2024 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)
July 15-19, 2024. Boston, MA, USA

Transfer Learning in CNC Milling Machines for Chatter Detection using LSTM-AutoEncoders

Eugene Li, Sanjeev Bedi, and William Melek
Mechanical and Mechatronics Engineering
University of Waterloo
Waterloo, Canada
{eugene.li, sbedi, wmelek}@uwaterloo.ca

Abstract—In this work, machine learning is applied to develop a LSTM-AutoEncoder for anomaly detection in three-axis CNC machines. This anomaly detection network is then transferred to another three-axis CNC machine for chatter detection, using significantly less data. This network is then extended to five-axis CNC machines by using the encoder from the three-axis CNC machine to develop an anomaly detection network using transfer and incremental ensemble learning. This approach is compared to a network trained from scratch, with comparable results observed. This approach demonstrates the feasibility of augmenting networks designed for three-axis CNC machines to five-axis CNC machines.

Index Terms—CNC machining, chatter, anomaly detection, machine learning

I. Introduction

In the production world there is an idea known as ”lights out” manufacturing. In this approach, all of the processes are fully automated and limited on-site human presence is required. Lights out manufacturing potentially allows for increased productivity and requires limited human presence on-site as parts are produced [1]. CNC machines are an ideal tool for these types of factories, as they are well equipped machines, capable of producing high performance parts with little human intervention. However, for these machines to properly contribute towards this goal of ”lights out” manufacturing, they need to not only produce quality parts, but also detect and respond to anomalies when they occur. Failing to respond to issues results in unacceptable parts being produced, or damage to the machine or its surroundings.

Given the importance of this issue, there has been significant effort in the literature to find a solution to this problem. Within the machining literature there are many different anomalies caused by different phenomena. Some of the most common and important machining anomalies are chatter, tool breakage or tool/part misalignment [2] [3]. Within the literature there have been different approaches taken to solve these problems, ranging from sensor based options, summarized in [4], to model based methods, summarized in [5], and more recently, model based methods, summarized in [6].

With recent advances in machine learning techniques, data driven approaches have shown remarkable promise for anomaly detection. Recent examples for anomaly detection in CNC machines are [7], [8], [9] and [10]. Although these methods work well, they require large amounts of data to be collected before they can be properly implemented. While this may be feasible in a lab setting, this can be a costly barrier to implementation for those in a production setting.

One solution to this problem is to apply a method known as transfer learning. In transfer learning a network is trained on an initial large data set in a source domain, and then the knowledge is transferred to a target domain [11]. This approach is beneficial as it significantly reduces the amount of training data required for a network by using the network weights from the source network as a starting point for new training. This approach has successfully been applied in machining applications in works such as [12], [13], [14] and [15] where transfer learning was applied for shallow types of networks.

Deep networks are a useful too that allow for more complex system dynamics to be captured. One particularly effective form of deep network used in time-series systems applications is the Long-Short-Term-Memory (LSTM) network. This network has been applied in machining applications for tool wear problems in works such as [16], and chatter detection problems such as [17]. While effective, a disadvantage of these deep networks is their need for large amounts of training data to correctly model a system's behaviour. These deep networks require substantially more data than their shallow counterparts, creating a major obstacle to their implementation.

In our previous work in [18] , the authors were able to develop an LSTM-AutoEncoder for anomaly detection on a source three-axis CNC machine and then use transfer learning to move the network to another three-axis CNC machine. This approach was demonstrated on several machines and was able to achieve 85% accuracy when classifying cutting conditions as either stable or chatter. The accuracy was then improved in [19] by implementing an incremental learning algorithm, which brought the accuracy to 96%, obtaining comparable scores to networks trained from scratch.

Many of the CNC machines implemented in industry are three-axis milling machines. These represent the vast majority of applications and, as such, the vast majority of research. However, in industrial applications, five-axis CNC machines are used when parts need to be produced with highly complex geometry and high precision. In spite of this, five-axis CNC machines are often more expensive than their three-axis counterparts, typically costing three times more than a three-axis machine. As a result of this price disparity, the amount of available training data for five-axis machines is typically substantially less, resulting in fewer open source data sets that can be leveraged for machine learning. This leaves the burden of collecting this data solely on the operators of these machines, adding another barrier to their implementation.

In this paper, we summarize our previous work on transfer learning for chatter detection using LSTM-AutoEncoders for three-axis CNC machines, and incremental ensemble learning for improving the accuracy of these systems. We then demonstrate how these approaches can be applied to five-axis CNC machines to transfer data from a three-axis CNC machine, and then achieve comparable results to networks trained from scratch on only five-axis CNC machine data.

This paper aims to provide the reader with the necessary background information to understand chatter dynamics of milling machines, LSTM-AutoEncoder networks, and the fundamentals of transfer and ensemble learning. The organization of this paper is as follows: In section 2, the necessary background information on chatter dynamics, LSTM Auto-Encoders and transfer learning will be outlined. In section 3, the methodology implemented will be discussed. In section 4, the experimental design and results will be presented. Section 5 will offer a discussion on the findings and section 6 will provide a conclusion.

II. Background

A. Chatter Detection in CNC Milling Machines

In precision machining applications the presence of chatter is detrimental, as it means that a particular part cannot be made within tolerance and must be scrapped. Chatter is a common problem that occurs due to self excited vibrations between the tool and the work piece. These vibrations grow until the tool jumps out of the cutting zone, or breaks due to exponentially growing dynamic displacements [20].

Established in works such as [21] and [22], traditional chatter analysis is done by establishing a stability lobe diagram (SLD). In a SLD, the spindle speed and cutting depth conditions are determined based on machine, material and cutting specific parameters. These conditions are then used to create a diagram with stable and unstable cutting regions, where unstable regions are where chatter occurs. These conditions are determined by completing cutting experiments, and then analyzing the frequency response to look for changes in amplitude, indicating a change in cutting conditions [23].

While these methods have proven effective, and are easy for humans to interpret, the SLD need to be re-obtained for every material and cutting combination of interest, and can change overtime as machine dynamics shift. In addition, the SLD is difficult to integrate into a machine controller, making automatic responses more challenging to implement.

B. LSTM-AutoEncoder

First developed by Hochreiter and Schmidhuber in [24], and further refined by [25] , the Long-Short-Term-Memory network is a type of recurrent neural network that, as its name suggests, is capable of storing both long term and short term memory of a system. This is accomplished by developing a ”memory” of the state of a system over time by using gates that control the flow of information in and out of a cell within the network.

In a LSTM network, each unit is composed of a cell which has an input gate, an output gate and a forget gate. Each cell retains the memory for a time interval, with the gates controlling the flow of information in and out of the cell. As outlined in [24], the LSTM architecture can be defined with matrices $W_q \in \mathbb{R}^{h\times d}$ and $U_q \in \mathbb{R}^{h\times h}$ and vector $b \in \mathbb{R}^h$, which contain, respectively, the weights of the input and recurrent connections, where the subscript $_q$ can either be the input gate i, output gate o, the forget gate f, or the memory cell c. These weights and bias vector parameters are learned during training, where the superscripts d and h refer to the number of input features and number of hidden units, respectively.

$$
\begin{aligned}
g_t = \tilde{C}_t &= tanh(W^g x_t + U^g h_{t-1} + b^g) \\
f_t &= \sigma(W^f x_t + U^f h_{t-1} + b^f) \\
i_t &= \sigma(W^i x_t + U^i h_{t-1} + b^i) \\
o_t &= \sigma(W^o x_t + U^o h_{t-1} + b^o)
\end{aligned}
$$

where the operator $\odot$ denotes the Hadamard product, otherwise known as element-wise multiplication, σ represents a sigmoid function and the subscript t indexes the time step. The variables used are summarised in Table. I

TABLE I
Variables used in LSTM

Parameter Name	Notation
x_t	Input vector to the LSTM unit
$g_t/\tilde{C}_t$	Cell input activation vector
C_t	Cell state vector
f_t	Forget gate’s activation vector
i_t	Input vector to the LSTM unit
o_t	Output gate’s activation vector

In an LSTM AutoEncoder, we develop a network with several layers for encoding, an embedding layer and a decoding layer. This network topology can be seen in Fig.1. This topology is used in anomaly detection by having the network trained to reconstruct the original signal, and seen in works such as [16], [26], [27] and [28].

In this configuration, if the network is provided with proper signals, it will correctly reconstruct the original

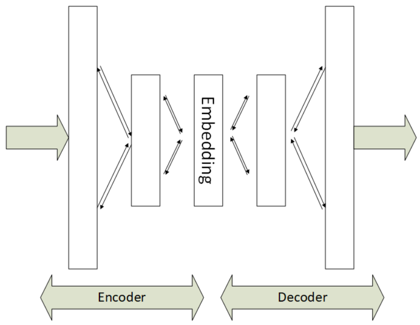

Fig. 1. Block diagram of LSTM AutoEncoder Architecture indicating encoder and decoder portions

signal with little error. However, if signals from an anomaly are provided, the network will reconstruct the signal poorly, resulting in a large reconstruction loss. This reconstruction loss can be used to determine if an anomaly has occurred, if the value crosses a pre-defined threshold.

The anomaly threshold is determined by plotting the loss distribution histogram from a given sample of stable conditions. From the histogram, we can determine the mean and three-sigma distance using traditional statistical methods. This three-sigma distance allows us to determine the threshold for reconstruction error, since reconstruction losses beyond three-sigma, most likely represent outliers and do not belong to the proper set, and thus represent anomalies.

C. Transfer Learning

As outlined in [11], transfer learning is a method where an existing model, trained on a source data set, is adapted to predict examples from a different target data set. This is favorable, as the target data set can be much smaller than the source data set, and the target network can be trained faster with less data.

Following the conventions defined by [29], we can mathematically define transfer learning as follows. Given a domain $\mathcal{D} = \{x, P(X)\}$, containing a feature space x and the probability distribution $P(X)$, where $X = \{x_1, ..., x_n\}$. A task can be presented by $\mathcal{T} = \{\mathcal{Y}, f(x)\}$, where y represents a label space and $f(x)$ a target function. These labels and target functions are learned from the training data consisting of pairs $\{x_i, y_i\}$, where $x_i \in X$ and $y_i \in \mathcal{Y}$.

For a given learning task $\mathcal{T}_S$, from source domain $\mathcal{D}_S$, and a target domain $\mathcal{D}_T$ and learning task $\mathcal{T}_T$, where $\mathcal{D}_S \neq \mathcal{D}_T$, or $\mathcal{T}_S \neq \mathcal{T}_T$, transfer learning aims to improve the predictive function $f_T(\cdot)$ in $\mathcal{D}_T$ using the knowledge in $\mathcal{D}_S$ and $\mathcal{T}_S$. This is accomplished by using the latent knowledge from $\mathcal{D}_S$ and $\mathcal{T}_S$. In most cases the size of $\mathcal{D}_S$ is much larger than $\mathcal{D}_T$ [29].

To ensure that the information from the source network can be properly transferred to the destination network, we evaluate the Maximum Mean Discrepancy (MMD) metric. The MMD metric was first described in [30] and has been used in works such as [31] and [32]. The MMD is a kernel based statistical test used to determine if two data sets represent the same distribution. In the context of transfer learning, if both the source and target training data have a low MMD score, then we can conclude that they both represent similar domains and transfer learning will be successful. For CNC milling machines, a low MMD score also indicates that both the source and target networks will respond to a given input in a similar manner, due to structural similarities between the two machines [33].

The MMD metric can be mathematically represented as follows. Given two distributions X and Y, we assume there is a feature map $\phi(*)$ that maps the distributions to a feature space F. The MMD is calculated as

$$MMD^2(X,Y) = \|\mu_X - \mu_Y\|_F^2 \tag{1}$$

where μ represents the mean of the distribution in the feature space.

This expression has a straightforward description, but is often difficult to implement, especially in large dimension datasets. Thus, we use an empirical estimation by representing the system with a kernel $k(x_i, x_j)$. For this implementation, we have chosen a linear kernel, as the data does not appear to have any specific distribution.

$$MMD^2(X,Y) = \frac{1}{m(m-1)} \sum_i \sum_{j \neq i} k(x_i, x_j) \tag{2}$$

$$-2\frac{1}{m.m} \sum_i \sum_j k(x_i, y_j) + \tag{3}$$

$$\frac{1}{m(m-1)} \sum_i \sum_{j \neq i} k(y_i, y_j) \tag{4}$$

III. Methodology

A. Transfer learning of Anomaly Detection for Three-Axis CNC Milling Machine

To measure the vibrations in a CNC milling machine, and thus detect chatter, the approaches outlined in [34] were followed. This approach identified an accelerometer as the optimal choice for measuring the system vibrations due to its ease of use, accuracy and strong correlation to the physical phenomena.

With this knowledge, cutting experiments were conducted and end mill vibrations were measured by mounting accelerometers on the CNC spindle. These cutting experiments are completed with stable and chatter conditions observed. Then, using the stable cutting data, and the procedure described above, an LSTM-AutoEncoder is trained. This is accomplished using mini-batch stochastic gradient descent following the methods detailed in [18]. Once the anomaly detection system for the source CNC machine is designed, the encoding layers are frozen, and new cutting experiments on a target machine can be completed. The decoding layers of the AutoEncoder network are then re-trained using the cutting data from the target machine. This newly trained network is substantially more capable of detecting anomalies such as chatter, as opposed to directly transporting the network without retraining. The experimental results and details can be found in [18].

B. Incremental Learning of Anomaly Detection for Three-Axis CNC Milling Machines

Although transfer learning is a powerful technique that allows networks to be trained quickly with less data, it presents a limitation by only training the target network based on a snapshot in time. This can result in reduced accuracy, and potentially impacts the generalization of the network depending on the data set used. One solution to this problem is to implement incremental learning. Incremental learning, as the name implies, allows the network to continually grow and learn as more data is presented [35].

Incremental learning can also be applied to improve the anomaly detection accuracy of system trained with transfer learning. This can be accomplished by implementing a boosting approach, outlined in the popular Learn++ algorithm, detailed in [36] and [37]. In this approach, weak learners are added to a network as time progresses, allowing the system to continually learn. This approach has been extended to LSTM networks through the work in [38].

This approach was implemented for anomaly detection in CNC machines in our work in [19]. In this work, we showed that by implementing ensemble incremental learning, we could improve the accuracy of a network trained via transfer learning by 25%. This was accomplished by adding and replacing weak learners to a pool of learners as more data become available. Each of these learners followed the AutoEncoder format that was used to design the rest of the network, with the output reconstruction error being averaged through a dense layer.

C. Transfer learning of Anomaly Detection for Five-Axis CNC Milling Machine

As outlined in the introduction, five-axis machines typically cost significantly more than their three-axis counterparts. As a result of this increased cost, there are typically fewer five-axis machines implemented in production, and less available data as a consequence. This makes training machine learning models challenging as an operator may not have an abundance of old data to draw from for training a network.

Of the many different types of configurations of five-axis CNC milling machines, one particular configuration provides a unique opportunity to leverage transfer learning to its advantage. For swivel head and rotary table style CNC machines, such as the Hurco VMX42SRTi, there is significant overlap in the geometry between a three-axis mill and the five-axis mill. For example, the Hurco VMX42SRTi and the Hurco VMX42Di are constructed with the same column casting, base case and y-saddle. This resemblance means that the two machines will respond to an input signal with greater similarity [39].

To leverage the likeness between the Hurco VMX42Di and the VMX42SRTi, we implement a transfer learning and incremental ensemble learning approach. In this approach, we use the encoder from the three-axis network to capture the known dynamics from the spindle. We then implement an ensemble learner to capture the new degree of freedom from the spindle and the table. The transferred portion and the ensemble learner is then fed into a decoder that is trained on data from the five-axis machine. This topology can be seen in Fig. 2.

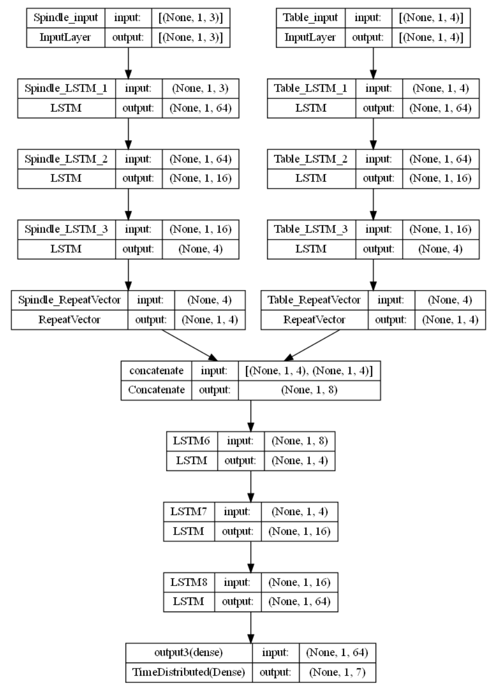

Fig. 2. Network Configuration for Five-Axis Anomaly Detection indicating layer sizes and types for stacked encoder and decoder

The accuracy and robustness of this network can then be improved by continually adding incremental learners to the system. This allows the network to compensate for limited initial training data, and improve its performance over time. A simplified representation of this network can be seen in Fig. 3 below. In this approach, we implement several weak learners, as outlined in the incremental learning portion, and average the output of them with a single dense layer.

IV. Experimental Design and Results

To experimentally develop and validate this approach, a Hurco VMX42SRTi five-axis CNC milling machine was instrumented with accelerometers to capture the spindle and table dynamics. To capture this data, a set of Erbessd EPH-V11E wireless tri-axial vibration sensor were modified to capture the necessary cutting information at a sampling rate of 10kHz, in a dynamic range of

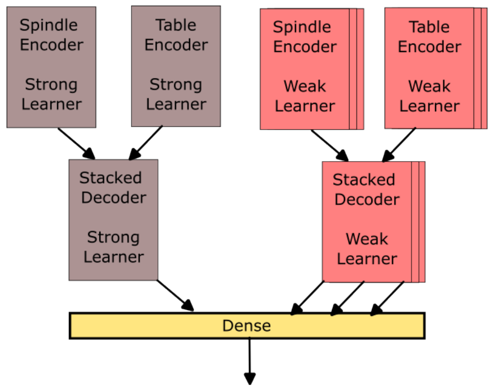

Fig. 3. Block diagram indicating network configuration for five-axis anomaly detection with ensemble incremental learning

± 8g with a sensitivity of 100 mv/G. The information was sent via the Erbessd Phantom gateway to a desktop computer, where it was recorded for post processing. The Erbessd sensors were also supplemented with additional single axis accelerometers, produced by PCB Piezotronics (352A21). This sensor has a ±15% sensitivity of 10 mV/g (1.0 mV/(m/s^2), measurement range of ±500 g pk (±4900 m/s^2 pk), broadband resolution of 0.004 g rms (0.04 m/s^2 rms), and ±5% frequency range of 1.0 to 10 kHz. This experimental setup can be seen in Fig. 4.

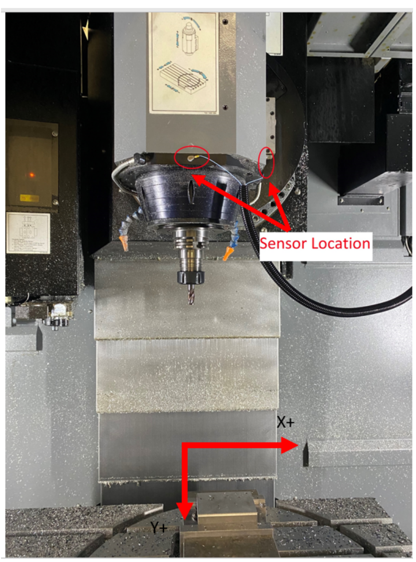

Fig. 4. Experimental Setup on Hurco VMX42SRTi indicating sensor mounting

Cutting experiments were completed by performing straight cuts in the X and Y direction, and then completing a "ramp" cut in the XYZ direction. The 3D model of this ramp can be seen in Fig.5, with the resulting cut part in Fig.6. These cutting experiments were conducted on 1020 mild steel with depths of cut ranging from 6.35mm to 31.75mm at increments of 6.35mm. These experiments also varied the cutting speeds from 3000 RPM to 5000 RPM, at increments of 500 RPM.

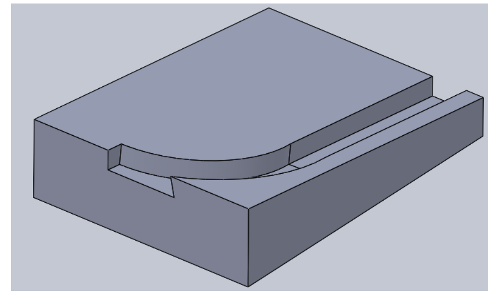

Fig. 5. Solidworks model of ramp cut with motion in the X,Y and Z axes

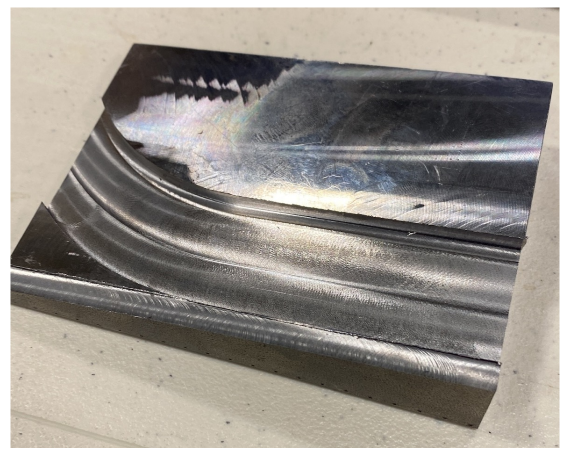

Fig. 6. Experimental results of ramp cut completed with chatter marks

The frequency response of the system was found from analyzing these cutting experiments. These experiments were then used to determine the ground truth for when chatter occurred, and verified by comparing the times indicating chatter, to the chatter marks seen in Fig.6. An example of one of these cutting signals from a ramp experiment can be seen in Fig. 7.

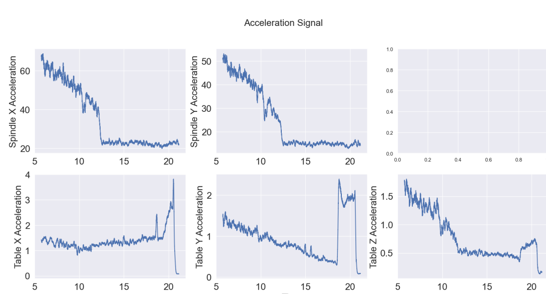

Fig. 7. Example Accelerometer signal for five-xxis machine

Once the cutting experiments were completed, an anomaly detection system was trained from scratch using the available data to develop an AutoEncoder based on the structure outlined above. This data set consisted of

150 000 data points, and the network was trained for 150 epochs over a span of 8 hours. This anomaly detection system was trained using all of the data available for stable cutting conditions, and then was provided cutting signals from both stable and chatter conditions. The performance of the system was evaluated based on the accuracy and recall, as defined as follows

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$
$$Recall = \frac{TP}{TP + FN}$$

Where TP are the true positive results, FP are the false positive results, TN are the true negative results, and FN represents the false negative results. These metrics are commonly used in different machine learning applications, and represent how accurate the system is, as well as its ability to detect the true positive rate. In this instance, the true positive rate represents the system's ability to catch chatter conditions if they truly occurred.

With this criteria in mind, the network trained from scratch is tested with the signal seen in Fig. 7. The result of this network can be seen in Fig. 8. As the figure shows, the system is able to capture many of the chatter dynamics, but has difficulty making a clear distinction at the 60 second mark. This results in an accuracy of 76.62 % and a recall of 78.75%, which acts as a baseline accuracy for the system.

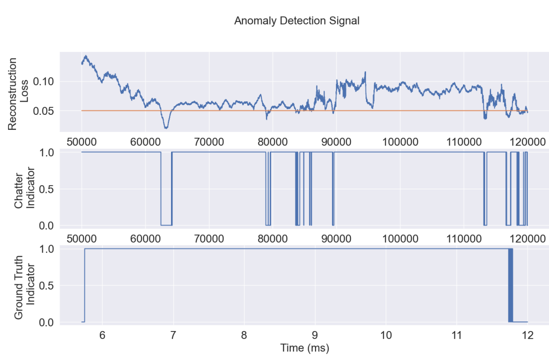

Fig. 8. Anomaly detection signal from directly trained network compared to ground truth

Following the approach described in the methodology section, an initial network is created using the transfer and ensemble learning approach. The three-axis portion of this network is transferred from a Hurco VMX42Di system trained under similar conditions as those listed in the experimental setup. This network was trained on 25 000 data points for 150 epochs, over a span of two hours. The results of this initial network are quite poor and can be seen in Fig. 9. This network only has an accuracy of 16.34 % and a recall of 13.28 %.

However, as additional data is provided to the system, the incremental learning methodology can be implemented. These weak ensemble learners allow the network

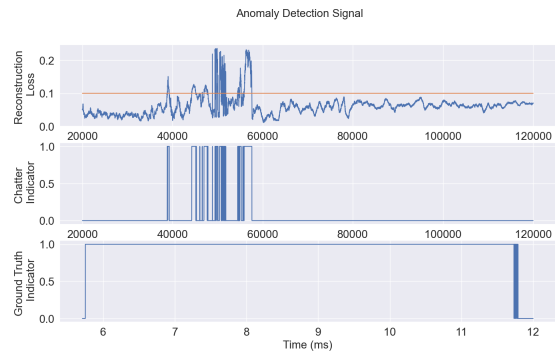

Fig. 9. Anomaly detection signal from network trained with only transfer learning compared to ground yruth

to gain a more complete understanding of the impact the additional aspects of the five-axis mill, namely the swivel and tilt, have on the system. This network was trained on a total of 75 000 data points, over a span of three hours. The results of this network can be seen in Fig. 10. By implementing this approach, we are able to obtain an accuracy of 72.87% and a recall of 74.85%, which is comparable to the network trained from scratch. The comparison of these metrics for the various networks can be seen in summary in Table II.

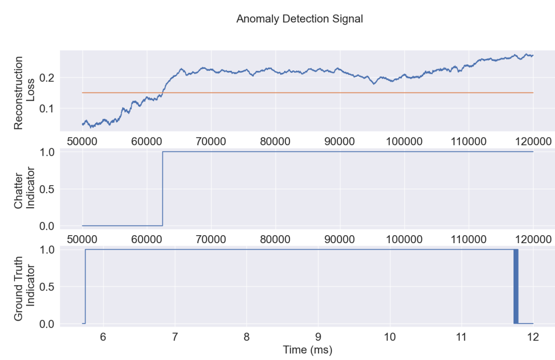

Fig. 10. Anomaly detection signal from network trained with transfer and ensemble incremental learning compared to ground truth

TABLE II
Comparison of network performance metrics

Network Configuration	Accuracy Score (%)	Recall Score (%)
Trained from Scratch	76.62	78.75
Transfer and Ensemble Learning Only	16.34	13.28
Transfer and Ensemble Learning with Incremental Learning	72.87	74.85

V. Discussion

In this work we developed an LSTM-AutoEncoder based anomaly detection system for three and five-axis CNC milling machines using on transfer learning. This approach allows us to capitalize on the training effort from a source system to reduce the required training data and time for a target network. This is accomplished by leveraging the similarities in structural design between certain three-axis CNC machines, and certain five-axis CNC machines, by implementing an ensemble-incremental learning algorithm. This ensemble learner is capable of learning the differences between the three-axis machine and the five-axis machine, and achieving comparable accuracy to a network trained from scratch by using incremental learning.

By following this method, a machine tool manufacturer, or end user, can deploy a rudimentary system when a machine is first installed. The system can continue to learn machine specific dynamics during its calibration and use incremental learning to obtain the same level of performance as a system that had been trained from scratch. This can provide significant benefits to users of five-axis CNC milling machines, where the same level of development as a three-axis machine does not exist. For example, in our work in [33], we developed anomaly detection systems using the methods described above on one three-axis CNC machine. This system was developed using a large data set with millions of data points. We were then able to quickly modify this system to three other three-axis machines from the same manufacturer, with varying size and drive specifications. This was accomplished using only a fraction of the necessary data. These experiments demonstrated the potential of this approach for three-axis CNC machines, and can likely be extended to five-axis CNC machines as well.

It should be noted that the accuracy for the five-axis milling machines was lower than the three-axis milling machines in this work. In many anomaly detection algorithms for three-axis milling machines, a typical accuracy would be greater than 85%, with the accuracy in this work being approximately 73%. As outlined in [40], the mechanistic model of a five-axis CNC machine is derived from not only the spindle displacement, but the tilt angle as well. Although we are able to infer some details of the tile angle by the relative acceleration between the X,Y and Z accelerometers, we do not have a measurement of the specific angle that the spindle is cutting at. We believe that this missing parameter impacts the network's ability to accurately predict the anomaly conditions for the five-axis CNC machine, but requires further investigation to confirm it.

VI. Conclusion

There is a drive to develop fully autonomous CNC machines and a key part to achieving this goal is the development of an anomaly detection system. There has been considerable development towards this effort using model based, sensor based and data driven approaches. In recent years, machine learning algorithms have provided powerful tools to achieve this goal, but require significant amounts of data to properly function. Transfer learning has emerged as a potential solution to this problem by transferring the weights from a source network to a target network as a starting point for training. This reduces the required training data and time for a network to become fully operation. Although the use of transfer learning has been effective in three-axis CNC machines, there has not been significant investigation into the feasibility of these techniques for five-axis CNC machines.

Capitalizing on the similarity between certain three-axis and five-axis CNC machines, an ensemble learning approach was implemented for a five-axis CNC machine. In this approach, the encoder from a three-axis CNC network is taken and combined with an ensemble learner, then fed into a combined decoder. This approach allows for key weights from the three-axis machine to be transferred. This system is then integrated with an incremental learning approach that allows the network to learn the missing dynamics of the five-axis CNC machine. This approach shows that the ensemble learner can obtain similar accuracy to a network trained from scratch. However, richer data from the five-axis CNC machine is likely required to improve the accuracy of the network.

Although the performance of the specific network demonstrated in this work is limited, the method shows great potential. There are often many machine learning systems that are designed for three-axis machines, and this approach demonstrates the potential of extending them to five-axis machines without the need for collecting a significant amount of user data. If this approach can be further developed, there may be significant advances made for five-axis CNC machines.

Acknowledgment

We would like to acknowledge the generous support provided by Hurco Companies Inc. Through their support we have had the opportunity to gain insight into the dynamics of their machines and to obtain a significant amount of training data.

References

[1] G. Erdoğan, "Land selection criteria for lights out factory districts during the industry 4.0 process," Journal of Urban Management, vol. 8, no. 3, pp. 377–385, 12 2019.

[2] Y. Liang, S. Wang, W. . Li, X. Lu, Y. C. Liang, and W. D. Li, "Data Driven Anomaly Diagnosis for Machining Processes Data-Driven Anomaly Diagnosis for Machining Processes," Journal of Intelligent Manufacturing, 2019. [Online]. Available: https://doi.org/10.1016/j.eng.2019.03.012

[3] M. Demetgul, Q. Zheng, I. N. Tansel, and J. Fleischer, "Monitoring the misalignment of machine tools with autoencoders after they are trained with transfer learning data," International Journal of Advanced Manufacturing Technology, vol. 128, no. 7-8, pp. 3357–3373, 10 2023. [Online]. Available: https://link.springer.com/article/10.1007/s00170-023-12060-2

[4] G. Serin, . B. Sener, . A. M. Ozbayoglu, and H. O. Unver, "Review of tool condition monitoring in machining and opportunities for deep learning," The International Journal of Advanced Manufacturing Technology, 2020. [Online]. Available: https://doi.org/10.1007/s00170-020-05449-w

[5] B. Shi and H. Attia, “Current status and future direction in the numerical modeling and simulation of machining processes: A critical literature review,” pp. 149–188, 4 2010.

[6] C. A. K. A. Kounta, L. Arnaud, B. Kamsu-Foguem, and F. Tangara, “Review of AI-based methods for chatter detection in machining based on bibliometric analysis,” pp. 2161–2186, 9 2022.

[7] G. Li, Y. Fu, D. Chen, L. Shi, and J. Zhou, “Deep anomaly detection for CNC machine cutting tool using spindle current signals,” Sensors (Switzerland), vol. 20, no. 17, pp. 1–18, 9 2020.

[8] Y. M. Al-Naggar, N. Jamil, M. F. Hassan, and A. R. Yusoff, “Condition monitoring based on IoT for predictive maintenance of CNC machines,” in Procedia CIRP, vol. 102. Elsevier B.V., 2021, pp. 314–318.

[9] B. Lindemann, B. Maschler, N. Sahlab, and M. Weyrich, “A survey on anomaly detection for technical systems using LSTM networks,” 10 2021.

[10] I. Kvinevskiy, S. Bedi, and S. Mann, “Detecting machine chatter using audio data and machine learning,” International Journal of Advanced Manufacturing Technology, vol. 108, no. 11-12, pp. 3707–3716, 6 2020.

[11] A. Burkov, The Hundred-Page Machine Learning Book. Andriy Burkov, 2019.

[12] Y. M. Kim, S. J. Shin, and H. W. Cho, “Predictive Modeling for Machining Power Based on Multi-source Transfer Learning in Metal Cutting,” International Journal of Precision Engineering and Manufacturing - Green Technology, vol. 9, no. 1, pp. 107–125, 1 2022.

[13] H. O. Unver and B. Sener, “Exploring the Potential of Transfer Learning for Chatter Detection,” in Procedia Computer Science, vol. 200. Elsevier B.V., 2022, pp. 151–159.

[14] M. C. Yesilli, F. A. Khasawneh, and A. Otto, “On transfer learning for chatter detection in turning using wavelet packet transform and ensemble empirical mode decomposition,” CIRP Journal of Manufacturing Science and Technology, vol. 28, pp. 118–135, 1 2020.

[15] M. Postel, B. Bugdayci, and K. Wegener, “Ensemble transfer learning for refining stability predictions in milling using experimental stability states,” International Journal of Advanced Manufacturing Technology, vol. 107, no. 9-10, pp. 4123–4139, 4 2020.

[16] G. Hao and Z. Kunpeng, “Pyramid LSTM auto-encoder for tool wear monitoring,” in 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE). Online Zoom Meeting: IEEE, 8 2020.

[17] R. K. Vashisht and Q. Peng, “Online chatter detection for milling operations using LSTM neural networks assisted by motor current signals of ball screw drives,” Journal of Manufacturing Science and Engineering, Transactions of the ASME, vol. 143, no. 1, 1 2021. [Online]. Available: https://dx.doi.org/10.1115/1.4048001

[18] E. Li, S. Bedi, and W. Melek, “Anomaly detection in three-axis CNC machines using LSTM networks and transfer learning,” The International Journal of Advanced Manufacturing Technology, vol. 127, pp. 5185–5198, 2023. [Online]. Available: https://doi.org/10.1007/s00170-023-11617-5

[19] E. Li, Y. Li, S. Bedi, W. Melek, and P. Gray, “Incremental learning of LSTM-autoencoder anomaly detection in three-axis CNC machines,” The International Journal of Advanced Manufacturing Technology, pp. 1–13, 12 2023. [Online]. Available: https://link.springer.com/article/10.1007/s00170-023-12713-2

[20] Y. Altintas, Manufacturing Automation. Cambridge University Press, 1 2012.

[21] J. Tlusty and F. Ismail, “Basic Non-Linearity in Machining Chatter,” CIRP Annals, vol. 30, no. 1, pp. 299–304, 1 1981.

[22] E. Budak and Y. Altintaş, “Analytical Prediction of Chatter Stability in Milling—Part I: General Formulation,” Journal of Dynamic Systems, Measurement, and Control, vol. 120, no. 1, pp. 22–30, 3 1998. [Online]. Available: https://dx.doi.org/10.1115/1.2801317

[23] C. Deng, J. Tang, J. Miao, Y. Zhao, X. Chen, and S. Lu, “Efficient stability prediction of milling process with arbitrary tool-holder combinations based on transfer learning,” Journal of Intelligent Manufacturing, vol. 34, no. 5, pp. 2263–2279, 6 2023.

[24] S. Hochreiter and J. Urgen Schmidhuber, “Long Short-Term Memory,” Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: http://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf

[25] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to Forget: Continual Prediction with LSTM,” Neural Computation, vol. 12, no. 10, pp. 2451–2471, 10 2000. [Online]. Available: https://direct.mit.edu/neco/article/12/10/2451-2471/6415

[26] D. Azzalini, L. Bonali, and F. Amigoni, “A Minimally Supervised Approach Based on Variational Autoencoders for Anomaly Detection in Autonomous Robots,” IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 2985–2992, 4 2021.

[27] M. Said Elsayed, N. A. Le-Khac, S. Dev, and A. D. Jurcut, “Network Anomaly Detection Using LSTM Based Autoencoder,” Q2SWinet 2020 - Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks, pp. 37–45, 11 2020. [Online]. Available: https://doi.org/10.1145/3416013.3426457

[28] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, and X. Chen, “Deep Transfer Learning Based on Sparse Autoencoder for Remaining Useful Life Prediction of Tool in Manufacturing,” IEEE Transactions on Industrial Informatics, vol. 15, no. 4, pp. 2416–2425, 4 2019.

[29] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, “A survey on deep transfer learning,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11141 LNCS, pp. 270–279, 2018.

[30] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, “A Kernel Method for the Two-Sample-Problem,” Neural Information Processing Systems, 2008. [Online]. Available: www.kyb.mpg.de/bs/people/arthur/mmd.htm

[31] J. Wang, B. Zou, M. Liu, Y. Li, H. Ding, and K. Xue, “Milling force prediction model based on transfer learning and neural network,” Journal of Intelligent Manufacturing, vol. 32, no. 4, pp. 947–956, 4 2021. [Online]. Available: https://link.springer.com/article/10.1007/s10845-020-01595-w

[32] W. D. Li and Y. C. Liang, “Deep transfer learning based diagnosis for machining process lifecycle,” Procedia CIRP, vol. 90, pp. 642–647, 1 2020.

[33] E. Li, Y. Li, S. Kakarla, S. Bedi, and W. Melek, “Transfer Learning Methods for Three-Axis CNC Anomaly Detection,” Advanced Manufacturing, vol. 1, no. 1, 2024.

[34] E. Kuljanic, M. Sortino, and G. Totis, “Multisensor approaches for chatter detection in milling,” Journal of Sound and Vibration, vol. 312, no. 4-5, pp. 672–693, 5 2008.

[35] G. M. van de Ven, T. Tuytelaars, and A. S. Tolias, “Three types of incremental learning,” Nature Machine Intelligence 2022 4:12, vol. 4, no. 12, pp. 1185–1197, 12 2022. [Online]. Available: https://www.nature.com/articles/s42256-022-00568-3

[36] R. Polikar, L. Udpa, S. S. Udpa, and V. Honavar, “Learn++: An incremental learning algorithm for supervised neural networks,” IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, vol. 31, no. 4, pp. 497–508, 11 2001.

[37] M. D. Muhlbaier, A. Topalis, and R. Polikar, “Learn++.NC,” IEEE Transactions on Neural Networks, vol. 20, no. 1, pp. 152–168, 1 2009. [Online]. Available: https://dl.acm.org/doi/10.1109/TNN.2008.2008326

[38] H. Wang, M. Li, and X. Yue, “IncLSTM: Incremental Ensemble LSTM Model towards Time Series Data,” Computers & Electrical Engineering, vol. 92, p. 107156, 6 2021.

[39] J. P. Hung, Y. L. Lai, C. Y. Lin, and T. L. Lo, “Modeling the machining stability of a vertical milling machine under the influence of the preloaded linear guide,” International Journal of Machine Tools and Manufacture, vol. 51, no. 9, pp. 731–739, 9 2011.

[40] J. Li, Z. M. Kilic, and Y. Altintas, “General Cutting Dynamics Model for Five-Axis Ball-End Milling Operations,” Journal of Manufacturing Science and Engineering, Transactions of the ASME, vol. 142, no. 12, 12 2020.