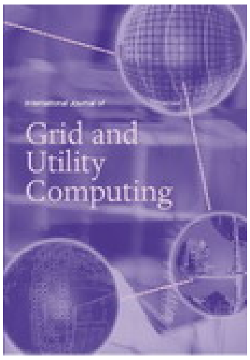


International Journal of Grid and Utility Computing

ISSN online: 1741-8488 - ISSN print: 1741-847X
https://www.inderscience.com/ijguc

A deep learning-inspired IoT-enabled hybrid model for predicting structural changes in CNC machines based on thermal behaviour

Thompson Stephan, Vinith Anand Thiyagu, Pavan Kumar Shridhar

DOI: 10.1504/IJGUC.2024.10061721

Article History:

Received:	07 June 2023
Last revised:	22 September 2023
Accepted:	04 October 2023
Published online:	19 February 2024

A deep learning-inspired IoT-enabled hybrid model for predicting structural changes in CNC machines based on thermal behaviour

Thompson Stephan*

Department of Computer Science & Engineering,
Graphic Era Deemed to be University,
Dehradun, Uttarakhand, India
Email: thompsoncse@gmail.com
*Corresponding author

Vinith Anand Thiyagu and Pavan Kumar Shridhar

Department of Computer Science Engineering,
M.S. Ramaiah University of Applied University,
Bengaluru, Karnataka, India
Email: tva1510@gmail.com
Email: pavankumarshridhar@gmail.com

Abstract: This research work introduces a hybrid model, BIG-LSTM, designed to enhance the precision of computer numerical control (CNC) machines in the manufacturing industry powered by the Internet of Things (IoT). Traditional models primarily focus on nut temperature's impact on thermal errors, often overlooking factors like bearing and ambient temperatures, and tend to ignore the intercept in the temperature-error relationship. The presented model addresses these gaps by incorporating ambient and bearing temperatures, and considering both intercept and slope for predicting Z-axis thermal deformation. Integration of motor speed and coolant behaviour is also included, acknowledging the rise in temperature with increased speed. BIG-LSTM, combining LSTM, GRU, and Bi-LSTM models, demonstrates efficacy in experiments, achieving Root Mean Square Errors (RMSEs) within 0.9 μm for spindle thermal displacement under varied temperature conditions. These findings highlight the model's potential in significantly improving accuracy and robustness in spindle thermal displacement predictions in the IoT era.

Keywords: CNC machine tools; thermal errors; hybrid model; deep learning; LSTM; spindle thermal displacement; prediction accuracy.

Reference to this paper should be made as follows: Stephan, T., Thiyagu, V.A. and Shridhar, P.K. (2024) 'A deep learning-inspired IoT-enabled hybrid model for predicting structural changes in CNC machines based on thermal behaviour', *Int. J. Grid and Utility Computing*, Vol. 15, No. 1, pp.3–15.

Biographical notes: Thompson Stephan, an Associate Professor in Computer Science & Engineering at Graphic Era Deemed to be University, India, completed his PhD from Pondicherry University in 2018. He excels in applying AI techniques in practical settings. Honoured as one of the world's top 2% most influential scientists in 2023 by Elsevier and Stanford University, he also received the IEEE Best Researcher Award-2020 and the Protsahan Research Award in 2023. His work, published extensively in prestigious journals, focuses on AI's practical applications.

Vinith Anand Thiyagu, a proficient and skilled Artificial Intelligence and Machine Learning Engineer, holds a Bachelor's degree in Electronic and Communication Engineering from Bangalore Institute of Technology, Bangalore (2021), establishing a solid foundation in electronic systems. Furthering education, he pursued a Master's degree at Ramaiah University of Applied Sciences, Bangalore (2023), specialising in Artificial Intelligence and Machine Learning. This academic journey reflects a deep commitment to advancing knowledge in cutting-edge fields.

Pavan Kumar Shridhar is a dynamic professional in Data Science and Engineering, graduating in Mechanical Engineering from PES University, Bangalore, in 2021, and obtaining an MTech from MS Ramaiah University of Applied Sciences, Bangalore, in 2023. Seamlessly transitioning between fields, he excels at the intersection of Mechanical Engineering and Data Science, particularly in AI and ML applications. His passion for leveraging cutting-edge technologies is evident in his practical problem-solving skills and proficiency in data science techniques.

1 Introduction

With the advancement of Industry 4.0, the demand for Computer Numerical Control (CNC) machines is on the rise. Achieving ultra-precision in component manufacturing is the primary objective of the industry, but thermal error plays a significant role when errors are generated within microns. Geometric and kinematic errors are corrected for the maximum error during an operation, but thermal error induced by the deformation or distortion of machine elements due to heating and temperature rise is one of the primary reasons for inaccuracy. Traditional methods are inadequate in reducing thermal error (Ramesh et al., 2000). The spindle, critical to machine tool performance, generates substantial amounts of heat during high-speed operation and is a key contributor to heat sources that have significant implications for machine design and operation (Haitao et al., 2007). Thermally generated errors have become one of the most common sources of error in precision and high-precision machining, causing up to 75% of the overall geometrical defects in machined workpieces (Mayr et al., 2012). To decrease thermal error, researchers have suggested redesigning machine tools from thermal stability materials, but this approach is expensive and has additional drawbacks such as chatter and decreased machine tool acceleration (Tanabe et al., 1986). To establish the link between temperature and spindle deformation in a machine tool and provide the foundation for a controller to account for spindle thermal displacement, mathematical machine-learning models for spindle thermal displacement have been presented (Zhang et al., 2016). Previous modelling techniques have used Principal-Based Models (PBM) and Empirical-Based Models (EBM) to study the correlations between temperatures and thermal errors based on numerical simulations or experimental data (Li et al., 2015). EBM is a type of 'black box' technique that operates under the assumption that thermal faults may be viewed as a function of certain crucial thermal discrete temperature points on the machine. Of all the EBM techniques, the use of regression analysis or artificial neural networks, also known as ANNs, is the most frequently employed (Ye et al., 2020). Li et al. (2019) used principal component analysis to choose thermal important points and performed k-means clustering. The field of deep learning has made remarkable progress and now operates at the cutting edge in several industries. Deep learning can gain knowledge of complex features by combining and learning basic characteristics from data. It outperforms alternative methods in various machine-learning problems by using massive data sets and computationally efficient training techniques (Xu et al., 2019). In this study, we propose a hybrid BIG-LSTM deep learning approach to model thermal error in CNC machines. The model comprises three Gated Recurrent Unit (GRU) layers, two Bi-LSTM layers, and two Long Short-Term Memory networks (LSTM) layers. We incorporate IoT into the model by considering the ambient temperature, bearing temperature and coolant behaviour. The spindle motor speed and coolant temperature variations are used to model temperature changes. Thermal error testing with various temperature rises and drops demonstrated that the suggested model's Root Mean Square Errors (RMSEs) were within 0.9 μm. The proposed hybrid model can serve as a base for implementing spindle thermal displacement correction in a machine tool to preserve machining precision.

In recent years, Song et al. (2022) explored various approaches to enhance the performance and precision of manufacturing processes. Hossain et al. (2023) highlighted the broader spectrum of research aimed at improving precision and efficiency in manufacturing processes. The preheating phase in machine tool operations, which can be time-consuming, consumes a significant amount of energy. Its purpose is to achieve thermal stability, ensuring precise machining. Temperature fluctuations can lead to thermal errors, affecting the accuracy of machining. To address this, a strategy and method for modelling and predicting thermal errors in machine tools are presented. This approach eliminates the need for lengthy preheating, maintaining high precision by adjusting the tool's position in real-time based on temperature variations. Thermal error compensation enhances precision, reduces scrapped parts, and allows for increased machining speed. It's crucial in industries with strict quality requirements. Compensation is tailored to each machine's unique temperature-displacement behaviour, ensuring consistent accuracy during operation. In many sectors, especially those that depend on precise and accurate machinery and procedures, thermal error correction is essential. A crucial component of contemporary manufacturing businesses is thermal error compensation in CNC machines. The accuracy and quality of the products produced by CNC machines can be severely impacted by heat changes at micron level, which are employed in precision machining and manufacturing. High-precision material cutting and shaping is accomplished by CNC machines through the use of pre-programmed instructions. The components of the machine, however, may expand or contract as a result of temperature changes in its surroundings, which could result in produced pieces with inaccurate dimensions. Real-time temperature monitoring of the machine and process adjustments is required for thermal error compensation. For correct fit and functionality, several industries demand that parts be manufactured with very tight tolerances. Manufacturing procedures are more dependable and accurate as a result of thermal error correction, which enables CNC machines to maintain these tight tolerances even in situations with changing temperatures. Hence, using

AI to model the thermal behaviour of the machine provides a cost effective solution.

The organisation of this paper is as follows. Section 2 provides a review of related works in the field of thermal error modelling, highlighting various models that have been employed in the past. The contributions of this paper are listed in Section 3. In Section 4, the working principles of GRU, LSTM and Bi-LSTM, the three types of recurrent neural networks used in the proposed thermal error prediction model, are described in detail. Section 5 presents the proposed thermal error prediction model, BIG-LSTM, including the architecture and design considerations. Performance evaluation of the model is discussed in Section 6, which includes comparison with traditional models and analysis of prediction accuracy. Finally, Section 7 concludes the paper with a summary of the key findings and directions for future research.

2 Related work

Cheng et al. (2022) proposed a Method for Analysing and Compensating for Thermal Deformation. This method involves finite element analysis to simulate the thermal deformation of the milling head, followed by using a polynomial fitting method to model and compensate for the deformation. The paper provides a detailed explanation of the method and presents experimental results that demonstrate the effectiveness of the proposed method in reducing thermal deformation of the milling head. Zhu et al. (2022) proposed a Robustness of Machine Tool Workpieces Using Random Forest Algorithm. The accuracy and robustness of machine tool workpieces are significantly influenced by thermal errors. However, the nonlinearity of these errors limits the prediction model's robustness and accuracy. To address this issue, the paper proposes a novel thermal error modelling method based on the random forest algorithm. The model's hyper-parameters can be easily optimised by the grid searching method integrated with cross-validation, with the temperature features measured as the model input. The proposed model can evaluate temperature feature importance based on the out-of-bag data generated during the modelling process. In order to improve the model's accuracy and lower computational costs, it is also taken into account to choose important temperature points in order to eliminate features that are unnecessary and the hysteresis effect between temperature and deformation. Over 90% prediction accuracy is maintained across a variety of operating situations, demonstrating the proposed model's accuracy and robustness through experimental validation. The suggested model delivers greater prediction accuracy and more resilience compared to standard machine learning approaches while requiring less training data, parameter adjustment that is quicker and more intuitive and stronger robustness. A Transfer Learning-Based Error Control Method was proposed by the authors in (Liu et al., 2022). The precision of machining complicated components is severely hampered by thermal problems. Data-based models' limited resilience and poor prediction accuracy, however, have restricted their application. This paper suggests a transfer learning-based error control system that makes use of a Long Short-Term Memory Network (LSTMN) for error prediction in order to increase resilience. It is suggested to use an improved filter to eliminate singular values and high-frequency noise. The deep residual network is constructed with a pre-activated residual block incorporated inside it. To increase robustness, a transfer learning model is created and a novel error prediction model is presented. The actual machining experiments validate the high predictive abilities of the transfer learning models. Overall, the proposed method has the potential to improve the accuracy and efficiency of machining processes and lead to better product quality. Yin et al. (2019) proposed a Method that Combines the Fuzzy C-Means Clustering Algorithm and Correlation Analysis to Select Temperature-Sensitive Points. This study proposes a modelling method for predicting the thermal error of machine tool spindles using a selective ensemble of BP neural networks. The method combines the fuzzy c-means clustering algorithm and correlation analysis to select temperature-sensitive points. Individual BP neural network models with unstable prediction performance are generated using different training sets and random initial parameters. Each model is then assigned a weight, which is evolved using a genetic algorithm. The ensemble model is formed by selecting individual models based on a threshold value. The method is tested on a horizontal machining centre THM6380, and its performance is compared with single BP neural network, multiple linear regression and least-square support vector machine models. Results demonstrate the superiority of the proposed method in predicting and compensating for thermal errors in machine tool spindles. The study provides a new approach to thermal error modelling and compensation. Tan et al. (2021) proposed a Segment Fusion Least Squares Support Vector Machine (SF-LSSVM) Method. In order to predict and compensate for thermal errors in CNC machines, it is important to consider key temperature points as input variables. However, these temperature points can change over time, which can negatively impact prediction accuracy. To overcome this issue, the SF-LSSVM thermal error modelling method is proposed. This involves dividing temperature and thermal error data into different time segments and using the LSSVM model to build sub-models for each segment. Key temperature points for each segment are selected using genetic algorithms. The sub-models are then fused together to create a final thermal error model that incorporates both local and global prediction characteristics. The model was tested on a horizontal machining centre, and the mean RMSE on 5 spindle speeds after compensation was only 3.1 μm. The proposed method outperformed traditional thermal error models by up to 51%, providing new insights into key temperature points and thermal error prediction methods.

Wei et al. (2022) proposed a Thermal Error Modelling Method Based on Gaussian Process Regression (GPR). Precision CNC machine tools are susceptible to thermal errors, which can negatively impact their accuracy. To minimise this effect, thermal error modelling and compensation are commonly used. However, existing models only provide point predictions of thermal error, ignoring the stochastic nature of these errors and the need for reliable risk analysis. A novel technique for predicting thermal inaccuracy that is based on Gaussian Process Regression (GPR) is put forth to overcome these constraints. Interval forecasts of thermal error are provided by the GPR model, which has good prediction accuracy and resilience. To improve its ability to predict outcomes, the model employs numerous batches of experimental data and adaptively chooses Temperature-Sensitive Points (TSPs) during training. Moreover, the model's interval predictions of thermal errors enable reliable risk analysis. Experimental results demonstrate that the proposed GPR model outperforms existing models under various working conditions. Thermal error compensation experiments were also conducted, confirming the effectiveness of the proposed model. This study presents a novel and reliable approach to thermal error modelling and compensation for precision CNC machine tools. Li et al. (2021) proposed a Beetle Antennae Search Algorithm (BAS) Method. High-speed motorised spindles can experience thermal errors due to heating, which can adversely affect the accuracy of machine tools. A thermal error model for high-speed motorised spindles is suggested as a solution to this problem in order to account for thermal faults and enhance machining precision. The Beetle Antennae Search algorithm (BAS) is used to optimise the model to increase its accuracy. Temperature and axial thermal drift data are gathered for the A02 motorised spindle at various speeds as part of an investigation on thermal characteristics. Gray relational analysis and fuzzy clustering are used to find temperature-sensitive locations. The BP neural network's weights and thresholds are optimised using BAS to create the BAS-BP thermal error prediction model. Results reveal that at various speeds, BAS-BP has a greater prediction accuracy than BP and GA-BP models. The BAS-BP model is therefore appropriate for spindle thermal error prediction and adjustment. Zimmermann et al. (2020) proposed a method that utilises Thermal Adaptive Learning Control to select optimal inputs for compensation models. This approach automatically adapts the number and individual inputs for each considered thermal error. To handle time series of missing data thermal error measurements, it combines k-means clustering with Time Series Cluster Kernel. The robustness of the compensation model is greatly increased by this adaptive sensor selection strategy, according to experimental findings on a 5-axis machine tool. About 40% less productivity is lost as a result of on-machine measurements. For modelling thermal error correction for the displacement of the cutter position in the *Y*- and *Z*-axes of CNC machine tools, Chen and Hung (2021) suggested a Backward Elimination (BE) Algorithm Method. The feature selection technique for the multiple regression models uses the BE algorithm based on mean squares of *K*-fold errors reduction. The model's performance is assessed using K-fold Cross-Validation (KCV) on a small set of training data. By choosing characteristics for the *Y*- and *Z*-axes, the multiple regression model is created. Test results show that the method effectively reduces the peak-to-peak value of thermal error in both directions. Specifically, the peak-to-peak value of thermal error is reduced from about 55 μm to below 14 μm in the *Y*-direction and from about 74 μm to below 19 μm in the *Z*-direction. By automatically starting on-machine measurements when unknown thermal circumstances arise, the One-Class Support Vector Machines that Zimmermann et al. (2021) improves the self-optimisation capacity of thermal error compensation models and novelty detection method based on this detects these circumstances, which are not reflected in the training data of the compensatory models. The outcomes show that the trade-off between accuracy and productivity for thermal error correction is eliminated by the autonomously initiated on-machine measurements used with a 5-axis machine tool. Without significantly impacting the accuracy of the thermal error correction, the time required to identify a deviation from preset limits is decreased by 78%.

3 Contribution

To address the research gaps in the existing literature, this research paper makes the following contributions:

1 Proposed a hybrid model, called BIG-LSTM, for predicting *Z*-axis thermal deformation in CNC machines.

2 BIG-LSTM incorporates LSTM, GRU and bi-LSTM models, and takes into account the impact of ambient temperature and bearing temperature on thermal errors.

3 Achieved root mean square errors within 0.9 μm in experiments measuring spindle thermal displacement under varying temperature changes, indicating high accuracy and robustness of the proposed model.

4 Overview of recurrent neural network for thermal error modelling

4.1 Gated recurrent unit

Recurrent Neural Networks (RNNs) are a subclass of neural networks designed to handle sequential data, such as time-series data, audio and natural language text. One of the most widely used RNNs is the Gated Recurrent Unit (GRU). GRUs have a gating mechanism that allows them to selectively remember or forget information from previous time steps as shown in Figure 1. This mechanism is controlled by two gates: the update gate and the reset gate. The update gate determines how much of the new information should be retained, while the reset gate decides how much of the previous state should be discarded. These functions are described in the below equations.

$$r_t = \text{sigmoid}\left(W_r \cdot \left[h_{t-1}, x_t\right]\right) \tag{1}$$

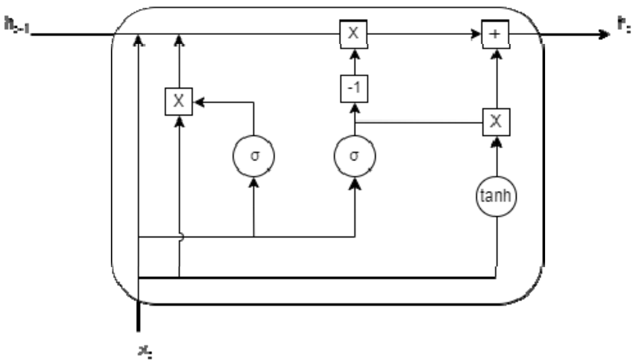
$$z_t = \text{sigmoid}\left(W_z \cdot \left[h_{t-1}, x_t\right]\right) \tag{2}$$

$$h_{t'} = \tanh\left(W_h \cdot \left[r_t \cdot h_{t-1}, x_t\right]\right) \tag{3}$$

$$h_t = \left(1 - z_t\right) \cdot h_{t-1} + z_t \cdot h_{t'} \tag{4}$$

where: r_t is the reset gate at time step t, z_t is the update gate at time step t, h_{t-1} is the hidden state at the previous time step, x_t is the input at time step t, $h_{t'}$ is the candidate activation at time step t, W_r, W_z and W_h are learnable weight matrices, [;] represents concatenation of vectors along the second dimension (i.e., horizontal concatenation).

Figure 1 GRU cell architecture



4.2 Long short-term memory

Long Short-Term Memory (LSTM) is a type RNN commonly used in applications that involve sequential data and Natural Language Processing (NLP). Traditional RNNs suffer from vanishing gradients, which make it difficult for the model to learn long-term dependencies. As the gradient is propagated backward through time, it becomes very small. To address this issue, LSTM introduces a long-term memory cell and three gates (input, forget and output) that regulate the flow of information into and out of the cell. The input gate, forget gate and output gate are responsible for determining which information to input into the cell, discard from the cell and output from the cell, respectively. These gates are controlled by sigmoid activation functions, which take the input and the cell's previous state as inputs, as shown in Figure 2. The below equations describe these functions.

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right) \tag{5}$$

$$i_t = \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right) \tag{6}$$

$$g_t = \tanh\left(W_c\left[h_{t-1}, x_t\right] + b_c\right) \tag{7}$$

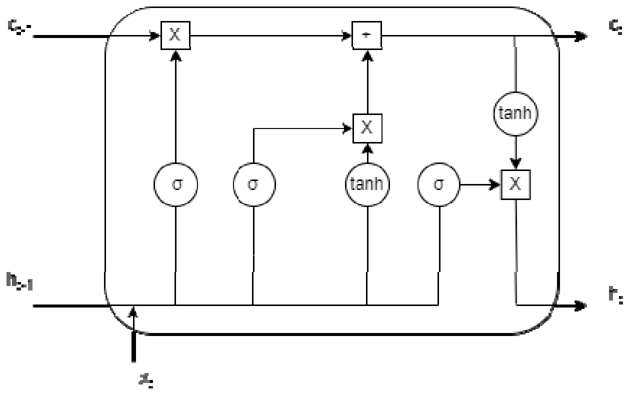
$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \tag{8}$$

$$c_t = f_t * c_{t-1} + i_t * g_t \tag{9}$$

$$h_t = o_t * \tanh\left(c_t\right) \tag{10}$$

where f_t, i_t and o_t are the forget, input and output gates respectively. g_t is the candidate memory cell content. c_t is the memory cell state. h_t is the output of the LSTM cell at time t. x_t is the input at time t and h_{t-1} is the output of the LSTM cell at the previous time step $t-1$. W and b are the weight and bias matrices/parameters, respectively. σ is the sigmoid activation function and tanh is the hyperbolic tangent activation function.

Figure 2 LSTM cell architecture



4.3 Bi-directional long short-term memory

Bi-directional Long Short-Term Memory (Bi-LSTM) is a Recurrent Neural Network (RNN) architecture widely used in Natural Language Processing (NLP) tasks such as speech recognition, machine translation, sentiment analysis and text categorisation. A Bi-LSTM network consists of two LSTM layers, one processing the input sequence forward and the other backward, as depicted in Figure 3. This bidirectional processing allows the network to capture both past and future context of the input sequence, leading to better performance in many NLP tasks. During training, Bi-LSTM networks use backpropagation through time to adjust the weights and minimise the error signal as it propagates backwards through the network. The functions used in the Bi-LSTM operation are described in the below equations.

$$h_t^f = LSTM\left(x_t, h_{t-1}^f, c_{t-1}^f\right) \tag{11}$$

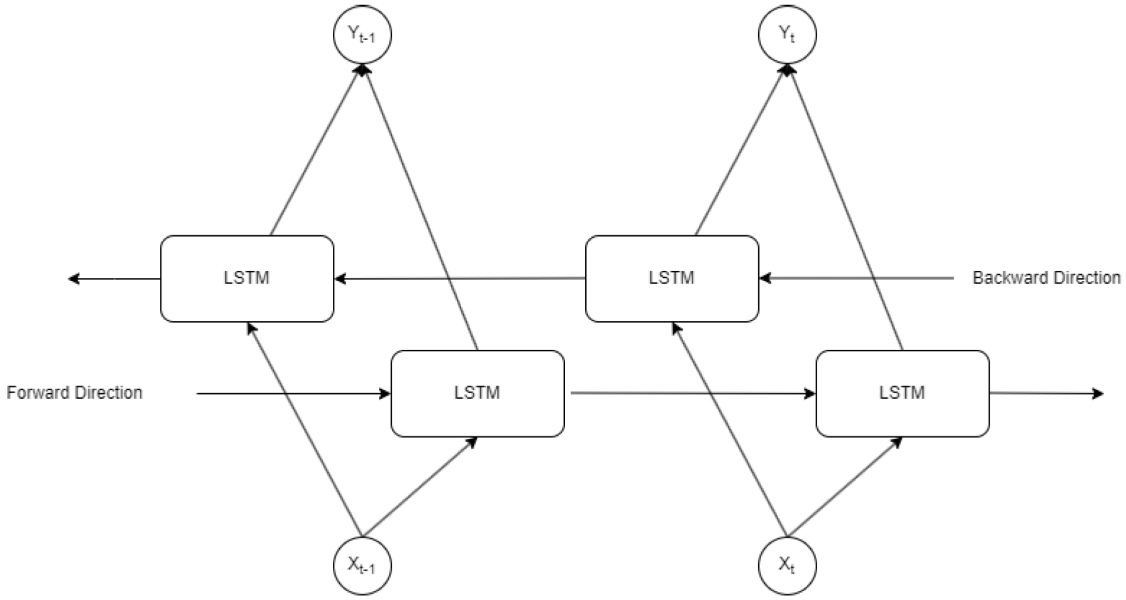
$$c_t^f = LSTM_c\left(x_t, h_{t-1}^f, c_{t-1}^f\right) \tag{12}$$

$$h_t^b = LSTM\left(x_t, h_{t+1}^b, c_{t+1}^b\right) \tag{13}$$

$$c_t^b = LSTM_c\left(x_t, h_{t+1}^b, c_{t+1}^b\right) \tag{14}$$

where x_t is the input at time step t, h_t^f and h_t^b are the hidden states of the forward and backward LSTM layers at time step t, c_t^f and c_t^b are the cell states of the forward and backward LSTM layers at time step t, and LSTM and LSTM_c denote the LSTM and cell state update functions, respectively.

Figure 3 Bi-LSTM cell architecture



The output of the Bi-LSTM layer is a concatenation of the forward and backward hidden states as shown in the below equation:

$$h_t = \left[h_t^f, h_t^b \right] \tag{15}$$

where $[,]$ denotes concatenation.

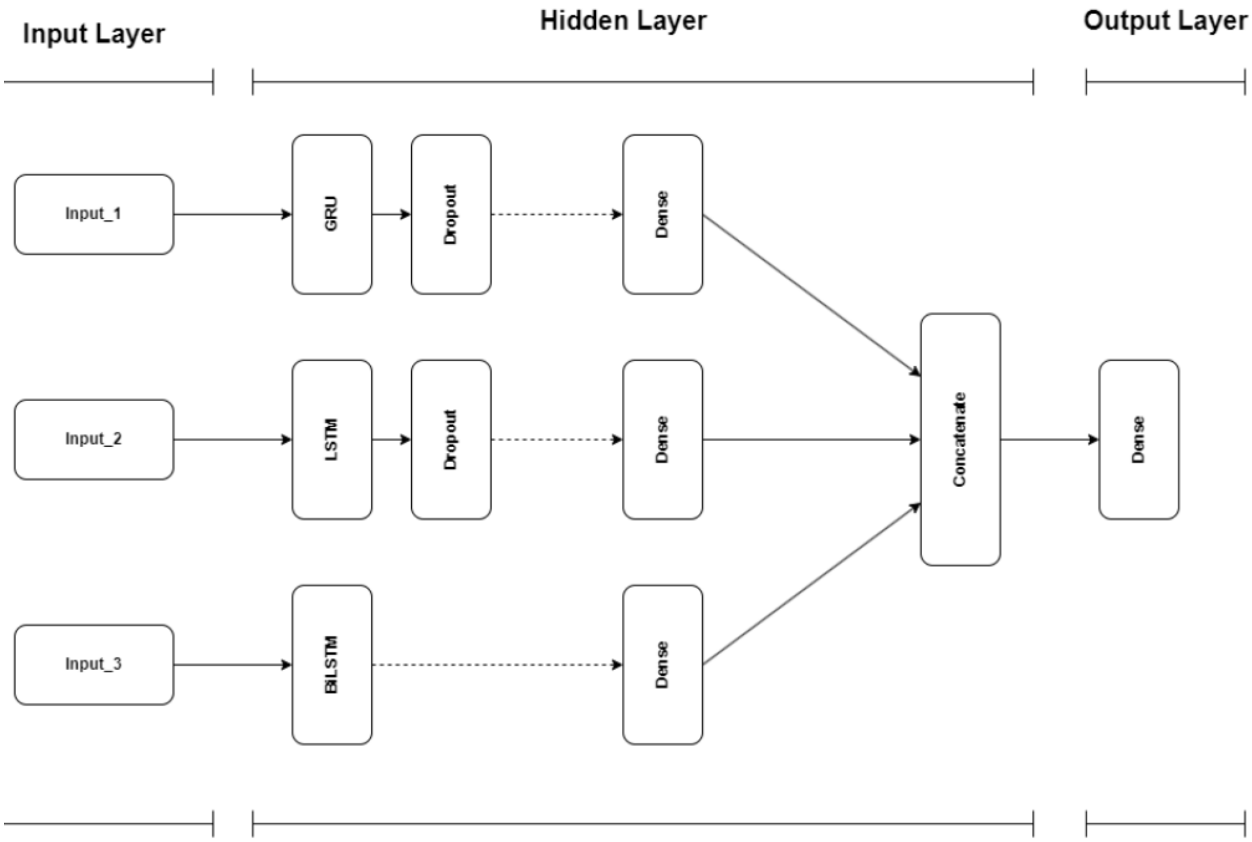
5 Proposed thermal error prediction model: BIG-LSTM

Changes in temperature result in dimensional errors or deviations in a system or component due to the cause-and-effect relationship between temperature and thermal error. The term 'thermal error' describes the difference between the actual measurements that need to be taken or the positions of the workpiece or cutting tool and the expected measurements or positions resulting from changes in the system caused by temperature variations. This relationship is particularly relevant in applications involving precision machinery, manufacturing and measurement. Most materials expand when heated and contract when cooled, a phenomenon known as 'thermal expansion and contraction'. When temperatures fluctuate, the materials used in mechanical systems and components, such as those in CNC machines, can either expand or contract, thereby altering the dimensions of the structure. Temperature variations can lead to changes in the shape of a machine or component, resulting in deviations from the original design parameters. These modifications can accumulate in complex systems or multi-component structures. This means that considering the combined effects of many components and materials, even minor temperature fluctuations can lead to significant errors. An uneven distribution of temperature in a system causes non-uniform thermal expansion, which can cause machine parts to warp, bend or twist. To mitigate the effects of thermal error, manufacturers and engineers employ thermal error compensation strategies. These methods involve real-time temperature monitoring and using the data to adjust the machine's operation to account for expected thermal discrepancies. Machine learning plays a crucial role in modelling the non-linearity and predicting thermal errors, enabling compensation in complex systems. By analysing temperature data from various regions of the machine, these temperatures serve as input features to predict the deviation between the workpiece and cutting tool, providing insight into the thermal characteristics of the machine.

5.1 Model framework

Figure 4 shows the proposed model framework, which contains an input layer with the model's input sequence. The output layer gives the predicted thermal error value. The hidden layer consists of seven layers, including three GRU layers, two LSTM layers and two bi-LSTM layers. Three different models are used to predict a single output value from a sequence of six inputs. The first model uses LSTM layers with a dropout layer in between, the second model uses two bidirectional LSTM layers, and the third model uses three GRU layers with dropout layers in between, as shown in Figure 4. These models are trained separately, and their outputs are concatenated and fed to a dense output layer with a single neuron and linear activation.

Figure 4 Proposed BIG-LSTM model architecture



The weights of the models are updated during training using backpropagation. During each iteration the model takes a batch of input sequences and makes predictions for each one. The difference between these predictions and the actual output values is measured using a loss function (mean squared error). The weights of the model are then adjusted using an optimiser (in this case, the Adam optimiser) to minimise this loss. This process is repeated for a specified number of epochs, which is the number of times the model sees the entire training data set.

Once the training is complete, the model can be used to make predictions on new input sequences. The input sequences are fed to the model, and the output value is computed by passing the concatenated output of the three models through the dense output layer with a single neuron and linear activation.

Overall, this model uses a combination of LSTM, bidirectional LSTM, and GRU layers to make predictions from input sequences. The output of these models is then combined to produce a final prediction. The weights of the models are updated during training using backpropagation and an optimiser to minimise the loss function, and the trained model can be used to make predictions on new input sequences.

5.2 Data collection and pre-processing for machine displacement behaviour analysis

Data on the machine's displacement behaviour is gathered every day. This entails the use of sensors to track the machine's motion and take the necessary measurements. Data from each day's machining operation is gathered over a number of days. This information offers a historical record of the changes in the machine's displacement behaviour over time. There are separate files where the information from each day's machining operation is kept. Each file includes details regarding the precise displacement behaviour that was seen on that specific day. These files contain timestamps, numbers and other pertinent information. This entails monitoring and documenting a machine's displacement behaviour over the course of many days of milling operations. This data collection and analysis method provides insightful information about the machine's performance and help make manufacturing environments more efficient. All these files have been re-structured to form sliding windows for training the model. As shown in Figure 5, the input format is [batch size, temperature points, a column of values for each temperature points]. The data set used in this paper contains 23,000 data points with nine features representing specific time, displacements on respective axes and temperature readings, including spindle rear, coolant fall, *X* bearing, ref rear, *Z* bearing, ref near transformer and ambient temperature. The data set also includes a column for the date and was used to investigate various aspects of the system under study. The change in displacement between work-piece and cutting tool with respect for time and input features is represented as shown in Figure 6.

Figure 5 Input format

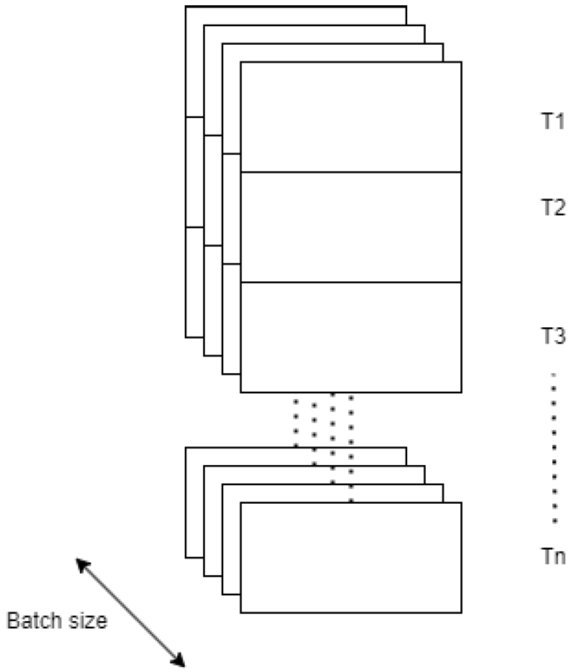
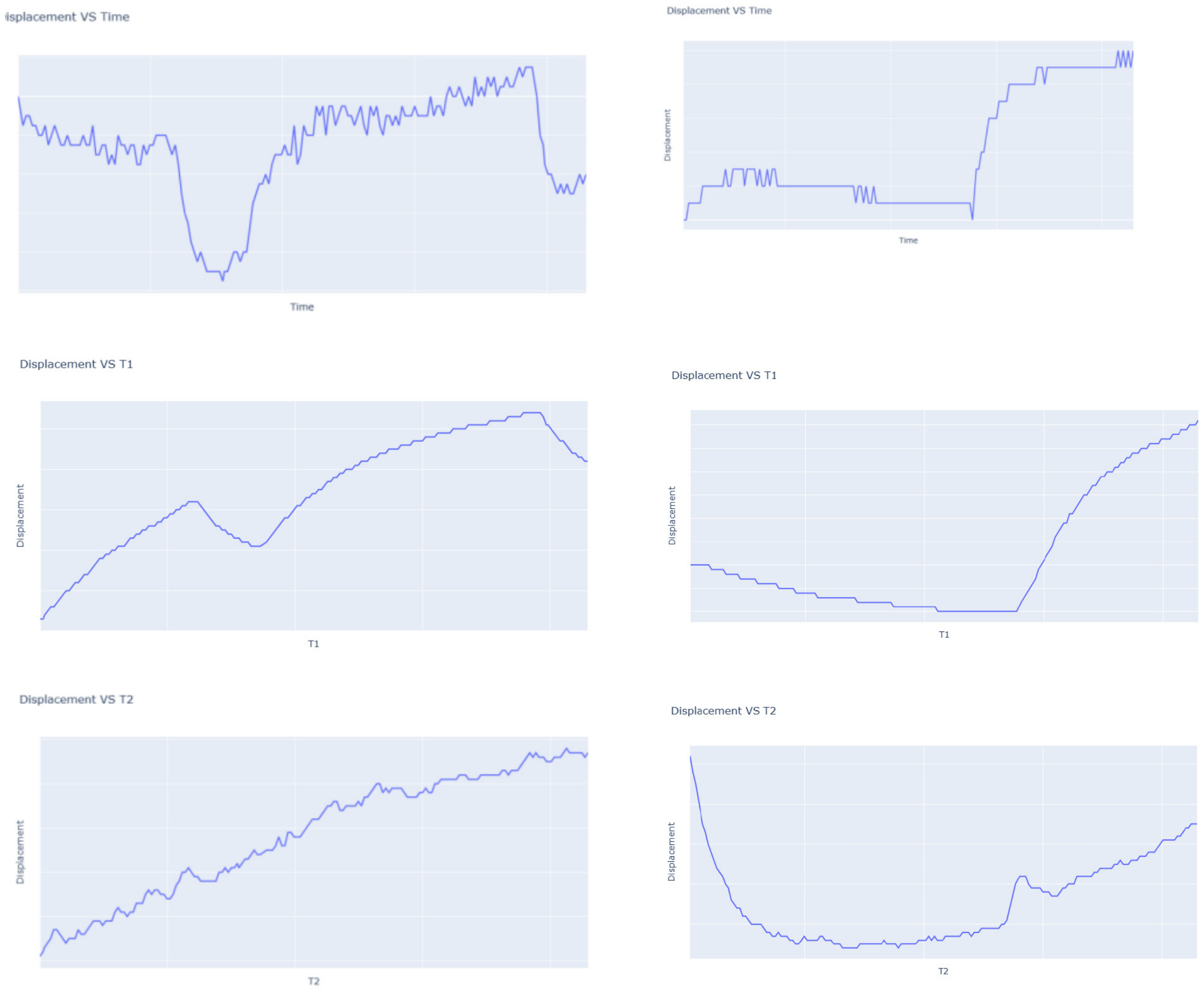


Figure 6 Change in displacement with respect to time



5.3 *Training model*

The process of selecting appropriate hyperparameters for the BIG-LSTM model outlined in this study was guided by careful consideration of the model's architecture and its impact on performance. Through defining three distinct models, specifically model 1, model 2, and model 3, diverse depths and architectures were explored to capture temporal dependencies and mitigate overfitting. In model 1, a two-layer LSTM configuration was employed, with the initial LSTM layer set to output sequences, followed by dropout layers to control model complexity. Model 2 featured bidirectional LSTM layers, enhancing the model's ability to capture both forward and backward temporal information, complemented by dropout layers for regularisation. Model 3 adopted a three-layer GRU structure, with dropout layers after the initial two layers to curb overfitting tendencies. These varied architectural choices enabled the models to capture different levels of temporal patterns, demonstrating a nuanced approach to handling complex data dependencies. The number of hidden units, denoted as 'LSTM 32 units' and 'GRU 64 units', played a pivotal role in shaping the models' capacity to learn from the input sequences. While higher numbers of hidden units can potentially enhance a model's ability to learn intricate relationships, they also introduce the risk of overfitting. Thus, moderate values for these parameters were meticulously selected to strike a balance between model complexity and generalisation ability. This deliberate choice aimed to maximise the models' predictive capabilities without compromising their ability to adapt to unseen data. Regarding learning rate optimisation, the Adam optimiser with a default learning rate was employed. The learning rate is a critical hyperparameter that can substantially influence the convergence speed and overall stability of training. The impact of these hyperparameter choices on the model's performance was substantial. Through the utilisation of a combination of multi-layer architectures, bidirectional structures and dropout layers, the trade-off between model complexity and overfitting was effectively managed, resulting in robust models capable of capturing intricate temporal dependencies. Notably, the choice of activation functions and the mean squared error loss function underscored the suitability of the model for regression tasks, aligning with the nature of the addressed problem. The description of the training model is provided below:

1. Data preparation is the initial step in model training, and it involves dividing the input data into training and validation sets. The former is utilised to train the model, while the latter is employed to evaluate the model's performance during training. To enhance the performance of deep learning models, normalisation is often applied, which involves scaling the input data to have zero mean and unit variance.
2. Randomly initialising the model's weights, passing input data through the model to generate a prediction, computing the gradients of the loss with respect to the weights using back propagation and updating the weights based on the adaptive learning rates computed by the Adam optimiser, which involves calculating the first and second moments of the gradients. The model is trained for a specified number of epochs, which refers to the number of times the model is trained on the entire training data set.
3. The loss function used for training the above code is the mean squared error (MSE) function, which calculates the average of the squared differences between the predicted and actual outputs. This function is widely used in regression problems and is known to be effective in minimising the difference between predicted and actual outputs. By minimising the MSE loss during training, the model can learn to make more accurate predictions on new data.

6 Performance evaluation

The computational requirements and scalability of the BIG-LSTM model were assessed on the Google Colab platform utilising a Python 3 environment. The experimentation was conducted on a system equipped with an NVIDIA T4 GPU and 12 GB of RAM. The model training process made use of the available GPU acceleration provided by the NVIDIA T4 GPU, enhancing computational efficiency for complex neural network architectures like BIG-LSTM. The presence of 12 GB of RAM facilitated efficient data handling during preprocessing, training and validation stages. The training time for the BIG-LSTM model was influenced by the network's architecture, the size of the data set and the complexity of the hyperparameters. On average, each epoch of training took approximately 1.02 minutes. It's worth noting that this time may vary depending on factors such as batch size, hyperparameter configuration and the convergence rate of the optimisation process. The BIG-LSTM model demonstrated promising scalability in the context of larger data sets and higher-dimensional input features. Although our analysis was performed on a limited scale due to the constraints of the experimental environment, the parallel processing capabilities of the NVIDIA T4 GPU showcased the potential to accommodate larger data sets efficiently. Additionally, the model exhibited adaptability to higher-dimensional input features, which is indicative of its capability to capture complex relationships within richer data representations. This section presents a comparative analysis of the thermal error prediction BIG-LSTM model developed in this paper and four existing neural network models for thermal error prediction, namely LSTM, RNN, GRU and BILSTM. The performance of each model was evaluated using a test set of data, and four widely used evaluation metrics in machine learning for regression problems were employed: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R^2-squared (R^2) and Mean Squared Error (MSE).

RMSE represents the square root of the sample average of the squared differences between the predicted value and the actual value, and is commonly used as a measure of the sample standard deviation of the prediction error. A smaller RMSE value indicates better prediction performance. MAE, on the other hand, measures the average value of the absolute difference between the predicted value and the actual value, providing a more intuitive understanding of the error in prediction. A smaller MAE indicates smaller errors.

R^2 is a measure of the goodness of fit, which ranges from 0 to 1, with a higher value indicating a stronger relationship between the variables and a better model fit. MSE is the expected value of the squared difference between the predicted and actual values, and a smaller MSE value indicates better accuracy of the predictions. The mathematical expressions for each metric are given by equations (16), (17) and (18).

- *Root mean squared error (RMSE):*

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{16}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value and n is the total number of data points. RMSE is a measure of the deviation between the predicted and actual values, with a smaller value indicating better performance.

- *Mean absolute error (MAE):*

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\, y_i - \hat{y}_i \,| \tag{17}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the total number of data points. MAE is a measure of the average magnitude of the errors, with a smaller value indicating smaller errors.

- *Coefficient of determination (R^2):*

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{18}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the actual values and n is the total number of data points. R^2 is a measure of the goodness of fit, with a value of 1 indicating a perfect fit and a value of 0 indicating no correlation between the predicted and actual values.

According to the prediction results shown in Figure 7, the performance of the proposed BIG-LSTM model was evaluated and compared with four other models in terms of three evaluation indices, as shown in Table 1.

Table 1 Comparison of model performance

Models	*RMSE (μm)*	*MAE (μm)*	*R2 Score*
GRU	1.915464	1.342984	0.979164
LSTM	1.777063	1.289625	0.982066
RNN	3.028327	2.191802	0.947921
Bi-LSTM	1.259673	0.940884	0.990989
BIG-LSTM	0.91749	0.628139	0.99522

The results shown in 1 indicate that both LSTM-based models (LSTM and BIG-LSTM) outperform the others in terms of accuracy, with BIG-LSTM being the most accurate. The RNN model lags behind in accuracy, while the GRU model also performs well but not as accurately as the LSTM-based models. These results suggest that for this specific application, LSTM-based models are the most suitable choices, especially BIG-LSTM, which provides the highest predictive accuracy. Figure8 indicates that the BIG-LSTM model outperforms the other four models in all three-evaluation metrics, suggesting a higher accuracy in predicting thermal error. Specifically, the BIG-LSTM model achieves the best performance in terms of both RMSE and R^2 score. RMSE, which measures the average difference between the predicted and actual values, is lower in the BIG-LSTM model, indicating a better model fit. The R^2-score, a statistical measure of the model's goodness of fit, approaches 1 in the BIG-LSTM model, indicating a more accurate fit to the data. Therefore, the superior performance of the BIG-LSTM model in both RMSE and R^2-score suggests that it is a more reliable and accurate model for predicting thermal error.

Figure 7 Comparison of predicted and actual thermal errors using different models (see online version for colours)

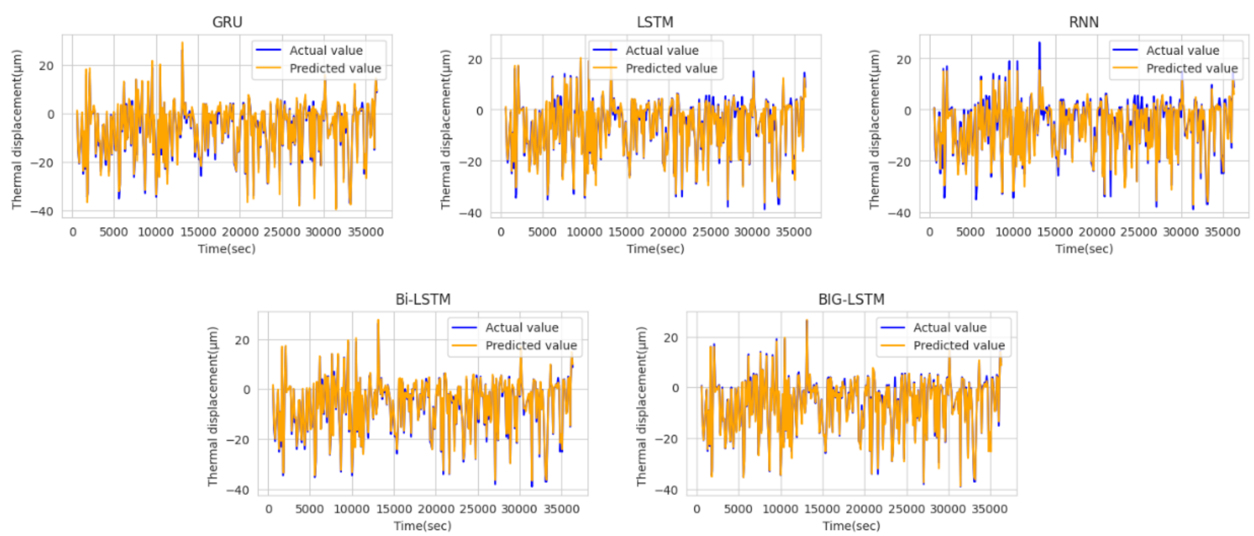
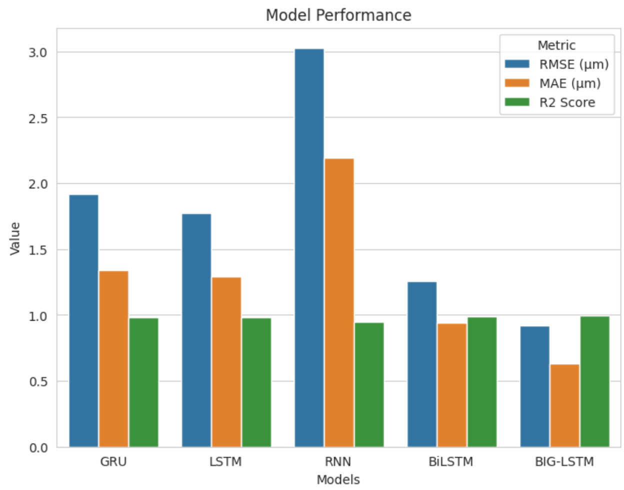


Figure 8 Comparison of model performance (see online version for colours)



A comprehensive ablation study was conducted to assess the significance of each constituent within the hybrid model, namely LSTM, GRU and Bi-LSTM, yielding valuable insights into the collective impact of the architecture on performance. This systematic analysis involved the controlled removal or modification of distinct components, providing a profound understanding of their individual contributions. The methodology comprised several crucial steps. Starting with the baseline hybrid model, which integrated all three components, each component was systematically eliminated or modified in isolation. This process allowed for the precise observation of each component's effect on the model's overall performance. The performance of each modified model was assessed using relevant evaluation metrics, such as Mean Squared Error (MSE) for regression tasks, or accuracy and F1-score for classification tasks. These metrics quantitatively gauged the impact of component changes on the model's predictive capabilities. The findings of the ablation study revealed pivotal insights. As observed, the removal of the LSTM component resulted in a noticeable degradation of long-range dependency capture. Similarly, the absence of GRU led to a diminished ability to capture short-term patterns, while omitting Bi-LSTM reduced the model's bidirectional sequence understanding prowess.

6.1 Real-time thermal error monitoring and control with the proposed model

The proposed model is designed to understand and predict a machine's displacement behaviour, particularly in situations where a cutting tool interacts with a workpiece. This interaction often generates heat due to friction and other mechanical processes. Consequently, the machine's internal temperature undergoes variations. During a cutting process, these temperature fluctuations, especially at specific crucial temperature points, significantly influence the machine's displacement behaviour. The model leverages these temperature changes to make predictions. Heat is generated throughout the machine as a result of the cutting operation. These specific points can be selectively identified as 'key temperature points', representing areas where temperature variations most accurately reflect the forces and stresses to which the machine is exposed during cutting. Once trained, the model uses the temperature data collected at these key locations to forecast how a particular machine will respond to a specific cutting procedure. The model's capacity to generalise its predictions to similar machines of the same type is crucial. In other words, if you have multiple machines of the same type, you can use the same model to predict displacement behaviour for all of them without the need for retraining. However, if you intend to use the model with a different type of machine, it must be retrained using data specific to that target machine. This necessity arises because various machine types may possess distinct mechanical characteristics, material compositions and structural arrangements. These precise nuances of each machine must be imparted to the model. The proposed model capitalises on the correlation between temperature changes at critical locations and a machine's displacement behaviour during cutting processes. It offers the advantage of transferability among machines of the same type, yet customisation becomes essential when working with different machine types due to their unique characteristics. In addition to the implementation methods discussed earlier, a crucial aspect of utilising the proposed model in CNC machines is its potential for real-time monitoring and control of thermal errors. This innovative application holds significant promise in industrial settings, offering both feasibility and several notable benefits. However, it also presents its own set of challenges that need to be addressed.

6.1.1 Challenges

While the concept is promising, several challenges must be considered when implementing real-time thermal error monitoring and control. First and foremost, ensuring the accuracy and reliability of temperature measurements is critical. Temperature sensors integrated into the CNC framework must provide precise and timely data. In cases where certain temperature points are not readily available within the CNC framework, external temperature monitoring equipment may be necessary. Managing the integration of these external sensors and ensuring their compatibility with the CNC system can be a complex task.

Additionally, the computational demands of running the model in real-time should not be underestimated. CNC machines require rapid decision-making and precise control. Therefore, optimising the model's execution to minimise latency is essential. Balancing the computational load with the CNC machine's processing capabilities can be a significant technical challenge.

6.1.2 Benefits

The benefits of implementing real-time thermal error monitoring and control using the proposed model are substantial. By continuously monitoring temperature variations and predicting their effects on machine displacement, manufacturers can achieve several advantages:

1 *Enhanced precision*: Real-time monitoring allows for immediate adjustments to compensate for thermal errors, resulting in higher machining precision and product quality.

2 *Reduced waste*: Minimising errors due to temperature fluctuations reduces material wastage and the need for rework, resulting in cost savings.

3 *Extended tool life*: Accurate control of thermal errors can extend the lifespan of cutting tools, reducing maintenance costs.

4 *Increased efficiency*: CNC machines can operate at optimal conditions, reducing downtimes and increasing productivity.

Although there are challenges to overcome, the implementation of real-time thermal error monitoring and control using the proposed model provides significant benefits to industrial CNC machining. Manufacturers willing to invest in advanced temperature monitoring and control systems can expect improved precision, efficiency and cost-effectiveness in their operations.

7 Conclusions

The proposed thermal error modelling method utilises a hybrid approach that employs BIG-LSTM deep learning. The model is designed to capture both short-term and long-term information in the temperature data and thermal error data by combining three GRU layers, two LSTM layers and two bi-LSTM layers in the hidden layer. GRU, LSTM and bi-LSTM are Recurrent Neural Networks (RNNs) that are commonly used for time-series data analysis. RNNs are suitable for sequential data, where each data point is related to the previous data points in the sequence. GRU and LSTM are two types of RNNs that were designed to overcome the vanishing gradient problem that can occur with traditional RNNs. Bi-LSTM is a variation of LSTM that can capture information from both past and future data points in a sequence. By combining these three types of layers in the hidden layer, the BIG-LSTM model can effectively capture both short-term and long-term information, allowing it to extract the temporal and spatial characteristics of the dynamic time series temperature data and thermal error data. This results in a superior learning ability of the model, which is reflected in its accuracy in predicting thermal error. During validation tests, the proposed BIG-LSTM model was compared with four traditional models based on RNN, GRU, LSTM and Bi-LSTM. The results showed that the BIG-LSTM model outperformed all four models in all evaluation criteria, including RMSE and R^2-score prediction accuracy. This indicates that the BIG-LSTM model is a reliable and accurate model for thermal error prediction. Specifically, the RMSE value of the BIG-LSTM model was only 0.9 μm, which is a very small error and demonstrates the high accuracy of the model.

Possible future directions for this research could include the exploration of different architectures and hyperparameters for the BIG-LSTM model to improve its performance further. Additionally, incorporating other relevant features such as tool wear and cutting parameters could enhance the accuracy of thermal error predictions. It may also be useful to investigate the transferability of the proposed model to other manufacturing processes, materials and cutting tools. Finally, the implementation of real-time monitoring and control of thermal error using the proposed model could be a valuable application in industrial settings.

Acknowledgement

All authors are thankful to M.S. Ramaiah University of Applied Sciences, for support to conduct study in this paper

References

Chen, C-C. and Hung, W-C. (2021) 'Thermal error modeling of CNC machine tool spindle based on multiple regression and features selection', *Proceedings of the IEEE 3rd Eurasia Conference on IOT, Communication and Engineering (ECICE)*, IEEE, pp.583–587.

Cheng, Y., Zhang, X., Zhang, G., Jiang, W. and Li, B. (2022) 'Thermal deformation analysis and compensation of the direct-drive five-axis CNC milling head', *Journal of Mechanical Science and Technology*, Vol. 36, No. 9, pp.4681–4694.

Haitao, Z., Jianguo, Y. and Jinhua, S. (2007) 'Simulation of thermal behavior of a CNC machine tool spindle', *International Journal of Machine Tools and Manufacture*, Vol. 47, No. 6, pp.1003–1010.

Hossain, I., Ahmed, Y.E. and Sharma, S.K. (2023) 'The use of both metallic-insert and tio2/water nanofluid to improve the performance of photovoltaic modules; experimental study', *Thermal Science and Engineering Progress*. Doi: 10.1016/j.tsep.2023.101944.

Li, B., Tian, X. and Zhang, M. (2019) 'Thermal error modeling of machine tool spindle based on the improved algorithm optimized bp neural network', *The International Journal of Advanced Manufacturing Technology*, Vol. 105, pp.1497–1505.

Li, Y., Zhao, W., Lan, S., Ni, J., Wu, W. and Lu, B. (2015) 'A review on spindle thermal error compensation in machine tools', *International Journal of Machine Tools and Manufacture*, Vol. 95, pp.20–38.

Li, Z., Zhu, B., Dai, Y., Zhu, W., Wang, Q. and Wang, B. (2021) 'Research on thermal error modeling of motorized spindle based on bp neural network optimized by beetle antennae search algorithm', *Machines*, Vol. 9, No. 11. Doi: 10.3390/machines9110286.

Liu, J., Ma, C., Gui, H. and Wang, S. (2022) 'Transfer learning-based thermal error prediction and control with deep residual LSTM network', *Knowledge-Based Systems*, Vol. 237. Doi: 10.1016/j.knosys.2021.107704.

Mayr, J., Jedrzejewski, J., Uhlmann, E., Donmez, M.A., Knapp, W., Härtig, F., Wendt, K., Moriwaki, T., Shore, P. and Schmitt, R. et al. (2012) 'Thermal issues in machine tools', *CIRP Annals*, Vol. 61, No. 2, pp.771–791.

Ramesh, R., Mannan, M.A. and Poo, AN. (2000) 'Error compensation in machine tools' a review: part i: geometric, cutting-force induced and fixture-dependent errors', *International Journal of Machine Tools and Manufacture*, Vol. 40, No. 9, pp.1235–1256.

Song, H., Park, J., Sharma, S.K. and Lee, J. (2022) 'A comparative study on ultrasonic propagation characteristics and defect detection of metal material additive manufacturing using deep learning algorithm', *Biennial International Conference on Future Learning Aspects of Mechanical Engineering*, Springer, pp.157–168.

Tan, F., Yin, G., Zheng, K. and Wang, X. (2021) 'Thermal error prediction of machine tool spindle using segment fusion LSSVM', *The International Journal of Advanced Manufacturing Technology*, Vol. 116, Nos. 1/2, pp.99–114.

Tanabe, I., Takada, K. and Tsutsumi, M. (1986) 'Thermal deformation of machine tool structures using epoxy resin concrete', *Proceedings of the Twenty-Sixth International Machine Tool Design and Research Conference: held in Manchester 17th–18th September 1986*, Springer, pp.245–252.

Wei, X., Ye, H., Miao, E. and Pan, Q. (2022) 'Thermal error modeling and compensation based on Gaussian process regression for CNC machine tools', *Precision Engineering*, Vol. 77, pp.65–76.

Xu, J., Guo, L., Jiang, J., Ge, B. and Li, M. (2019) 'A deep learning methodology for automatic extraction and discovery of technical intelligence', *Technological Forecasting and Social Change*, Vol. 146, pp.339–351.

Ye, W-H., Guo, Y-X., Zhou, H-F., Liang, R-J. and Chen, W-F. (2020) 'Thermal error regression modeling of the real-time deformation coefficient of the moving shaft of a gantry milling machine', *Advances in Manufacturing*, Vol. 8, pp.119–132.

Yin, Q., Tan, F., Chen, H. and Yin, G. (2019) 'Spindle thermal error modeling based on selective ensemble bp neural networks', *The International Journal of Advanced Manufacturing Technology*, Vol. 101, pp.1699–1713.

Zhang, H., Zhao, W., Du, C., Liu, H. and Zhang, J. (2016) 'Dynamic modeling and analysis for gantry-type machine tools considering the effect of axis coupling force on the slider–guide joints' stiffness', *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, Vol. 230, No. 11, pp.2036–2046.

Zhu, M., Yang, Y., Feng, X., Du, Z. and Yang, J. (2022) 'Robust modeling method for thermal error of CNC machine tools based on random forest algorithm', *Journal of Intelligent Manufacturing*, pp.1–14.

Zimmermann, N., Breu, M., Mayr, J. and Wegener, K. (2021) 'Autonomously triggered model updates for self-learning thermal error compensation', *CIRP Annals*, Vol. 70, No. 1, pp.431–434.

Zimmermann, N., Lang, S., Blaser, P. and Mayr, J. (2020) 'Adaptive input selection for thermal error compensation models', *CIRP Annals*, Vol. 69, No. 1, pp.485–488.