine-tuning process included:
   - **Model Fine-Tuning**: For transfer, the last two layers of the LSTM model were unfrozen, allowing the model to specifically adjust to the target material's characteristics. The layers retained from the initial training phase continue to capture general temporal dependencies, while the fine-tuned layers adapt to the target dataset.

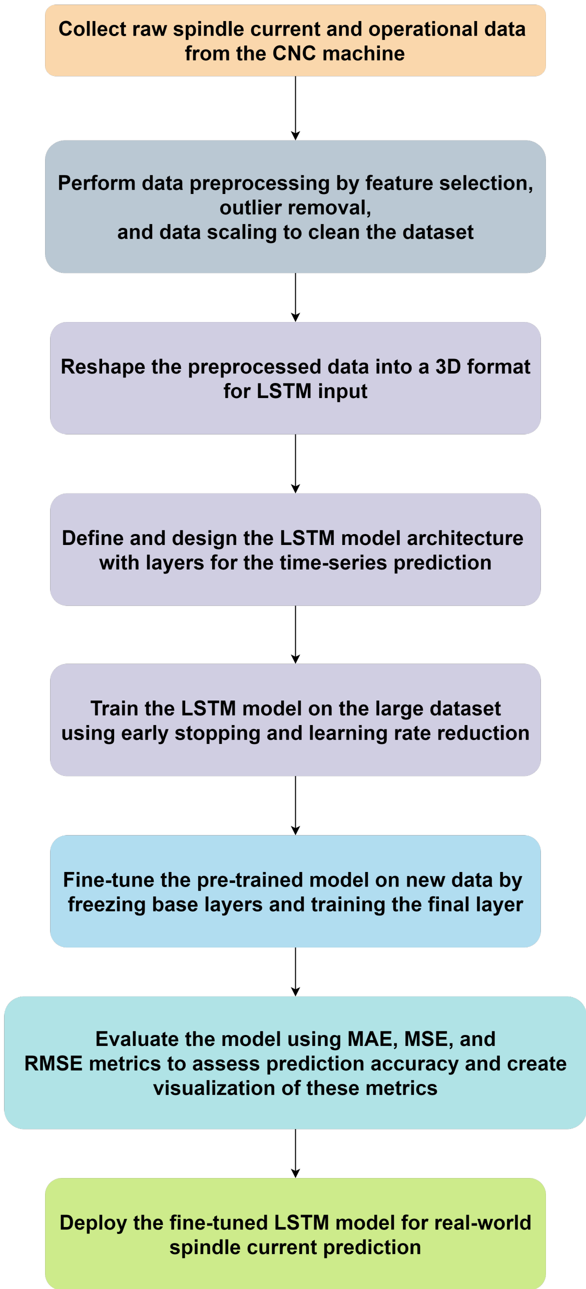

Figure 2: Implementation Workflow

- **Parameter Retention and Training Procedure**: The previously learned weights were retained in the initial layers to preserve temporal dependencies. Training was conducted on the target data for up to 50 epochs with a lower learning rate (0.0001), allowing gradual, stable adaptation to the new material without drastic updates to the weights.

The above-mentioned process allows the model to retain necessary knowledge from one dataset and adapt to the other.Figure 2 provides an overview of the entire implementation workflow, detailing each step from data processing to the deployment of the fine-tuned LSTM model.

4 Experimental Setup

This section outlines various experiments carried out during the implementation. It includes a description of the data handling procedures, identification of best model training parameters, and transfer learning stages to assess the model's performance.

4.1 Outlier Removal

Outliers were identified and removed using the Z-score[3, 1] method to enhance data quality and improve model training. This method assesses how far each data point is from the dataset's mean in standard deviations. The Z-score z_i for a data point x_i can be calculated using equation 1:

$$z_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

where: x_i is the data point, μ is the mean of the dataset, and σ is the standard deviation.

In our analysis, data points with high absolute Z-scores were flagged as potential outliers. Typically, values with $|z| > 3$ are considered outliers. However, to ensure that the data cleaning process did not excessively reduce the dataset, we dynamically adjusted the threshold to maintain less than 1% of the dataset as outliers. This involved testing Z-score thresholds ranging from 1 to 4 and identifying the threshold that preserved over 99% of the data.
Steps Involved:

1. **Calculate Z-scores**: For each data point in the target variable (`curr_sp`), the Z-score was calculated to determine the extent of deviation from the mean.
2. **Select Optimal Threshold**: We tested thresholds incrementally from 1 to 4 (at 0.1 increments) to identify a Z-score that retained over 99% of the data and identified 3.9 as the optimal threshold.
3. **Remove Outliers**: Data points with Z-scores exceeding the optimal threshold were removed from the dataset, leaving a cleaner and more reliable dataset.

The optimal threshold was found to be 3.9, leading to the removal of outliers in the available datasets as shown in Table 1.

4.2 Model Training

4.2.1 Data splitting

To ensure a robust and reliable evaluation of the model, the available dataset was divided into distinct subsets for training and testing. The training set, comprising 70% of the data, was used

Dataset	Steel		aluminum	
	#rows	#outlier	#rows	#outlier
Bautiel 1 with air cut	80231	399	135234	384
Bautiel 2 with air cut	12172	121	19170	187
Bautiel 1 without air cut	68922	308	136352	377
Bautiel 2 without air cut	12410	186	19388	124

Table 1: Table of outliers removed from the dataset

to learn the underlying patterns within the data, allowing the model to adjust its parameters to minimize prediction error. The remaining 30% served as the test set, which was held back from the model during training and used exclusively for final evaluation. This separation helps to assess the model's generalization performance on unseen data, providing a more realistic measure of its effectiveness in real-world applications. Dividing data into these subsets helps prevent overfitting, ensuring that the model does not simply memorize the training data but learns to generalize well. This approach aligns with best practices in machine learning[6].

4.2.2 Hyperparameters

1. **Batch Size**: Set to 32, as a balance between computational efficiency and model stability during training.
2. **Learning Rate**: Initialized at 0.001, and dynamically reduced by half if validation loss plateaued for five epochs, a technique that allows the model to make finer adjustments for better convergence as it approaches a local minimum.
3. **Dropout Rate**: We tested dropout rates of 10%, 20%, 30%, and 40% to determine the best balance between model complexity and generalization for all the datasets. A dropout rate of 20% showed the lowest loss, offering an effective trade-off between reducing overfitting and retaining essential data patterns. At 10%, the dropout rate was too low, allowing too many neurons to remain active, leading to overfitting. In contrast, the higher dropout rates of 30% and 40% might have overly disrupted training without introducing useful new information. Thus, the 20% dropout rate provided the optimal compromise for minimizing validation loss while maintaining model performance as seen in Figure 3 & 4
4. **Optimizer**: The Adam optimizer, known for adaptive learning rate adjustments, was used to improve performance on sequential data like that used in LSTM models.

4.3 Evaluation and Metrics

Several metrics were employed to assess the model's performance in predicting spindle current, each providing unique insights into different aspects of model accuracy and robustness. Here, we define three metrics that were used to evaluate the model's prediction performance for spindle

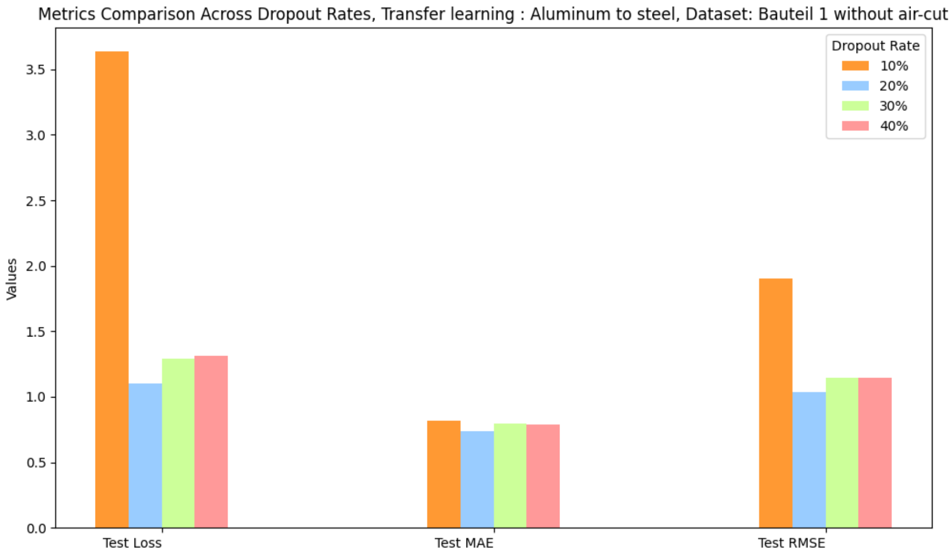

Figure 3: Dropout rate comparison based on testing loss, MAE and RMSE - Transfer Learning Method: aluminum to steel for Bautiel 1 without air cut

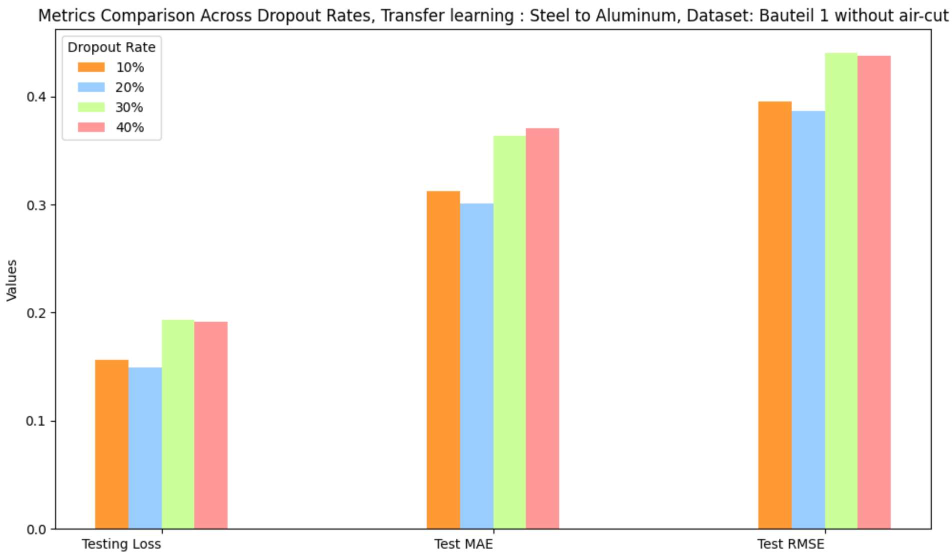

Figure 4: Dropout rate comparison based on testing loss, MAE and RMSE - Transfer Learning Method: steel to aluminum for Bautiel 1 without air cut

current and generalization capacity: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). MSE was chosen because it is highly sensitive to big mistakes, making it effective for spotting significant deviations in predictions that are crucial in CNC operations. RMSE interprets these mistakes in the same unit as the target variable, resulting in an intelligible prediction error scale. In contrast, MAE provides an absolute average error, allowing for an intuitive interpretation of error magnitude without penalizing larger errors disproportionately, as MSE does. These measurements provide a balanced perspective of model performance, covering both the magnitude and impact of errors on CNC milling prediction accuracy. The following section will have a detailed description of the metrics.

4.3.1 Mean Squared Error (MSE)

MSE metric[7, 4] calculates the average squared difference between the predicted spindle current ($\hat{y}_i$) and the actual spindle current (y_i), which focuses on penalizing larger errors. MSE can be calculated using the equation 2.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

where:

- y_i- Actual spindle current for the i-th observation.
- $\hat{y}_i$- Predicted spindle current for the i-th observation.
- n- Total number of data points.

In our model, lower MSE values indicate larger prediction errors are effectively minimized, which is useful for maintaining accuracy in real-time spindle operations.

4.3.2 Mean Absolute Error (MAE)

MAE[7, 4] metric calculates the average of the absolute differences between the predicted and actual spindle currents. The difference from MSE is that it is less sensitive to larger errors by only taking absolute error values, offering a more balanced assessment of the model's overall accuracy. MAE can be calculated using the equation 3.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

4.3.3 Root Mean Squared Error (RMSE)

RMSE[7, 4] metric calculates the square root of MSE, and provides results in the same units as spindle current, making it easier to understand. This metric is more valuable for our project because it gives more weight to larger errors than MAE does, making it suitable for detecting major deviations in predicted spindle currents.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

The lower RMSE value makes the model efficient and its predictions are closely aligned with the actual spindle current measurements, for accurate real-time feedback control in CNC milling operations.

4.3.4 R Squared Error

R Squared Error(R^2)[2], also known as the coefficient of determination, is a metric that measures the proportion of the variance in the target variable that is explained by the model's predictions. An R^2 value of 1 indicates that the model fits well and explains all the variance in the data. The value of 0 explains none of the variance, whereas negative values indicate that the model performs worse than a mean-based model.

$$\mathrm{R}^2 = 1 - \frac{\mathrm{SS}_{\mathrm{res}}}{\mathrm{SS}_{\mathrm{tot}}} \tag{5}$$

where:
Residual sum of squares (error between actual and predicted values)

$$\mathrm{SS}_{\mathrm{res}} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{6}$$

Total sum of squares (variance in the actual values)

$$\mathrm{SS}_{\mathrm{tot}} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{7}$$

- y_i: Actual values.
- $\hat{y}_i$: Predicted values from the model.
- $\bar{y}$: Mean of actual values.

A high value of R^2 reflects that a model is better, with its predictions being closer to the actual measurements of spindle current. This further strengthens the model's reliability for precise real-time feedback control in the operations of CNC milling to optimize the process and its accuracy.

4.4 Transfer Learning Setup

The model's performance was evaluated mainly through conducting experiments on 2 different cases. The following sections will explain the various cases.

4.4.1 Case: Transfer Learning between materials

Here we have trained the model using a dataset where the material used is steel and then tried to transfer the knowledge to the dataset that used material as aluminum and vice versa.

- Setup 1: Transfer learning - steel to aluminum

- Setup 2: Transfer learning -aluminum to steel

Training vs. Validation Loss: Figures 5 and 6 show the training and validation loss across epochs for both Setup 1 and Setup 2 for **bautiel 1 without air-cut dataset**, highlighting the model's convergence patterns. The validation loss consistently follows the training loss, indicating minimal overfitting and stable learning. These plots confirm the effectiveness of early stopping, as training halts when validation loss does not improve for 10 consecutive epochs.

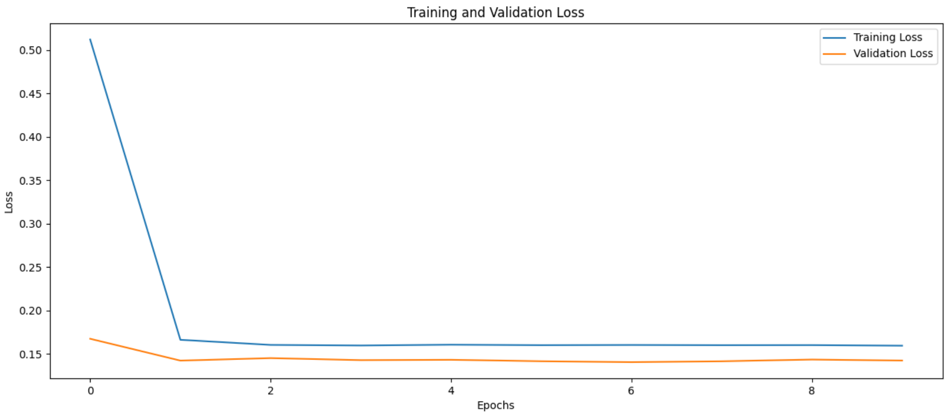

Figure 5: Testing Loss for the transfer learning scenario: Steel to aluminum (bautiel 1 without air-cut)

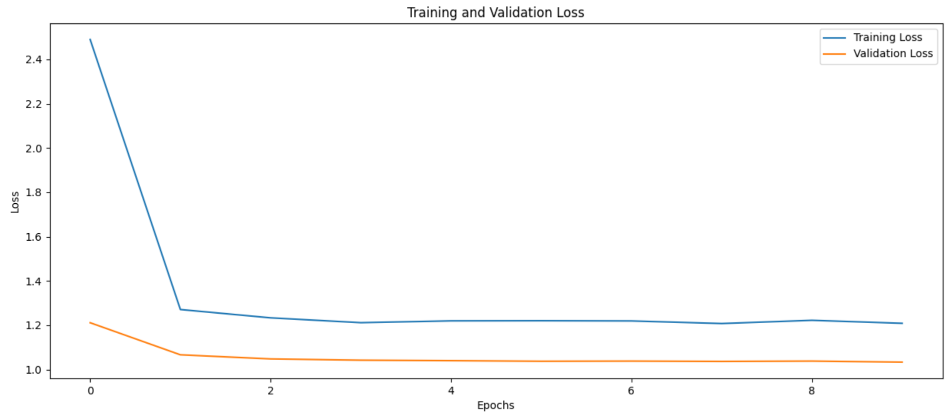

Figure 6: Testing for the transfer learning scenario: aluminum to Steel (bautiel 1 without air-cut)

MAE and RMSE Comparison: Figure 7 & 8 illustrate the RMSE and MAE for setup 1 and Figure 9 & 10 illustrate the RMSE and MAE for setup 2.

Setup 1 displays higher error rates across both metrics, confirming that training on steel provides more generalizable features when tested on aluminum. In both setups, the error rates remain within acceptable ranges, validating the model's adaptability across materials.

R Squared Error: The R^2 metric states that the percentage of variation in a dependent variable being explained by independent variables is a good indicator of performance for the

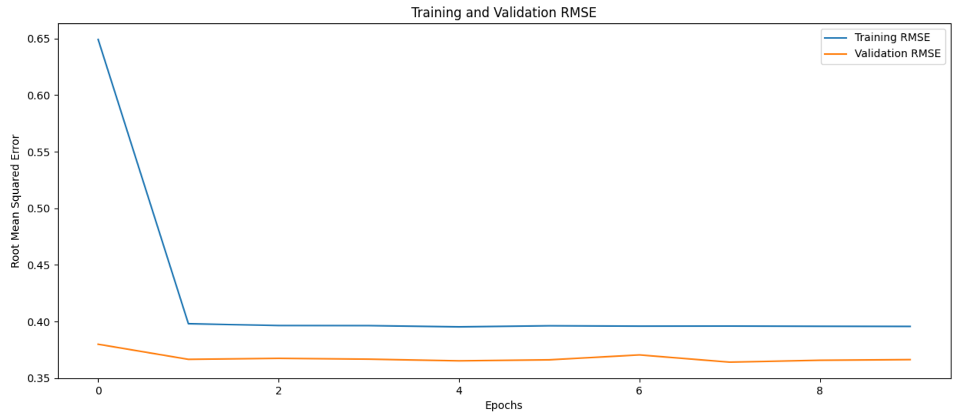

Figure 7: RMSE for the transfer learning scenario: Steel to aluminum (bautiel 1 without air-cut)

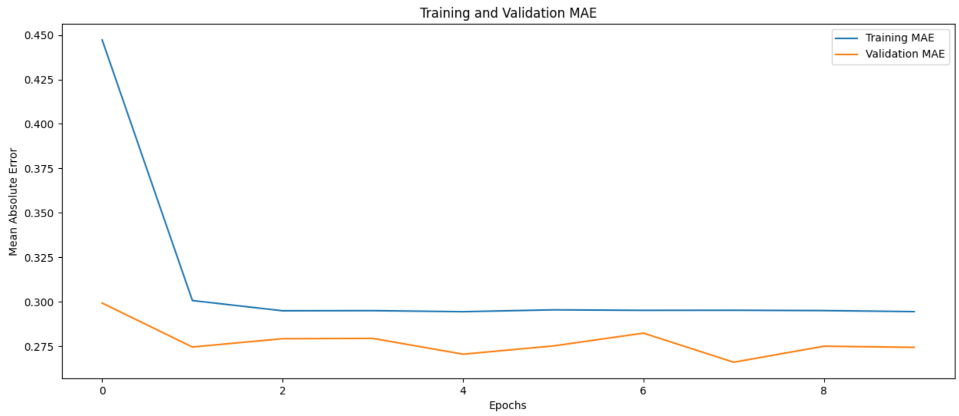

Figure 8: MAE for the transfer learning scenario: Steel to aluminum (bautiel 1 without air-cut)

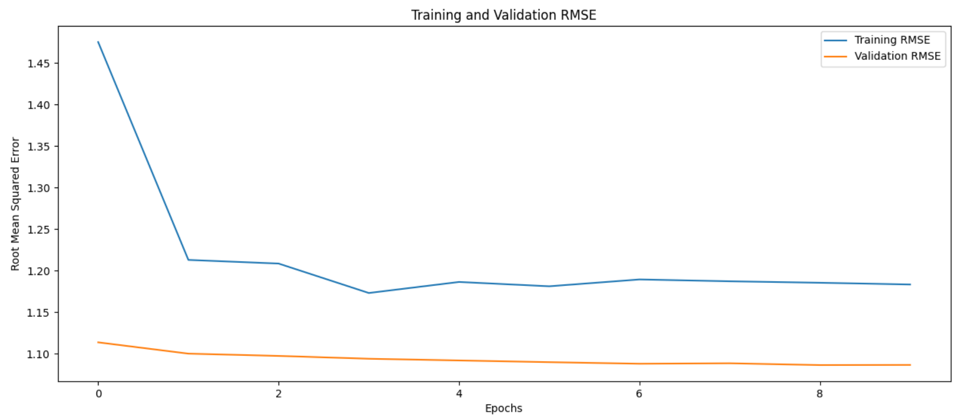

Figure 9: RMSE for the transfer learning scenario: aluminum to Steel (bautiel 1 without air-cut)

model. When the value of R^2, typically 0.95 or 0.93, is high, it shows that the model has a good prediction capability, as it can cover a great extent of the data's variability. The residual variance that is unexplained could be due to the latent factors or the noisy data. Figure 11

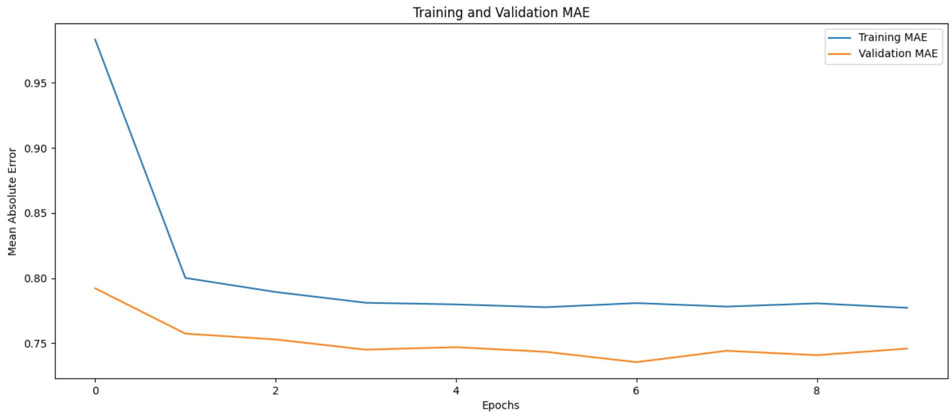

Figure 10: MAE for the transfer learning scenario: aluminum to Steel (bautiel 1 without air-cut)

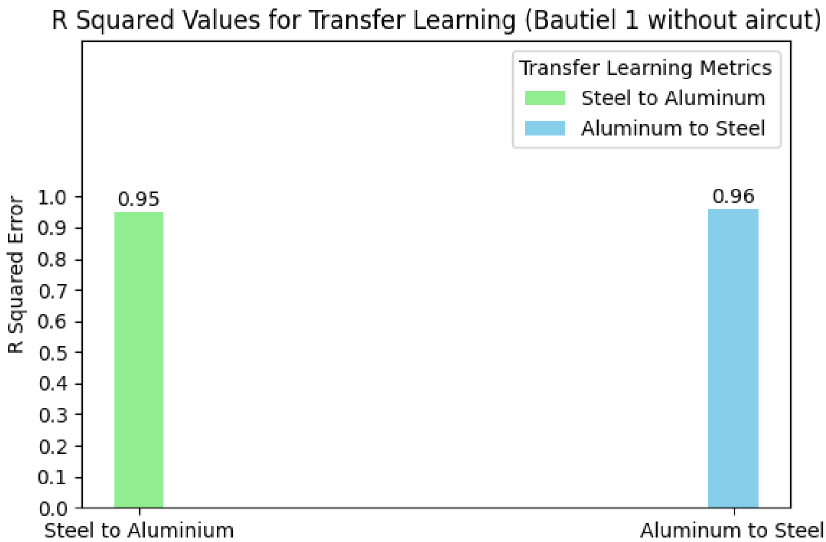

Figure 11: R Squared metrics for bautiel 1 without air-cut

shows the metric value for both setups. The model overfitting assessment should also be considered by taking into account other complementary metrics like MAE, RMSE, and loss function even though having smaller differences in the R^2 values. Besides, the interpretation of R^2 depends on the domain because the thresholds of acceptable values are different for various fields. Hence, the consideration of every relevant metric of the model should be considered when deriving the conclusion of the assessment.

4.4.2 Case: Transfer Learning between bautiels

Here, we have trained the model using a dataset for bautiel 1 and then tried to apply the learned knowledge to the data set for bautiel 2 and vice versa.

- Setup 1: Transfer learning - bautiel 1 to bautiel 2 (material: steel)

- Setup 2: Transfer learning -bautiel 2 to bautiel 1 (material: steel)

MAE and RMSE Comparison: Figure 12 & 13 illustrate the RMSE and MAE for setup 1 and Figure 14 & 15 illustrate the RMSE and MAE for setup 2.

Setup 2 displays higher error rates across both metrics, confirming that training on bautiel 1 provides more generalizable features when tested on bautiel 2. In both setups, the error rates remain within acceptable ranges, validating the model's adaptability across materials.

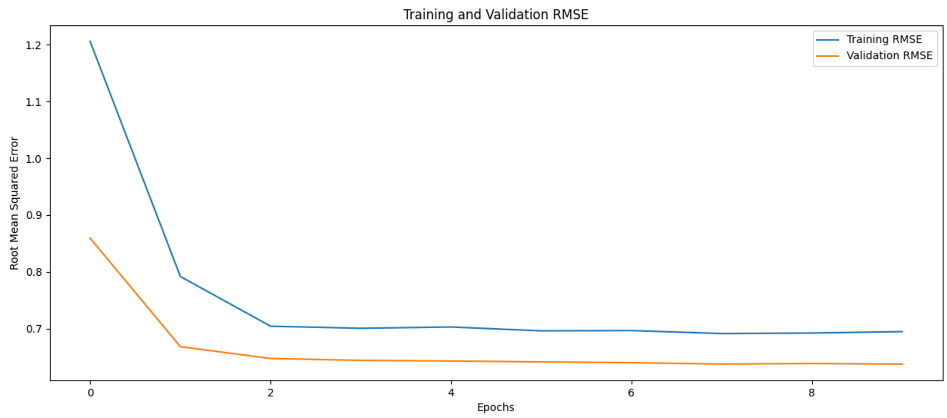

Figure 12: RMSE for the transfer learning scenario: bautiel 1 to bautiel 2 (steel without air-cut)

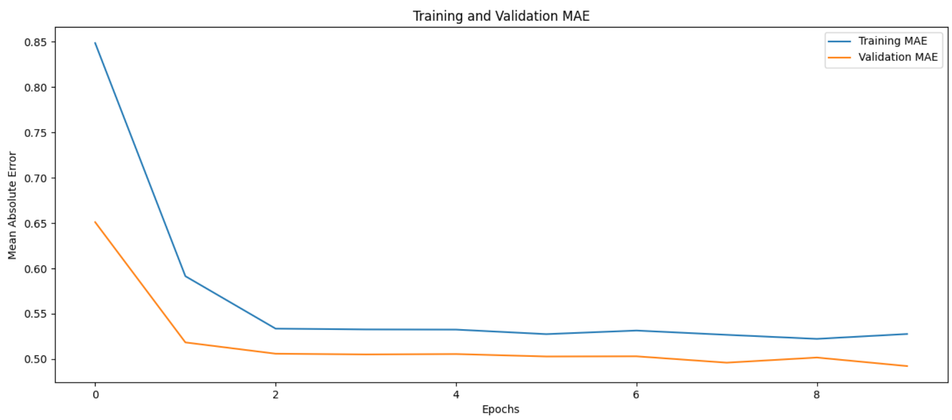

Figure 13: MAE for the transfer learning scenario: bautiel 1 to bautiel 2(steel without air-cut)

R Squared Error: Figure 16 shows the metric value for both setups.

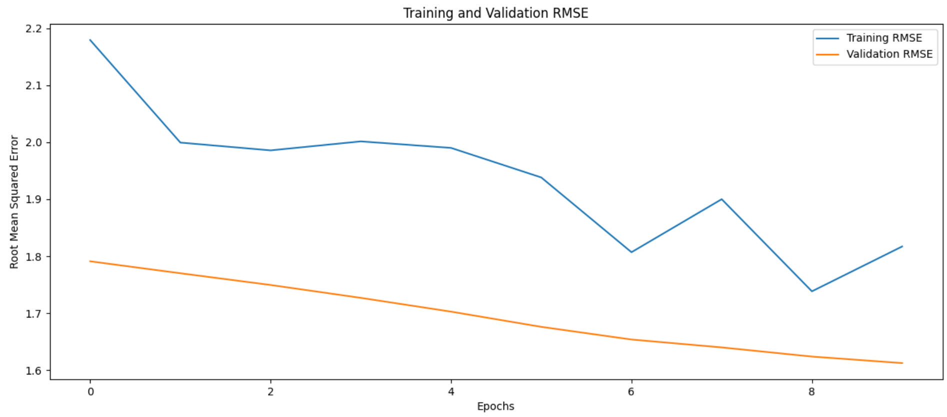

Figure 14: RMSE for the transfer learning scenario: bautiel 2 to bautiel 1(steel without air-cut)

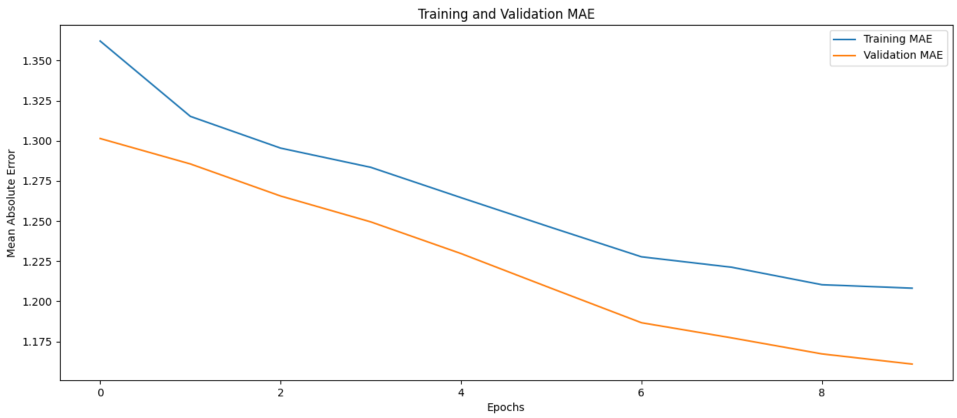

Figure 15: MAE for the transfer learning scenario: bautiel 2 to bautiel 1 (steel without air-cut)

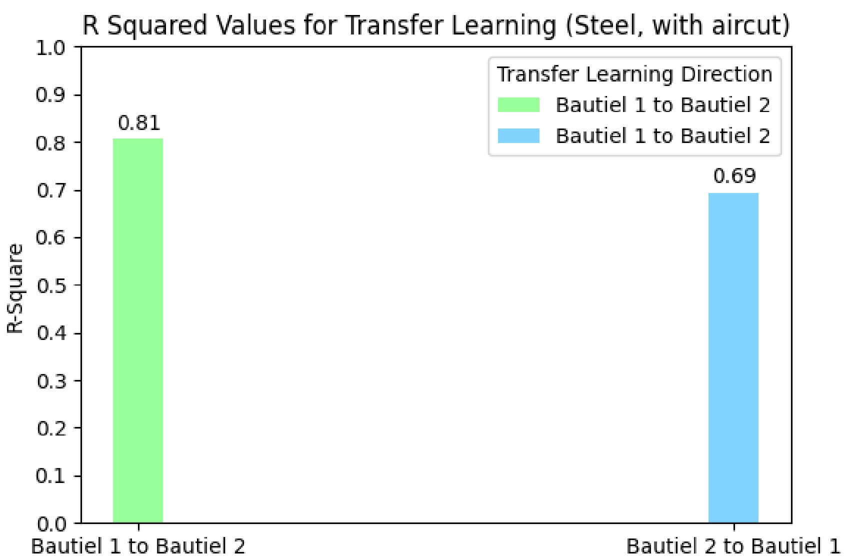

Figure 16: R Squared metrics for steel without air-cut

5 Results and Discussion

5.1 Results

As mentioned earlier in the experimental setup, we had 2 different cases of experiment. Case 1, between the materials, and case 2, which held between the items bautiel 1 and bautiel 2

5.1.1 Transfer Learning between materials - Result Analysis

- Firstly, we trained on steel data and tested on aluminum for Bautiel 1 without air-cut. The results indicated a notable difference in model performance between the two setups. Training on steel and testing on aluminum produced lower error rates. Steel data tends to exhibit more variability and complexity due to its heterogeneous properties, which presents a challenge in achieving high generalizability during the initial stage. But effectively transferred the Knowledge.
- The experiment was conducted in reverse where the model was trained on aluminum first and tested on steel data. Conversely, from the previous result, the model achieved higher stability and slightly higher error rates (measured by MAE and RMSE) when trained on aluminum and tested on steel. This outcome can be attributed to the homogeneous nature of aluminum, which generally produces more consistent and less complex data patterns. These characteristics allow the model to generalize better during the initial training phase, with slightly less effective transfer learning results when tested on steel. The results of both setups can be seen in Figure 17.

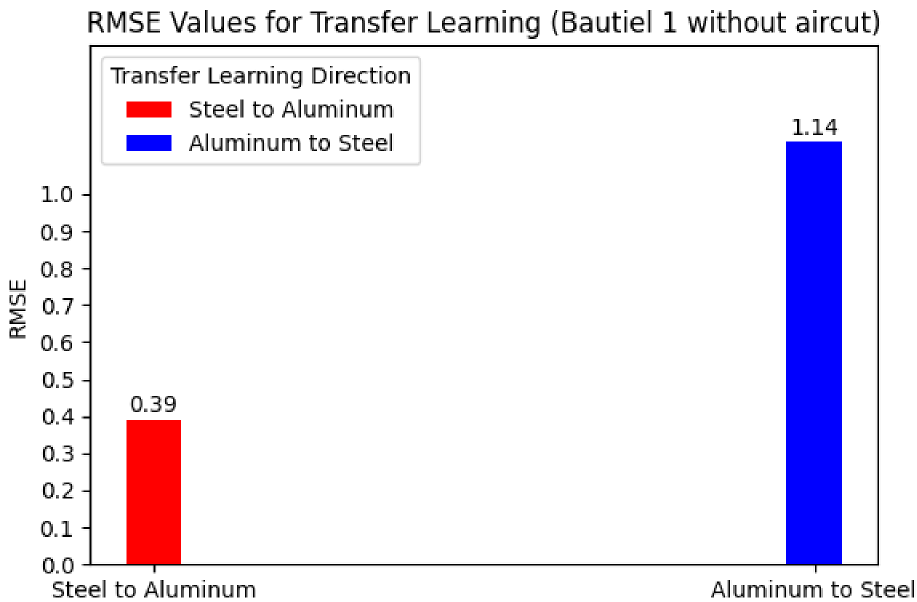

Figure 17: RMSE metrics for Transfer Learning on Bautiel 1 without air-cut

- When tried with Bautiel 1 with air-cut (Steel to aluminum) and (aluminum to Steel) datasets to understand the influence of air-cut in terms of losses during transfer learning, the insights from Figure 18 show that the trend in losses does not vary drastically from the Bautiel 1 without air-cut. The material exhibits its properties and affects the learning pattern and knowledge transfer similarly.

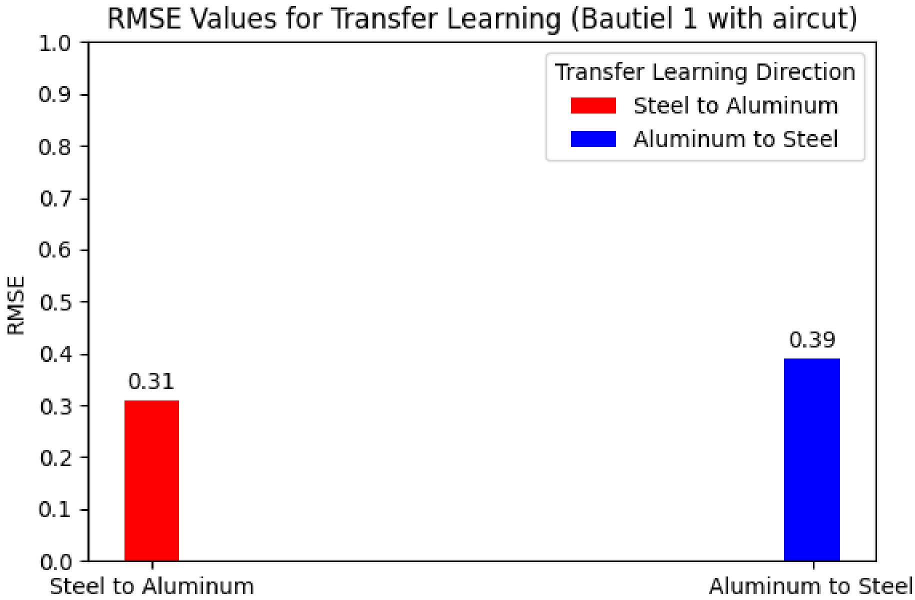

Figure 18: RMSE metrics for Transfer Learning on Bautiel 1 with aircut

- A similar experiment was conducted on Bautiel 2 with and without air-cut (aluminum to Steel) and (Steel to aluminum) which depicts that does not greatly alter loss trends, or variability associated with it when the knowledge is being transferred from one material to another material. See Figure 19 and 20.

5.1.2 Transfer Learning between bautiels - Result Analysis

- we train the model on one bautiel (Bautiel 1 - Steel without air-cut) and transfer it to another bautiel (Bautiel 2 - Steell without air-cut) and vice versa. Since both bautiel's material is steel, the material properties do not vary and its influence in the transfer learning of knowledge is negligible.
- However, when transferring knowledge from Bautiel 1 to Bautiel 2 has significantly less loss in comparison with Bautiel 2 to Bautiel 1 can be seen in Figure 21.
- The same experiment was conducted on Bautiel 1 (aluminum) to Bautiel 2 (aluminum) and vice versa to derive stable insights on this approach. See Figure 22. It shows that the variation in loss while transferring between Bautiel 1 and Bautiel 2 is evident not only in steel but also in aluminum, pointing out the principles of complexity in the data of different Bautiels, generalizability, and training dynamics.

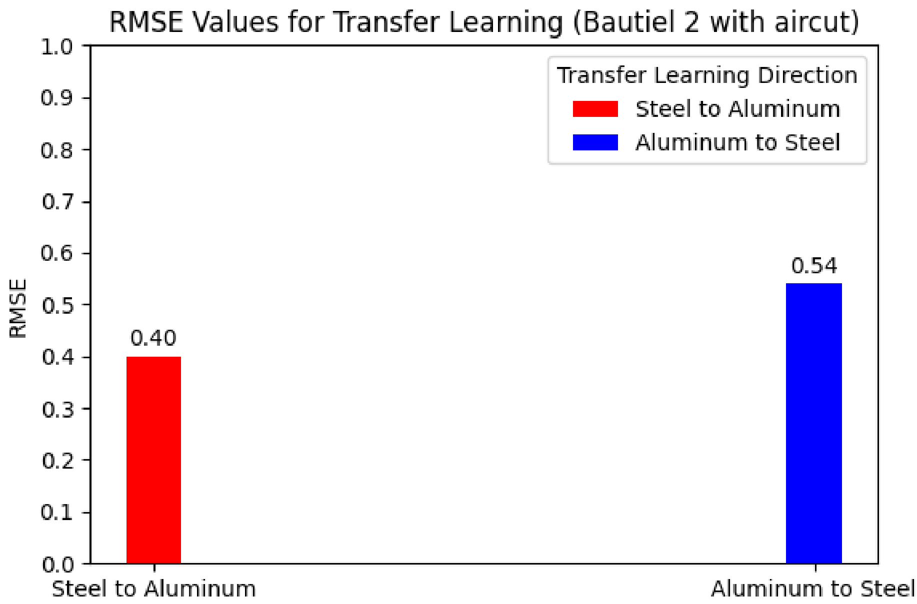

Figure 19: RMSE metrics for Transfer Learning on Bautiel 2 with air-cut

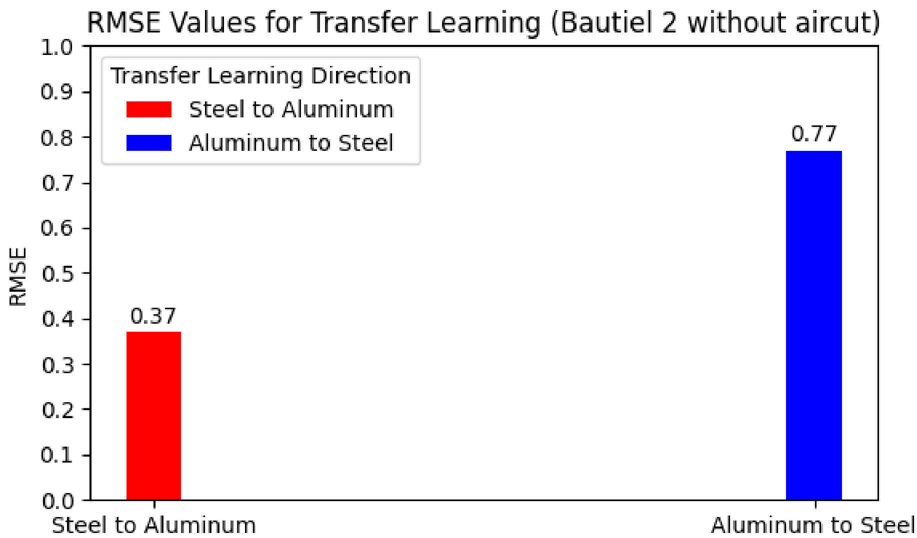

Figure 20: RMSE metrics for Transfer Learning on Bautiel 2 without air-cut

- Bautiel 1 has always exhibited more structured and homogeneous data patterns irrespective of the material. This simplicity enables the model, which is trained on Bautiel 1, to learn features generalizable across the more complicated and variable patterns of Bautiel 2.
- In contrast, Bautiel 2 brings in higher variability and complexities of a noisy dataset. A model trained on Bautiel 2 overfits these specific patterns. It was then much harder for such a model when applied to the simpler pattern of Bautiel 1, yielding higher loss in reverse transfer.

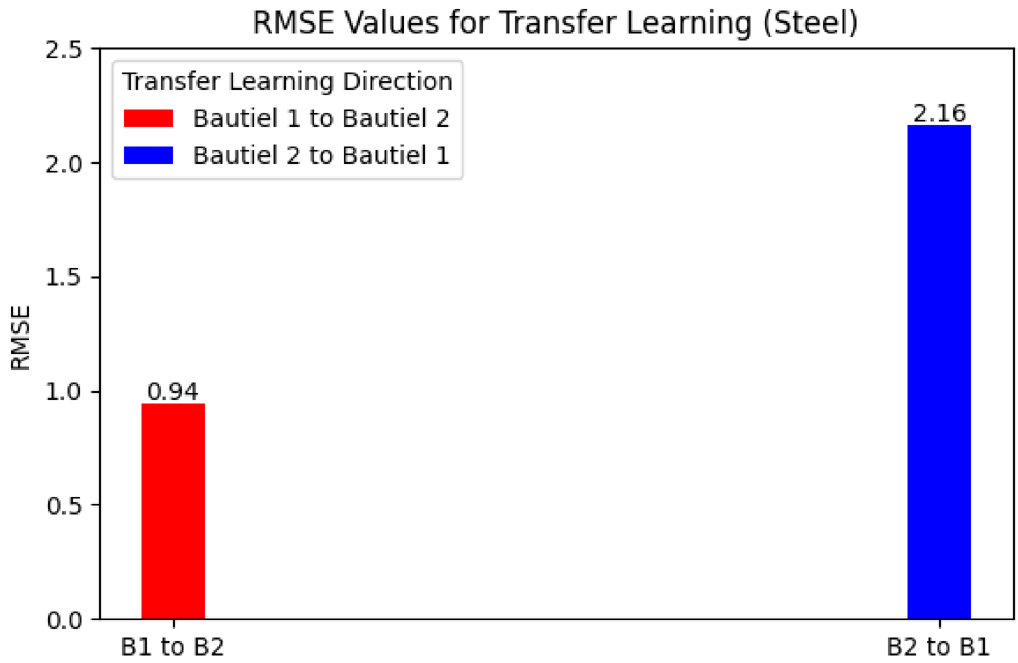

Figure 21: RMSE metrics for Transfer Learning on Steel material

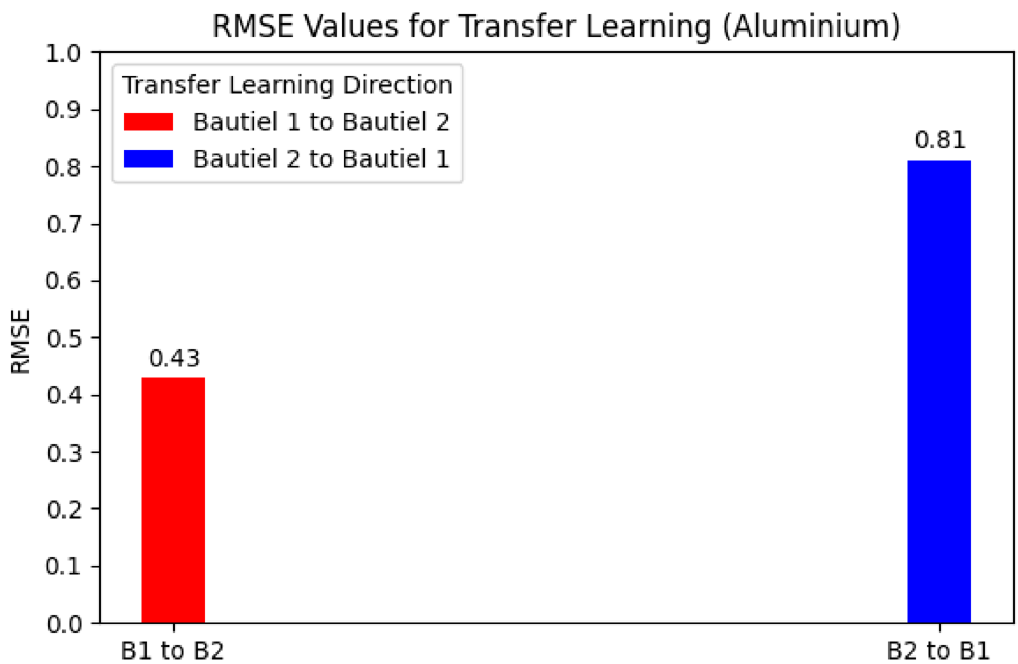

Figure 22: RMSE metrics for Transfer Learning on aluminum material

The detailed overview of metrics for all the experiment setup is shown in Table 2. Across every setup, the model achieved satisfactory prediction accuracy, with lower overall error rates in the transfer learning setup compared to models trained and tested on individual materials. The

Setup	**Loss**	**MAE**	**RMSE**	R^2
Bautiel 1 with Aircut - Al to St	0.155321	0.296639	0.394108	0.940240
Bautiel 1 with Aircut - St to Al	0.169429	0.350679	0.311618	0.973617
Bautiel 1 without Aircut - Al to St	1.298848	0.796005	1.139670	0.967642
Bautiel 1 without Aircut - St to Al	0.149376	0.301067	0.386492	0.958944
Bautiel 2 with Aircut - Al to St	0.288125	0.131530	0.536773	0.960251
Bautiel 2 with Aircut - St to Al	0.156187	0.077871	0.395205	0.879812
Bautiel 2 without Aircut - Al to St	0.592237	0.595616	0.769569	0.881021
Bautiel 2 without Aircut - St to Al	0.139771	0.100917	0.373860	0.963056
Bautiel 1 to Bautiel 2 with Aircut (Steel)	0.557446	0.321063	0.746623	0.694028
Bautiel 2 to Bautiel 1 with Aircut (Steel)	0.467664	0.421058	0.683860	0.872155
Bautiel 1 to Bautiel 2 without Aircut (Steel)	0.877750	0.673217	0.936883	0.817496
Bautiel 2 to Bautiel 1 without Aircut (Steel)	0.662653	1.314537	2.159318	0.920349
Bautiel 1 to Bautiel 2 without Aircut (aluminum)	0.186260	0.180396	0.431579	0.352247
Bautiel 2 to Bautiel 1 without Aircut (aluminum)	0.663001	0.574395	0.814249	0.959479
Bautiel 1 to Bautiel 2 with Aircut (aluminum)	0.516547	0.289911	0.718712	0.542594
Bautiel 2 to Bautiel 1 with Aircut (aluminum)	0.997408	0.487697	0.998703	0.543916

Table 2: Result metrics for different experiment setups

transfer learning approach demonstrated reliable cross-material adaptability, reducing the need for extensive material-specific data.

5.2 Comparison to Other Studies

When compared to existing studies in CNC maintenance prediction, this approach aligns with previous work that emphasizes the importance of capturing temporal dependencies in predictive maintenance. However, it expands upon conventional methods by introducing a cross-material generalization component, where the model learns generalized temporal trends from one material and adapts to others. This differs from typical approaches that train models on single-material datasets, which often lack transferability. The results show that using transfer learning can enhance the versatility and applicability of CNC predictive maintenance models, an improvement over models that are narrowly trained and evaluated within isolated datasets.

When compared to similar works in CNC maintenance prediction, the results from this study align well with the findings of Dong et al. (2023)[8], who demonstrated the effectiveness of transfer learning in environments with high variability and limited labeled data. Like our approach, they found that transfer learning improves prediction accuracy while lowering data requirements. However, unlike Dong et al., who focused on RUL prediction using Bi-LSTM networks, our study emphasizes energy consumption prediction through spindle current analysis, leveraging a simpler LSTM architecture optimized for sequential data in CNC settings.

Additionally, Obara and Nakamura's (2022)[4] research on LSTM-based transfer learning in oceanographic forecasting also corroborates our results. They observed stable cross-environment predictions, similar to our model's stable performance. This consistency underscores LSTM's robustness in handling complex temporal patterns, validating our architecture choice for CNC applications. In summary, this project reinforces the transfer learning approach's advantages observed in existing research, confirming its suitability for cross-material prediction tasks in CNC maintenance and extending its applicability to spindle current prediction.

5.3 Challenges

1. **Data Variability Between Materials**: CNC spindle current data varies significantly depending on material properties. For example, aluminum typically generates more homogeneous data due to its relatively uniform properties, while steel data often exhibits greater variability, influenced by its heterogeneous internal structure. This variability makes it challenging for the model to generalize across materials, as the differences can lead to inconsistencies in the predictive performance, especially in the setup where the model was trained on steel and tested on aluminum.

2. **Feature Correlation**: Certain features, such as spindle position and spindle velocity, may exhibit high correlations with spindle current, which can cause the model to rely heavily on these variables at the expense of other potentially relevant features. This feature correlation can reduce model robustness, especially when transferring from one material to another, where the relationship between spindle current and certain features may shift. Identifying and addressing multicollinearity remains a challenge in this context.

3. **Hyperparameter selection**: Adjusting hyperparameters (e.g., batch size, learning rate) during fine-tuning to prevent overfitting on smaller datasets.

4. **Noise in CNC Data**: CNC data is prone to noise due to sensor inaccuracies, environmental conditions, and varying machine settings. While preprocessing techniques like outlier removal help mitigate these effects, residual noise can still affect model performance, particularly in transfer learning applications where accurate feature representation is essential for cross-material adaptability.

5.4 Limitations

1. **Limits of Transfer Learning in Complex Multi-Material Tasks**: Transfer learning proved effective in adapting the model across two materials, but further complexity, such as incorporating multiple material types or drastically different machining conditions, could challenge the model's adaptability. Transfer learning approaches may require more advanced techniques (e.g., domain adaptation or multi-task learning) to generalize well across a broader range of materials and operational settings.

2. **Parameter Sensitivity and Generalization**: The model's performance was sensitive to hyperparameter tuning, particularly in dropout rates and learning rate adjustments during fine-tuning. Although iterative tuning minimized overfitting, this sensitivity highlights a potential limitation in scaling the model without re-optimizing parameters for new data. This parameter dependency limits the model's immediate applicability to unseen materials or CNC configurations.

3. **Potential Over-Reliance on Temporal Dependencies**: While the LSTM model is designed to capture temporal dependencies, its architecture may overemphasize sequential patterns, potentially overlooking non-temporal relationships in the data. This limitation is particularly relevant when transferring from one material to another, as the model may rely on temporal patterns that differ between materials, affecting its adaptability.

6 Conclusion

This study demonstrates the effectiveness of transfer learning with Long Short-Term Memory (LSTM) networks in predicting spindle current across different materials in CNC milling operations. The results show that a transfer learning approach enables the model to achieve high prediction accuracy with minimal labeled data from the target material, significantly reducing the need for costly and time-intensive data collection. Key findings include:

- The transfer learning model demonstrated superior performance when trained on steel and tested on aluminum, thereby generalizing spindle current predictions from one material to another. In real-world CNC applications where material diversity is prevalent, this adaptability highlights the model's capacity to manage changes in material properties.
- The model helps optimize energy usage, a crucial component of CNC machines' operational efficiency, by precisely forecasting spindle current. Based on these projections, predictive maintenance may be able to decrease energy waste, prevent equipment wear, and lower maintenance expenses.
- Transfer learning achieved dependable predictions even with insufficient target material data by reducing the data dependency commonly associated with machine learning models. In industrial settings where gathering large amounts of labeled data is either expensive or impracticable, this data efficiency is extremely advantageous.

7 Future Work

While the current study provides promising results, there are several areas for potential improvement and further research to enhance the robustness and applicability of the transfer learning model for CNC spindle current prediction.

Other time-series models like Gated Recurrent Units (GRU), Convolutional Neural Networks (CNN) for time series, or hybrid architectures combining CNN and LSTM layers could be investigated in future studies. These models might provide more generalizability and increased accuracy for materials with different attributes. Furthermore, by including attention mechanisms, LSTM models may become easier to read and be able to concentrate on pertinent temporal features.

The model may be optimized by using feature selection methods or dimensionality reduction strategies (such as Principal Component Analysis or L1 regularization), which find the most pertinent information, cut down on noise, and increase prediction accuracy. Future research could potentially look into deep learning-based automated feature extraction to better capture intricate dependencies.

In conclusion, the integration of alternative algorithms, or advanced feature selection methods would contribute to a more robust, adaptable, and interpretable model. These enhancements would allow the transfer learning approach to support a wider range of industrial CNC maintenance applications, contributing to energy efficiency, operational cost reduction, and overall improved predictive maintenance practices.

References

[1] Vaibhav Aggarwal, Vaibhav Gupta, Prayag Singh, Kiran Sharma, and Neetu Sharma. Detection of spatial outlier by using improved z-score test. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 788–790, 2019.

[2] Zeng Chen, Huan Xu, Peng Jiang, Shanen Yu, Guang Lin, Igor Bychkov, Alexey Hmelnov, Gennady Ruzhnikov, Ning Zhu, and Zhen Liu. A transfer learning-based lstm strategy for imputing large-scale consecutive missing data and its application in a water quality prediction system. *Journal of Hydrology*, 602:126573, 2021.

[3] David C. Hoaglin. Volume 16: How to detect and handle outliers. 2013.

[4] Yukie Obara and Ryota Nakamura. Transfer learning of long short-term memory analysis in significant wave height prediction off the coast of western tohoku, japan. *Ocean Engineering*, 2022.

[5] K. Ramesh, B. Rajarao, Hitesh E. Chaudhari, S. Angel Latha Mary, M. Venkatanaresh, and Tarun Kumar Dhiman. Transfer learning approach to reduce similar iot sensor data for industrial applications. *Measurement: Sensors*, 2023.

[6] Sebastian Raschka. Model evaluation, model selection, and algorithm selection in machine learning, 2020.

[7] Farah Shahid, Aneela Zameer, and Muhammad Muneeb. A novel genetic lstm model for wind power forecast. *Energy*, 223:120069, 2021.

[8] Dong Shaojiang, Xiao Jiafeng, Hu Xiaolin, Fang Nengwei, Liu Lanhui, and Yao Jinbao. Deep transfer learning based on bi-lstm and attention for remaining useful life prediction of rolling bearing. *Reliability Engineering & System Safety*, 2022.

[9] Muhammad Tahir, Sufyan Ali, Ayesha Sohail, Ying Zhang, and Xiaohua Jin. Unlocking online insights: Lstm exploration and transfer learning prospects. *Annals of Data Science*, 2024.

[10] Manuel Weber, Maximilian Auch, Christoph Doblander, Peter Mandl, and Hans-Arno Jacobsen. Transfer learning with time series data: A systematic mapping study. *IEEE Access*, 9:165409–165432, 2021.

[11] Chen Zeng, Huan Xu, Jiang Peng, Shan en Yu, Guang Lin, Igor Vyacheslavovich Bychkov, Alexey E. Hmelnov, Gennady M. Ruzhnikov, Ning Zhu, and Zhen Liu. A transfer learning-based lstm strategy for imputing large-scale consecutive missing data and its application in a water quality prediction system. *Journal of Hydrology*, 602:126573, 2021.

[12] Gang Zou, Zhen Yan, Chengying Zhang, and Lei Song. Transfer learning with cnn-lstm model for capacity prediction of lithium-ion batteries under small sample. *Journal of Physics: Conference Series*, 2258, 2022.