Thermal Error Prediction for Heavy-Duty CNC Machines Enabled by Long Short-Term Memory Networks and Fog-Cloud Architecture

Y.C. Liang[1], W.D. Li[1,2]†, P. Lou[3]†, J.M. Hu[3]

[1] Faculty of Engineering, Environment and Computing, Coventry University, CV1 5FB, U.K.

[2] School of Logistics Engineering, Wuhan University of Technology, 430070, China

[3] School of Information Engineering, Wuhan University of Technology, 430070, China

† Email: weidong.li@coventry.ac.uk, ping.lou@whut.edu.cn

Abstract

Heavy-duty CNC machines are important equipment in manufacturing large-scale and high-end products. During the machining processes, a significant amount of heat is generated to bring working temperatures rising, which leads to deformation of machine elements and further machining inaccuracy. In recent years, data-driven approaches for predicting thermal errors have been actively developed to adaptively compensate the errors on the fly to improve machining accuracy. However, it is challenging to adopting the approaches to support heavy-duty CNC machines due to their low efficiency in processing large-volume thermal data. To tackle the issue, this paper presents a new system for thermal error prediction on heavy-duty CNC machines enabled by a Long Short-Term Memory (LSTM) networks and a fog-cloud architecture. Innovative characteristics of the system include the following aspects: (1) data-based modelling is augmented with physics-based modelling to optimise the number/locations of thermal sensors deployed onto machine elements and minimise excessive data to facilitate computation; (2) a LSTM networks with a data pre-processor is developed for modelling thermal errors more effectively in terms of prediction accuracy and computational efficiency; (3) A fog-cloud architecture is designed to optimise the volume of transferred data and overcome low latency of the system. The system was validated using an industrial heavy-duty CNC machine. Practical case studies show that the system reduced the volume of transmitted data for 52.63% and improved the machining accuracy for 46.53%, in comparison with the processes without using the designed system.

Keywords: Thermal error prediction, Heavy-duty CNC machine, LSTM, Fog computing

1. Introduction

Heavy-duty CNC machines are highly demanded in some important applications for manufacturing large-scale and high-end products, such as steam turbines, large nuclear pumps, marine propellers, and large aircraft wings (Huang et al., 2015). For machining processes conducted on heavy-duty CNC machines, precision is an essential technical requirement. It is paramount to develop effective error prediction modelling and then compensation technologies to minimise machining errors. The thermal error is a principal one accounting for about 40% – 70% of the total machining errors (Bryan, 1990). It is caused by heat and temperature rising during machining processes, leading to deformation of machine elements and further machining inaccuracy (Blaser et al., 2017). There are mainly two types of heat sources generating thermal errors: (1) internal sources - heat is generated from cutting, friction and moving processes of machine elements; (2) external sources - heat is related to surrounding environmental temperature variations. To improve machining accuracy for heavy-duty CNC machines, it is prominent to develop an effective model for thermal error prediction, which can represent relationships between heat sources and the impact on thermal errors. Based on the model, an appropriate compensation strategy can be further developed to enhance the overall machining accuracy.

In the past, physics-based modelling has been actively investigated to predict thermal errors for CNC machine systems (Huang and Hoshi, 2001; Li et al., 2015). In the modelling process, physical mechanisms of machining elements are established. Finite Element Method (FEM), Finite Difference Method (FDM), or Finite Difference Element Method (FDEM), is then applied to establish thermal deformation fields of machine elements. However, due to the large-scale structures and dynamic operation environments of heavy-duty CNC machines, it is challenging to developing accurate physical mechanisms for analysing and predicting thermal errors preciously. In recent years, data-based modelling has been actively explored as an increasingly popular solution for thermal error prediction. By leveraging the latest deep learning technologies, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), thermal error related features can be mined from a large volume of thermal data collected from machine elements through deployed sensors. In comparison with physics-based modelling, data-based modelling is

more flexible and accurate in supporting the machining processes conducted through heavy-duty CNC machines.

However, data-based modelling could be severely hindered by the less efficient processes of collecting and handling large-volume thermal data. To mitigate the challenge, this paper presents an improved data-driven system for thermal error prediction on heavy-duty CNC machines. Innovative characteristics of the research are given below:

- A large number of thermal sensors is required to install into the large-scale structure of a heavy-duty machine in order to collect sufficient data for accurate thermal modelling. Nevertheless, sensors could be excessively deployed to generate redundant data if without an appropriate installation guidance provided. To optimise the number of deployed sensors and minimise excessive data, this research is innovative in integrating physics-based modelling to facilitate data-based modelling. That is, FEA is conducted based on the structure of a heavy-duty CNC machine and its heat sources. The results of FEA, which show the heat distributions over the entire structure of the heavy-duty machine, provide sensible instructions to minimise unnecessary sensors and data;
- Meanwhile, it is essential to design a more effective deep learning algorithm for predicting thermal errors on machines. The LSTM networks, which is an improved RNNs, is robust for time-series data processing and relatively insensitive to unknown gap lengths in the data. Based on the characteristic, in this research, a LSTM networks is designed to build the relationship between key heat sources and the thermal deformation fields of machine elements for thermal error prediction. This LSTM design is justified by benchmarking with different LSTM designs and other intelligent algorithms. Furthermore, to reduce data collinearity and computational workloads to better support the LSTM, an improved Grey Relational Analysis (*i*GRA) is developed to pre-process thermal data;
- A cloud architecture with associated computational resources on a cloud server is a popular solution to support data-based modelling. However, the efficiency of communicating monitored data could be seriously affected by the limited bandwidth and high latency of the industrial Internet. In this research, a fog-cloud architecture based on the "divide-and-conquer" strategy is designed to tackle the issue. In the new architecture, *i*GRA-based data pre-processing and LSTM-based prediction modelling are

conducted locally on a fog layer to expedite decision making. FEA and LSTM training, which are computationally intensive activities, are processed on a cloud layer to leverage its computational resources. Quantitative analyses are conducted to showcase the advantage of the fog-cloud design for thermal error prediction of heavy-duty machines.

In this paper, the developed system was validated using a heavy-duty CNC machine namely ZK5540, which was made by the Wuhan Heavy Duty Machine Tool Group Corporation. Industrial case studies demonstrated that, by adopting the system, the volume of transmitted data was minimised by 52.63% and the thermal errors were reduced by 46.53%, in comparison with processes without using the developed system. Based on the case studies and benchmarking analyses, the benefit of adopting the system in terms of data processing efficiency and machining accuracy improvement is clearly exhibited.

The rest of the paper is organised as follows: In Section 2, related works are reviewed. In Section 3, research methodologies and system design are discussed. In Section 4, system deployment and case studies are presented. In Section 5, conclusions are drawn and future research directions are outlined.

2. Related Work Review

2.1 Data-based modelling for thermal error prediction

According to the survey by Liu et al. (2018) and Li et al. (2019), data-based modelling for thermal error prediction has been becoming an increasingly popular approach. There are usually two main steps to build a data-based thermal model: (1) to optimise deployment points of sensors, and (2) to develop thermal error prediction models. Miao et al. (2015) developed a principal component regression algorithm to remove data points with collinearity relationships. Cheng et al. (2015) proposed a grey system theory to analyse the data similarity among temperature sensors. Based on the analysis, sensors mostly sensitive to heat were selected. Case studies demonstrated the temperature sensors were decreased from 24 to 7 after this analysis. Abdulshahed et al. (2015) proposed a method based on grey modelling and fuzzy c-means clustering to determine key temperature points on machines. 525 discrete spots were classified into 8 groups for further modelling and data minimisation. Liu et al. (2015) designed an optimal selection method of temperature-sensitive measuring points. In the method, the degree of temperature sensitivity was defined and used to select measuring points with high sensitivity to thermal error. The selected points were then classified with

fuzzy clustering and grey correlation grades, and temperature-sensitive measuring points were selected based on the analysis of temperature sensor locations. Results showed that the number of the measuring points was reduced from 27 to 5 by using the method. Abdulshahed et al. (2016) utilised grey system models and a grey neural network model to predict the thermal error of a gantry-type 5-axis machine tool. The grey system models can transform original data to a monotonic series of data, so that the randomness of data can be reduced. A particle swarm optimisation algorithm was designed to optimise the parameters of the grey neural network model. The method was tested and 85% thermal error reduction after compensation was achieved. Ma et al. (2016) developed the genetic algorithm and particle swarm optimisation algorithm to optimise the parameters of an artificial neural networks to build up a thermal error prediction model. The model was tested to show effectiveness in the aspects of accuracy, convergence performance and robustness. Li et al. (2017) devised a clustering method to select the most suitable sensor points for data analysis. Li et al. (2019) developed a thermal error prediction model based on a hybrid particle swarm optimisation and artificial neural network. Temperature measurement points were clustered by a SOM neural network, and an analysis was conducted to explore the correlation between the thermal sensitive points and the thermal error. Fujishima et al. (2019) developed a novel thermal displacement compensation method using a deep learning algorithm. In the algorithm, reliability of thermal displacement prediction was evaluated and compensation weights were adjusted adaptively. Yao et al. (2020) designed an optimal composite model (OM) for modelling spindle thermal error prediction. The grey model and the least squares support vector machine (LS-SVM) were used to establish the thermal error prediction model. Then, the OM model was used to adjust the weighting coefficients of LS-SVM to fine-tune the prediction model based on practical thermal error data. The above works are also summarised in Table 1.

Table 1: A brief summary of some research works of data-based modelling for thermal error prediction.

Research	**Functions**	**Intelligent algorithms**
Miao et al. (2015)	To remove data points with collinearity relationship	Principal component regression algorithm
Cheng et al. (2015)	To remove data points by analysing data similarity among temperature sensors	Grey system theory
Abdulshahed et al. (2015)	To determine key temperature points	Grey modelling and fuzzy c-means clustering
Liu et al. (2015)	To remove data points by analysing data similarity among temperature sensors	Fuzzy clustering and grey correlation grade
Abdulshahed et al. (2016)	Transform original data to a monotonic series of data and predict thermal error	Grey system models and Grey Neural Network Model

Ma et al. (2016)	Thermal error prediction modelling	Genetic algorithm, particle swarm optimisation, artificial neural networks
Li et al. (2017)	To select the most suitable sensor points for data analysis	Fuzzy clustering
Li et al. (2019)	Thermal error prediction modelling	Hybrid particle swarm optimisation and artificial neural networks
Fujishima et al. (2019)	Thermal displacement compensation	A deep learning algorithm
Yao et al. (2020)	Thermal error prediction modelling	Grey model and the least squares support vector machine, composite model

The aforementioned methods for sensor point selection are based on correlation analysis of collected data to reduce redundant sensors. On the other hand, for heavy-duty CNC machines, it is challenging to deciding initially deployed sensor points to create a suitable initial dataset for further optimisation. Trials are usually expensive and therefore a guidance on initially deployed points is required. Nevertheless, limited research was conducted in this area.

2.2 LSTM networks for data-based manufacturing applications

Owing to the advantage of effectively solving the common problem of "vanishing/exploding gradients" in deep learning and revealing patterns for time-series data, the LSTM networks for data-based manufacturing applications has been widely researched (Qu et al., 2018; Chen et al., 2020). Zhang et al. (2018) designed a LSTM networks for machine remaining life prediction. Li et al. (2018) proposed a Quantum weighted LSTM networks to predict degradation trend of rotating machinery. In the research, the minimum prediction error using the LSTM networks proved to be 1.54% while other comparative algorithms such as RNNs and Support Vector Machine (SVM) were more than 1.90% in prediction error. Yang et al. (2019) developed a LSTM networks to detect and isolate faults for electro-mechanical actuators. The method achieved accuracy for more than 99% and good robustness, while other algorithm such as SVM was 94.4% in accuracy. An et al. (2020) designed data-driven modelling for remaining useful life prediction for machining cutters using a convolutional and stacked LSTM networks. The prediction accuracy of the proposed method was up to 90% in comparison with 86% accuracy using CNNs. Xia et al. (2020) proposed an ensemble framework based on convolutional bi-directional LSTM networks to predict remaining useful life of machine tools, and the proposed method was tested with multiple datasets to demonstrate the accurate and robust prediction performance of the designed LSTM. The Root Mean Square Error (RMSE) can

achieve 12.66, and other algorithms such as Deep Neural Networks (DNNs) and CNNs can achieve 19.05 and 15.50, respectively. Liu and Zhu (2020) developed a two-stage approach for predicting the remaining useful life of machine tools utilising a bidirectional LSTM network. The developed method was compared with different algorithms, such as Support Vector Machine (SVM) and decision tree, for benchmarking. The proposed method achieved 4.18% in the Mean Absolute Percentage Error (MAPE), which was the lowest compared to 7.16%, 6.29 and 7.08% achieved by using the decision tree, SVM and boosted tree respectively.

The above research works have proved that LSTM exhibits good performance in comparison with other comparative algorithms especially in processing time-series data. It is therefore worth exploring how to design an appropriate LSTM networks to facilitate heavy-duty machine applications. Meanwhile, the data collinearity issue for input would degrade the performance of LSTM, thereby inspiring design of a sensible strategy to further improve the performance of the LSTM networks for manufacturing applications.

2.3 Information architecture for data-based modelling

For data-based modelling to support heavy-duty CNC machines, thermal data are usually collected and transmitted to a cloud server for training and applying intelligent algorithms. However, based on the limited bandwidth of the industrial Internet in manufacturing companies, data transmission could be severely congested due to the heavy data traffic from sensors deployed on machine elements to a cloud server. To expedite system efficiency, a fog (From cOre to edGe) computing model has been actively to adopt a "divide-and-conquer" strategy in avoiding low latency of data transmission (Hu et al., 2017; Tao et al., 2018; Bellavista et al., 2019). A fog model consists of gateways with certain computing capacities to process monitored data. Only data that are necessary for central processing are transmitted to a cloud server. That is, the most computationally intensive jobs are undertaken on cloud while the fog model is responsible for local processing. Thus, the overall data transmission and computational performance can be significantly enhanced in a fog architecture (Liang et al., 2019; Xu et al., 2020). Wu et al. (2017) proposed a fog computing system to achieve process monitoring and prognosis for manufacturing company. The system is capable of collecting and processing large-scale data in real time. A machine learning algorithm was implemented to predict tool wear through CNC machining operations. Sood and Mahajan (2018)

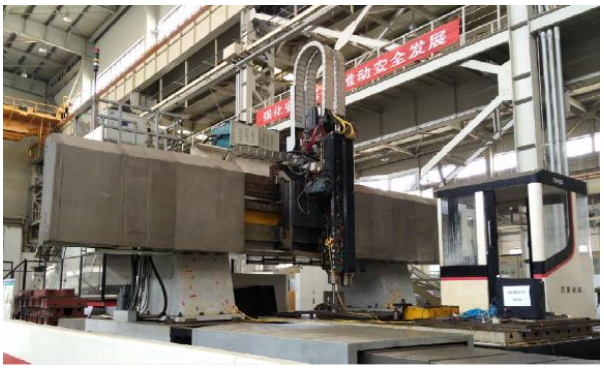
proposed a fog system for the healthcare sector to distinguish, detect and prevent mosquito borne diseases (MBDs). The system can assimilate, analyse and share medical information among users and healthcare service providers.

For thermal error prediction on heavy-duty machines, there is no reported work yet to leverage the research progress of a fog-cloud architecture for improving data transmission and computing efficiency. Thus, in this research, investigations will be conducted on how to design an appropriate fog-cloud architecture to optimise this application.

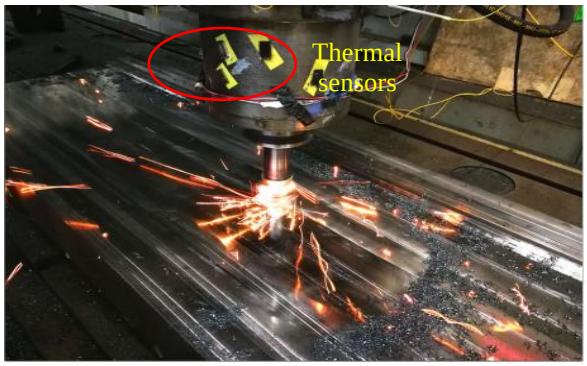
3. Research Methodologies and System Design

3.1 FEA analysis for optimised deployment of sensors

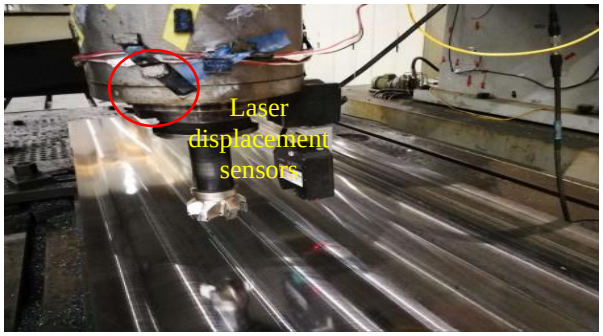
In this research, a heavy-duty CNC machine, namely ZK5540, is used as an example to explain and verify the developed methodologies. In the machine, thermal sensors and laser displacement sensors are installed to collect thermal data and spindle deformation data respectively. The collected data are used to train a LSTM networks designed in this research to predict thermal errors for adaptive compensation on the fly. The ZK5540 machine, deployed sensors, and data processing system are shown in Fig. 1.



(a) The heavy-duty CNC machine ZK5540



(b) Sensors and machining process



(c) Sensors and machining process

(d) Data processing system

Fig. 1: The heavy-duty CNC machining system and sensor deployment.

To provide a guidance on how to place thermal sensors more effectively, an FEA model is designed to analyse the thermal distribution of the machine. To build up the FEA model, the main internal heat sources of the machine are identified as thermal loads. They are primarily from: (1) the current of the stator and rotor coils of the spindle motor and the servo motor, (2) the friction of the bearings of the spindle motor and the servo motor, and (3) the friction of the guide rails. To calculate heat generated during the machining process, these aspects are formulated in detail as follows.

Heat is generated due to the current in the stator and rotor coils of the spindle motor and servo motor (Ward, 2014):

$$P_e = I \cdot R^2 \tag{1}$$

where P_e is the thermal power due to the current; I is the current flowing through the stator and rotor coils; R is the resistance of the stator and rotor coils.

Meanwhile, heat is generated from the friction of the spindle and servo bearings of the spindle motor and servo motor (Zhao et al., 2007):

$$P_f = 0.0001047 \cdot n \cdot M \tag{2}$$

where P_f is the thermal power generated due to the friction of the spindle bearing; n is the rotational speed of the spindle; M is the total friction torque.

At the same time, heat is also generated because of the friction on the guide rails (Halliday et al., 2014):

$$P_m = \mu_f \cdot N \cdot V \tag{3}$$

where P_m is the thermal power due to friction on the guide rail; μ_f is the friction coefficient; N is the normal force; V is the velocity of translational travelling.

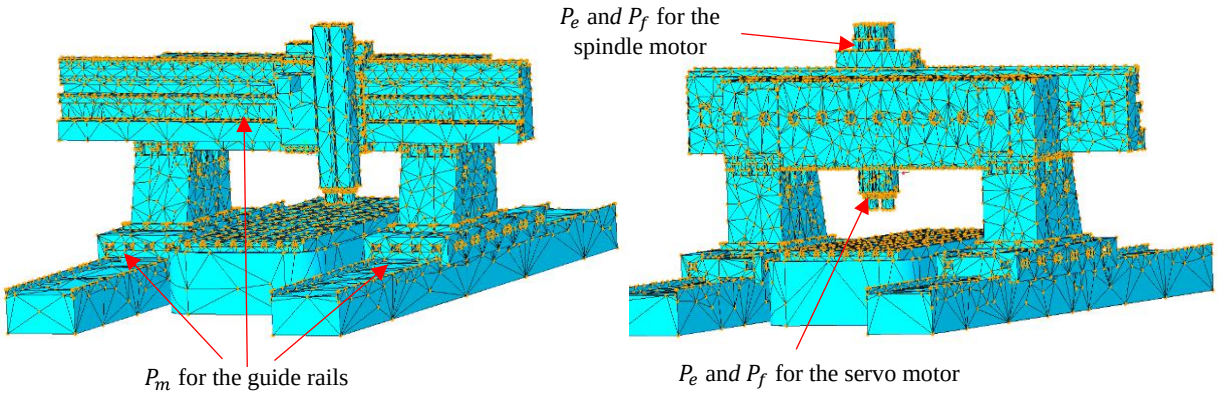
Table 1 and Table 2 show important properties and parameters of the ZK5540 machine. To build its FEA model, meshing and material properties of the machine are applied. Meanwhile, thermal loads defined as above are set. Fig. 2 shows the meshing result of the machine, in which the thermal loads are marked. Table 3 includes detailed meshing information.

Table 1: Properties of the heavy-duty machine.

Property	Main material	Working table width	Working table length	Travelling range for axis X	Travelling range for axis Y	Travelling range for axis Z
Value	Steel	2000mm	4500mm	5000mm	4000mm	650mm

Table 2: Important parameters for the heavy-duty machine.

Property	Spindle speed range	Spindle motor power	Feed rate X axis	Feed rate Y axis	Feed rate Z axis
Value	10-6000 r/min	31Kw	0-10,000 mm/min	0-10,000 mm/min	0-5,000 mm/min



(a) The front view of meshing (b) The back view of meshing

Fig. 2: The technical details and meshing of the heavy-duty machine.

Table 3: Meshing properties of the heavy-duty machine.

Property	Tetra quantity	Edge quantity	Face quantity	Min Edge Angle (degrees)	Max Edge Angle (degrees)	Max Aspect Ratio
Value	87329	139752	199783	0	173.98	12.51

Fig. 3 shows the heat distribution of ZK5540 analysed by FEA, which is used to identify appropriate positions to deploy sensors (some deployments based on the FEA result are shown in Fig. 4). According to the FEA result and the monitoring areas on the machine, the following observations are made:

- Temperature primarily increases around the spindle at about 80 °C the highest. The main reason is that heat is primarily generated there due to the current and friction of motors. Heat is difficult for dissipation due to the limited area for conduction. Therefore, majority of the temperature sensors, i.e., 52 thermal sensors, are installed all-round the spindle.
- Temperature also increases to approximately 35 °C around the columns due to the current and friction of motors. The temperature is not as high as the spindle because it has a much greater area for heat dissipation. Thus, 32 sensors are installed on columns (16 sensors on each column).
- Heat could be generated around the slide-ways and beam due to friction. Similarly, the temperature is not as high as that of the spindle. 32 and 8 sensors are installed on the slide-ways and beams, respectively.

- To detect the environment temperature, 4 sensors are installed on the edges of the slide-ways. The temperature difference between the environment temperature and temperature on machining components, including the spindle, columns, slide-ways and beam, can be calculated.

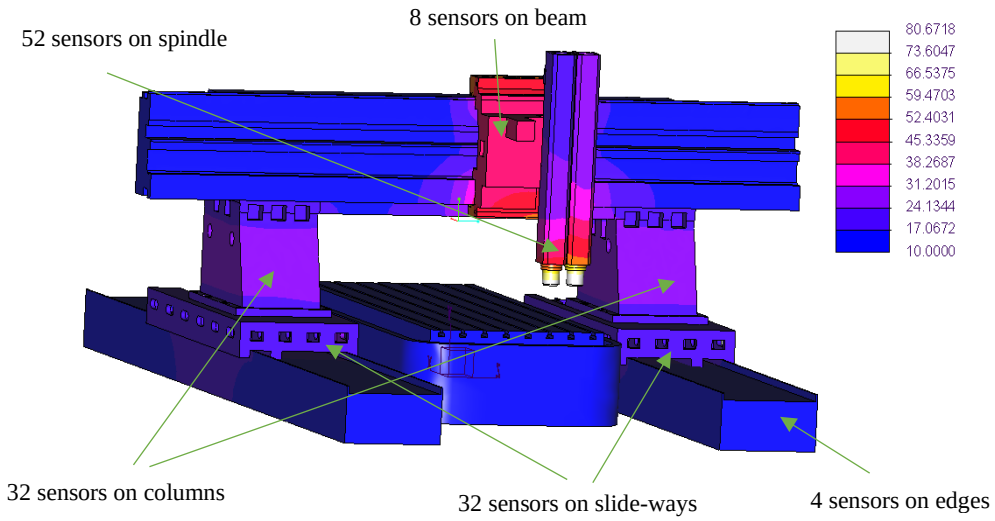


Fig. 3: The heat distribution analysed by FEA to guide the deployment positions for sensors.

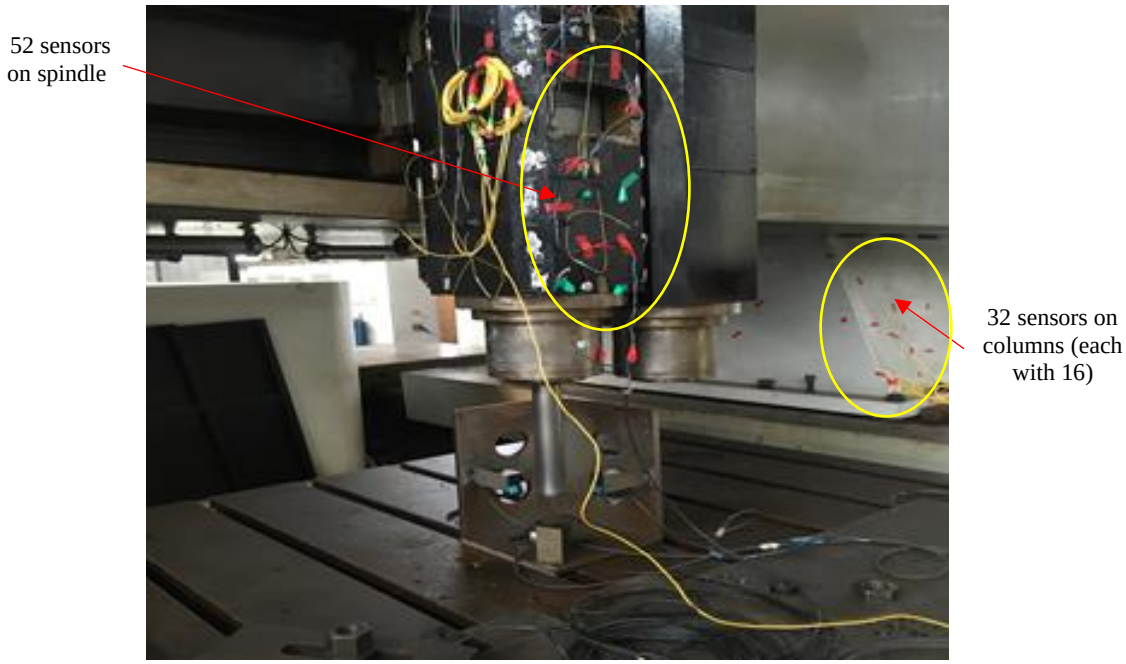


Fig. 4: Illustration of some deployed sensors into the machine.

3.2 Improved Grey Relational Analysis (*i*GRA)

To reduce data collinearity and processing workloads, collected thermal data are expected to be pre-processed. Grey Relational Analysis (GRA) is a useful approach to identify relationship coefficients among sensor nodes in order to remove duplication for data quality improvement (Kuo et al., 2008; Kumar and

Singh, 2020). However, the GRA process might be not adaptive in practical applications considering that it is difficult to identify a reference sensor, which refers to a sensor that can reflects the temperature distribution with the highest sensitivity. To tackle the issue, an improved GRA, i.e., *i*GRA, is developed in this research. Some critical steps for *i*GRA are depicted below.

1. Normalisation: t_{ij} represents a temperature measured by sensor node i at time j. All the measured temperatures are grouped as a set $T = \{t_{ij}\}$ ($i = 1,2,\dots,n$; $j = 1,2,\dots,m$), where n is the maximum sensor node index and m is the maximum time index. The temperature t_{ij} is normalised to t'_{ij} to facilitate further processing. The normalisation process is below:

$$t'_{ij} = \frac{t_{ij} - Min(T)}{Max(T) - Min(T)} \tag{4}$$

where $Max(T)$ and $Min(T)$ represent the maximum and minimum temperatures in the set T.

2. Calculation of the grey relational coefficient (including improvements over the GRA algorithm): The grey relational coefficient $g(t'_{rj}, t'_{ij})$ can determine how close the temperature of a sensor t'_{ij} is to the temperature of the reference sensor t'_{rj}. Its computation is below:

$$g(t'_{rj}, t'_{ij}) = \frac{\Delta_{min} + \mu \cdot \Delta_{max}}{\Delta_{ij} + \mu \cdot \Delta_{max}} \tag{5}$$

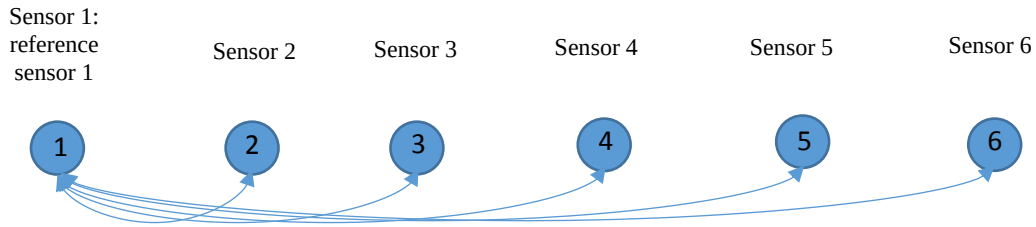
where $\Delta_{ij} = |t'_{rj} - t'_{ij}|$; $\Delta_{min} = Min(\Delta_{ij})$; $\Delta_{max} = Max(\Delta_{ij})$; μ , which is within 0 and 1, is a distinguishing coefficient.

μ is designed to expand or compress the range of the grey relational coefficient. In this research, to improve the performance of grey relational analysis, μ is set 0.5 by default (Kuo et al., 2008).

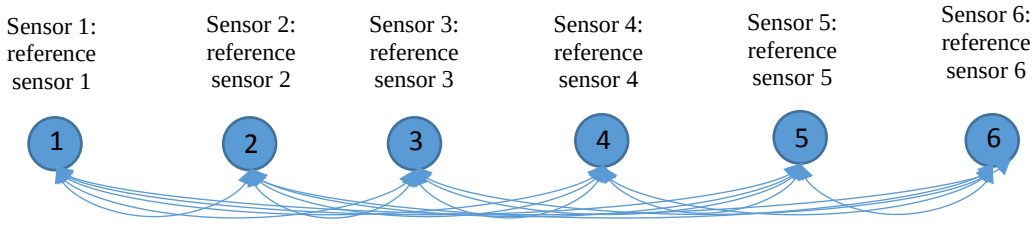
In the GRA algorithm, only one reference sensor is set to be compared with other sensors for their temperatures. However, it is laborious to identify which sensor will be appropriate to be such a reference. Therefore, robustness is improved in the *i*GRA algorithm. That is, there will not be just one reference, and all the sensor data are selected as references for comparisons. Meanwhile, the following is designed for computational minimisation in the *i*GRA algorithm:

As aforementioned, instead of setting up a fixed reference sensor, each sensor data is set as a reference along with the computation process. All the other data sequences are compared with the standard sequence according to the total grey relational grade $g(t'_{rj}, t'_{ij})$ based on Equations (5). For example, t'_{rj} is set as a

standard data sequence and compared with t'_{ij}. The same result will be computed again when t'_{ij} is set as the standard data sequence and compared with t'_{rj}. To minimise the redundant computation, the same pair of data sequences will only be computed once. Fig. 5 shows an example to illustrate the difference between the GRA and *i*GRA algorithms, in which each curved double arrow connector represents a grey relational coefficient between two sensors.



(a) Example of the GRA algorithm with 6 sensors.



(b) Example of the *i*GRA algorithm with 6 sensors.

Fig. 5: Example of illustrating the GRA and *i*GRA algorithms using 6 sensors.

Therefore, the total number of grey relational coefficients to be calculated is:

$$N = \frac{n \times (n-1)}{2} \tag{6}$$

where N is total amount of grey relational coefficients to be calculated; n is the number of sensors.

3. Calculation of the total grey relational grade: The total grey relational grade $G(T'_r, T'_i)$ is calculated based on $g(t'_{rj}, t'_{ij})$ according to the following formula:

$$G(T'_r, T'_i) = \sum_{j=1}^{m} (w_j \cdot g(t'_{rj}, t'_{ij})) \tag{7}$$

where w_j is a weight decided by a user; $\sum_{i=1}^{n} w_j = 1$.

4. Sensor selection: The above $G(T'_r, T'_i)$ is used to calculate the similarity between sensor nodes T'_r and T'_i. The higher $G(T'_r, T'_i)$, the closer T'_r and T'_i. If $G(T'_r, T'_i)$ is within a pre-defined threshold θ, T'_i can be

removed for minimising data volume and improving the training accuracy of the LSTM networks. That is, the threshold is decided when the following fitness achieves the highest value:

$$fitness = max\ (w_1 \cdot N_{data} + w_2 \cdot N_{accuracy}) \tag{8}$$

where N_{data} is the normalised saved data volume; $N_{accuracy}$ is the normalised training accuracy of the LSTM networks; w_1 and w_2 are the weights for the two objectives ($w_1 + w_2 = 1$); the normalisation process follows the same principle as Equation (4).

It is assumed that the number of sensor nodes is reduced from *n* to n' after this selection operation. The rest of data are grouped into a set $T'_{j'} = \left\{t'_{ij'}\right\}$ ($i = 1,2,\dots,n; j' = 1,2,\dots,m'$), which are further used for training and applying the LSTM networks for thermal error prediction.

3.3 LSTM design for thermal error prediction

Thermal data, which are pre-processed by *i*GRA, are fed into the designed LSTM networks to predict thermal errors. As the base for a LSTM networks, RNNs is a deep learning neural networks with a specially designed topology, through which the connections among data points within arrays can be built by hidden states. Information from previous data points generates impact on the following prediction, so that the prediction can be improved compared with traditional neural networks (Zhao et al., 2018; Vlachas et al., 2020). Owing to the capability of learning 'context' within data arrays, RNNs has been widely used in time-series data prediction, such as stock market forecasting, language modelling, machine degradation prediction, etc. Fig. 6 shows a typical structure of RNNs. To predict thermal errors by a RNNs, the vector in the hidden layer h_t and output vector E_t (thermal error at time t) can be calculated by the following equations (Cinar et al., 2018; Vlachas et al., 2020):

$$h_t = \sigma(W_h \cdot T'' + U_h \cdot h_{t-1} + b_h) \tag{9}$$

$$E_t = \sigma(W_E \cdot h_t + b_E) \tag{10}$$

where σ is a sigmoid activation function; t is the time step; T'' is an input vector at time t; h_{t-1} and h_t are hidden layer vectors at times *t*-1 and *t*, respectively; W_h and U_h are weight matrices for T'' and h_{t-1}, respectively; b_h is a bias matrix; E_t is an output vector; W_E and b_E are weight and bias vectors, respectively, to calculate E_t.

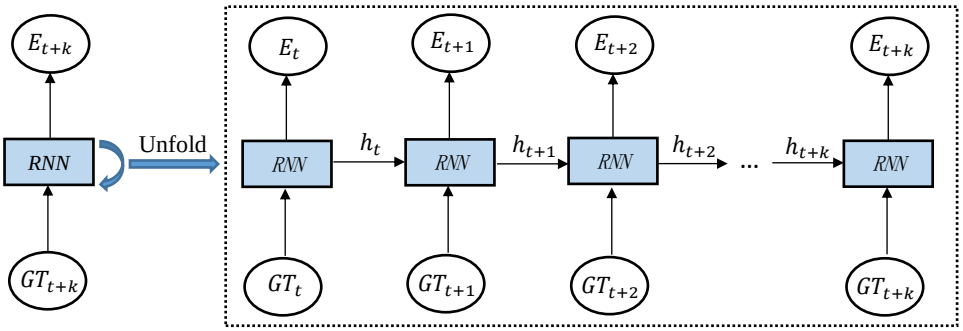


Fig. 6: The general structure of a RNNs.

However, since the hidden state can be extremely small or high with long steps of multiplication, prediction results of the RNNs are usually crippled by a vanishing or exploding gradients problem (Qu et al., 2018; Mohanty et al., 2020). To avoid extreme values through the multiplication process, a gating mechanism is required to control which information to be kept (i.e., memorising) and which information to be removed (i.e., forgetting). A LSTM networks is therefore developed to improve the RNNs design. In a LSTM networks, there are three gates, i.e., input gate, forget gate and output gate, to determine how much information can be fed forward to the next cell (Wu et al., 2018; Ellis and Chinde, 2020). A typical LSTM networks is shown in Fig. 7.

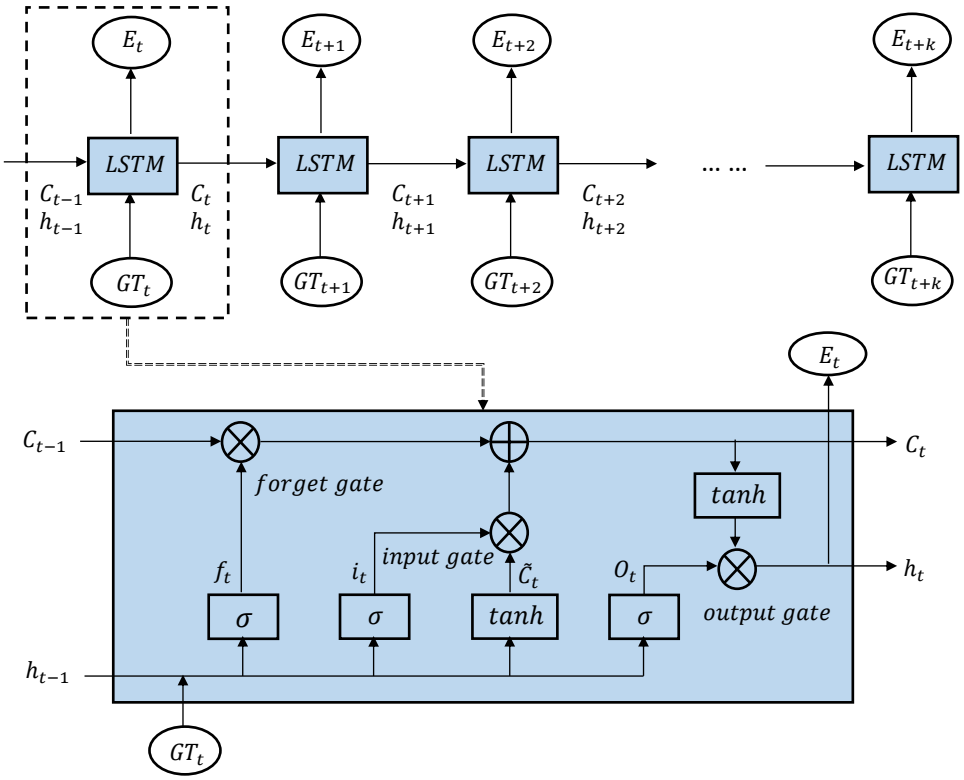


Fig. 7: The structure of a LSTM networks.

In the forget gate f_t, some information is removed by using a sigmoid activation function:

$$f_t = \sigma(W_{GTf} \cdot GT_t + W_{hf} \cdot h_{t-1} + b_f) \tag{11}$$

where σ is a sigmoid activation function; t is the time step; h_{t-1} is the short-term memory; W_{GTf} and W_{hf} are the weight matrices for GT_t and h_{t-1}, respectively; b_f is the bias in the forget gate.

In the input gate i_t, the information to input is selected by utilising sigmoid and tanh:

$$i_t = \sigma(W_{GTt} \cdot GT_t + W_{ht} \cdot h_{t-1} + b_t) \tag{12}$$

$$\tilde{C}_t = tanh(W_{GTc} \cdot GT_t + W_{hc} \cdot h_{t-1} + b_c) \tag{13}$$

where tan is the tanh; W_{GTt}, W_{ht}, W_{GTc} and W_{hc} are the weight matrices; b_t and b_c are the bias; $\tilde{C}_t$ is the output of the standard recurrent layer without the LSTM.

Based on C_{t-1}, $\tilde{C}_t$, f_t and i_t, the long-term memory C_t can be calculated:

$$C_t = f_t \otimes C_{t-1} \oplus i_t \otimes \tilde{C}_t \tag{14}$$

where $\otimes$ and $\oplus$ are element-wise multiplication and addition, respectively.

In the output gate O_t, the short-term memory h_t can be calculated below:

$$O_t = \sigma(W_{GTo} \cdot GT_t + W_{ho} \cdot h_{t-1} + b_o) \tag{15}$$

$$h_t = O_t \otimes tanh\,(C_t) \tag{16}$$

$$E_t = h_t \tag{17}$$

where W_{GTo} and W_{ho} are weight matrices; b_o is the bias; E_t is the predicted thermal deformation at time index *t*.

In this research, the designed LSTM networks has one layer. Its performance is benchmarked with different LSTM layers and the relevant results will be shown in Section 4.3. Table 4 shows the input and output data of the designed LSTM networks. The normalised temperature data $NT_1, NT_2 \ldots, NT_n$, which are grouped into GT_t, are fed into the LSTM networks as input. The output is the thermal error E_t.

Table 4: Input and output of the LSTM networks.

Input vector (GT_t)	**Output vector (E_t)**
Point 1 in normalised temperature data NT_1	Thermal error data on the *X* axis E_{t_X}
Point 2 in normalised temperature data NT_2	Thermal error data on the *Y* axis E_{t_Y}
… …	Thermal error data on the *Z* axis E_{t_Z}
Point *n'* in normalised temperature data $NT_{n'}$	

The LSTM networks will be trained and needs to be re-trained (calibration) after a period using historical data. Mean Square Error (MSE), which calculation is based on the difference between predicted thermal errors and practically measured thermal errors, is defined below:

$$MSE = \frac{\sqrt{(E_{t\_X} - P_{t\_X})^2 + (E_{t_Y} - P_{t\_Y})^2 + (E_{t_Z} - P_{t\_Z})^2}}{3} \tag{18}$$

where P_{t_X}, P_{t_Y} and P_{t_Z} are the practically measured thermal errors; E_{t_X}, E_{t_Y} and E_{t_Z} are the predicted thermal errors.

The LSTM networks will be re-trained by using an improved gradient descent optimiser (we improved the gradient descent optimiser in Step 4 with a decaying learning rate to improve the convergence speed). The optimisation process is explained as follows:

1. Define the weight W_{LSTM} and bias b_{LSTM} of the LSTM networks; initialise the W_{LSTM} and b_{LSTM} with random values;

2. Define the maximum epoch time *T*;

3. Optimise the value of W_{LSTM}, b_{LSTM} by minimising the loss function with gradient descent:

$${W_{LSTM}}^{t+1} = {W_{LSTM}}^{t} - \varepsilon \cdot (\frac{\partial(MSE)}{\partial(W_{LSTM})}) \tag{19}$$

$${b_{LSTM}}^{t+1} = {b_{LSTM}}^{t} - \varepsilon \cdot (\frac{\partial(MSE)}{\partial(b_{LSTM})}) \tag{20}$$

where t is the current epoch time; ε is the learning rate; ∂ is the gradient descent. Traditionally, the learning rate is a constant value.

4. To improve the optimisation process, the learning rate ε can be designed to decay over training epochs to approach optimal parameters:

$$\varepsilon = \varepsilon_0/(1 + k \cdot t) \tag{21}$$

where k is the hyper-parameter.

5. Repeat the above Steps 3 and 4 until reaching the maximum epoch time.

3.4 Fog-cloud architecture design

A cloud structure has been widely used to process computationally intensive tasks benefiting from its centralised storages and strong computational capacities on the cloud server side. However, the performance of the cloud solution could be severely compromised by the limited communication bandwidth

of the industrial Internet in manufacturing companies. Furthermore, data security in cloud is also a key concern for industrial users (Liang et al., 2019). To address the issues, a fog-cloud architecture, which consists of a terminal layer, a fog layer and a cloud layer, is deployed to support the thermal error prediction system. The architecture is illustrated in Fig. 8. Some details are explained below:

- *Terminal layer*: The terminal layer is integrated with the physical heavy-duty machine. In the layer, sensors are mounted on the machine (the details are explained in Sections 3.1 and 4.1). The machine undergoes machining processes, and thermal sensors and laser displacement sensors are installed on the machine to continuously collect data. The optimal placement of thermal sensors will be finally guided by the result of FEA, which shows heat distributions of the machine over the entire machining process. The data are transmitted to the fog layer through an Internet router for further processing.
- *Fog layer*: Cost-effective edge devices are deployed on the fog layer to provide certain computational capability for local processing on the transmitted data. Local processors include a pre-processor *i*GRA and a trained LSTM networks. Based on the correlation among thermal data, similar data are removed by *i*GRA to minimise data collinearity and optimise the overall computing efficiency. The trained LSTM networks is used for thermal error prediction. Considering dynamics in thermal changes throughout machining processes, after a period of system service, a pre-defined threshold, the LSTM networks will be re-trained on the cloud layer based on updated time-series sensor data.
- *Cloud layer*: The cloud layer hosts a cloud server providing intensive computational power and storage spaces, so that all the processing that requires high computational resources and data storage will be done on the cloud layer. FEA is processed on the cloud layer to identify the heat distribution over the entire structure of CNC machine. The results will be sent back to the terminal layer to provide an instruction to optimise thermal sensor deployment. The LSTM networks are re-trained on the cloud layer when necessary. A system coordinator, which connects the fog layer and the cloud layer, triggers the training process of the LSTM networks with an embedded improved gradient descent optimiser.

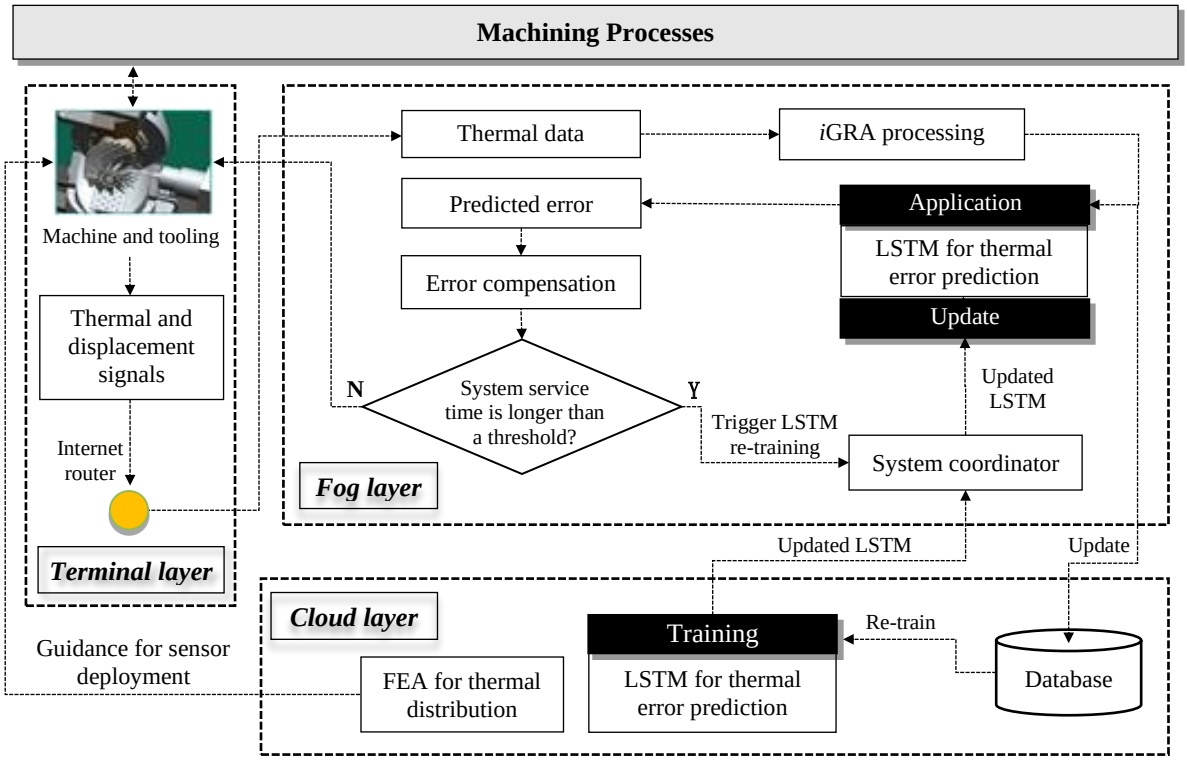


Fig. 8: The fog-cloud architecture enabled thermal error prediction and compensation.

To check the improvement of machining accuracy of the heavy-duty machine by using the proposed system, thermal errors without using the system on the *X*, *Y* and *Z* axes are measured as $E_{t\_without} = [E_{t_{X\_without}}, E_{t_{Y\_without}}, E_{t_{Z\_without}}]$, and the thermal errors with the proposed system on the *X*, *Y* and *Z* axes are measured as $E_{t\_with} = [E_{t_{X\_with}}, E_{t_{Y\_with}}, E_{t_{Z\_with}}]$. The improvement of the machining accuracy of the proposed system can be calculated:

$$Improvement = \frac{\sum_{i=1}^{3}(E_{t_{without_i}} - E_{t\_with_i})^2}{3} \tag{22}$$

In this research, specifications to implement the fog-cloud architecture are presented in Table 5. The method to quantify sensitivity and accuracy are as follows:

$$Sensitivity = \frac{Reflection\ wavelength\ shift}{\Delta T} \tag{23}$$

$$Accuracy = T_{true} - T_{sensor} \tag{24}$$

where ΔT is the change in temperature, T_{true} and T_{sensor} are the true temperature and the measured temperature by sensor, respectively. Resolution is the smallest change in temperature that the sensor can detect.

On the terminal layer, thermal sensors of Fibre Bragg gratings (FBGs) are deployed. The sensors are of small size, flexibility, immunity to electro-magnetic interference, etc. (Woyessa et al., 2020), and they also have good accuracy (⩽0.1°C). On the fog layer, the processor FPGA CycloneII EP2C5T144C8N, is deployed. FPGA exhibits good performances in executing the LSTM networks locally as it is of robust flexibility, reconfigurability and efficient parallel computing (Ahsan et al., 2012; Karakaya et al., 2018; Zairi et al., 2019; Sarić et al., 2020), while it is more cost effective in comparison with those processors deployed on the cloud layer. Training on the LSTM networks and FEA computations are computationally intensive, so that the LSTM training/re-training tasks and FEA are assigned to the cloud layer to leverage its better computational speed and larger memory.

Table 5: Detailed specifications for deployed sensors and data processing system.

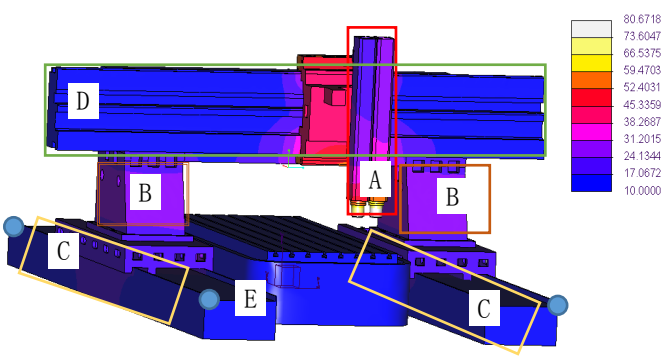
Terminal layer	**Sensor type**	**Sensitivity (pm/°C)**	**Accuracy (°C)**	**Resolution (°C)**
	Fiber Bragg grating (FBG)	38.1	≤0.1	≤0.05
Fog layer	**Processor**	**Speed**	**Memory size**	**RAM**
	FPGA: Cyclone II EP2C5T144C8N	260 MHz	119808 bit	1.1MB
Cloud layer	**Communications**	**Processor**	**Memory**	**Storage**
	2.4 GHz/5 GHz IEEE 802.11	4.2Ghz	32GB	1TB

4. Case Studies and Experimental Analysis

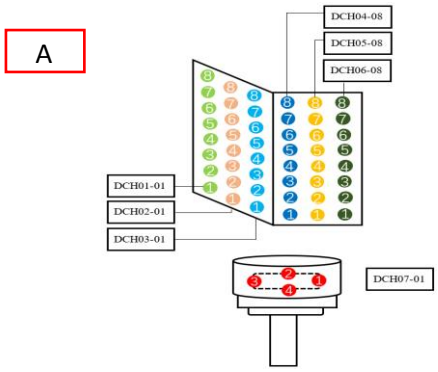
4.1 Sensor deployment

As aforementioned, the heavy-duty machine ZK5540 is used as an example to explain and verify the developed methodologies. After the FEA analysis for optimisation, there are 128 thermal sensors installed on ZK5540. The positions of the groups of sensors are marked in Fig. 9(a). More details are:

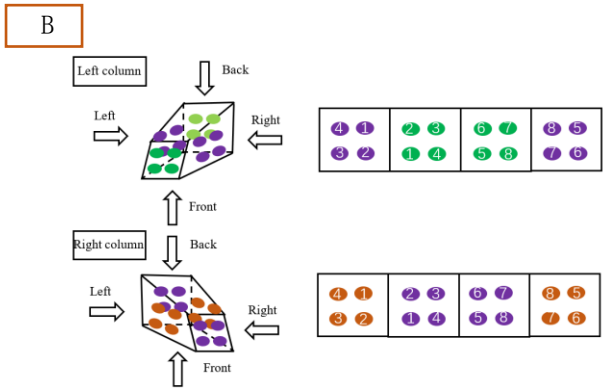
- 52 sensors are installed on the spindle (as shown in Fig. 9(b));
- 32 sensors are installed on the two columns, and each with 16 (as shown in Fig. 9(c));
- 32 sensors are installed on the two slide-ways, and each with 16 (as shown in Fig. 9(d));
- 8 sensors are installed on the beam (as shown in Fig. 9(e));
- 4 sensors are installed around machine tool to measure environmental temperature, and each side with one (as shown in Fig. 9(f)).



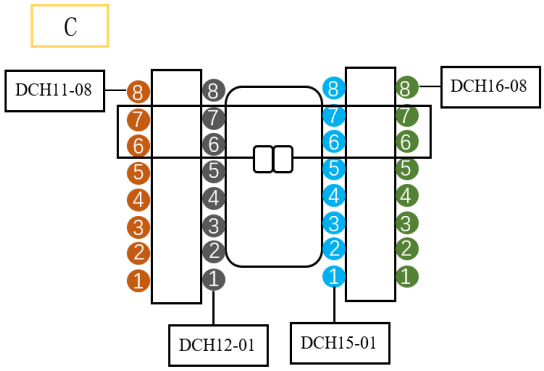
(a) The heat distribution of ZK5540 analysed by FEA



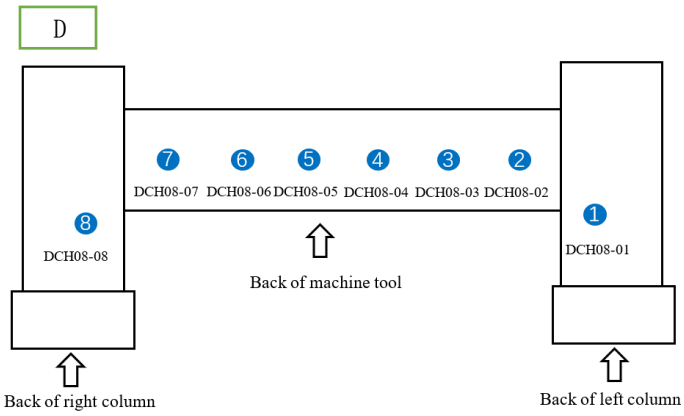
(b) Sensors on the spindle



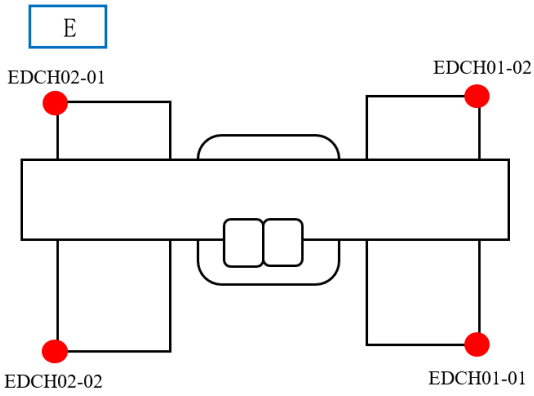
(c) Sensors on the columns



(d) Sensors on the slide-ways



(e) Sensors on the beams



(f) Sensors for the environment temperature

Fig. 9: Thermal sensor placement.

Laser displacement sensors are installed on the machining bed to measure the thermal deformation on the *X*, *Y* and *Z* directions.

All the sensor data points are presented in sequences $\boldsymbol{T} = [\sum_{j=1}^{m} t_{1,j}, \ldots, \sum_{j=1}^{m} t_{128,j}]$, where j is the time index and m is the maximum time index. Among them,

- $\sum_{j=1}^{m} t_{1,j}$ to $\sum_{j=1}^{m} t_{52,j}$ are the sensor data on the spindle,
- $\sum_{j=1}^{m} t_{53,j}$ to $\sum_{j=1}^{m} t_{84,j}$ are the sensor data on the columns,

- $\sum_{j=1}^{m} t_{85,j}$ to $\sum_{j=1}^{m} t_{116,j}$ are the sensor data on the slide-ways,
- $\sum_{j=1}^{m} t_{117,j}$ to $\sum_{j=1}^{m} t_{124,j}$ are the sensor data on the beam,
- $\sum_{j=1}^{m} t_{125,j}$ to $\sum_{j=1}^{m} t_{128,j}$ are the sensor data around the four corners of the machine.

Experiments were conducted over three months, and 9.28GB temperature data were collected. Each data node generates 0.0725GB temperature data over the period.

The counting sequence of sensors on each area is in the directions from left to right and from upward to downward. For example, the presentation of sensor data on the spindle is illustrated in Fig. 10. Fig. 11 shows that the laser displacement sensors measure the thermal deformation of the spindle during machining. The data are collected and transmitted to the fog layer through the Internet router for local processing. Data showing temperature and thermal errors on the spindle over time are displayed in Fig. 12. Some observations can be made as follows: (i) the thermal errors are sensitive to the temperature changes; (2) the changing patterns of temperatures and thermal errors over time are closely correlated; (3) a number of temperature sensors is duplicated in deployment as the measured temperatures and changing trends are similar, so that it is necessary to reduce the number of sensors by *i*GRA. The absolute thermal errors shown in Fig. 12 are in the range between 0.134mm and 0.009mm.

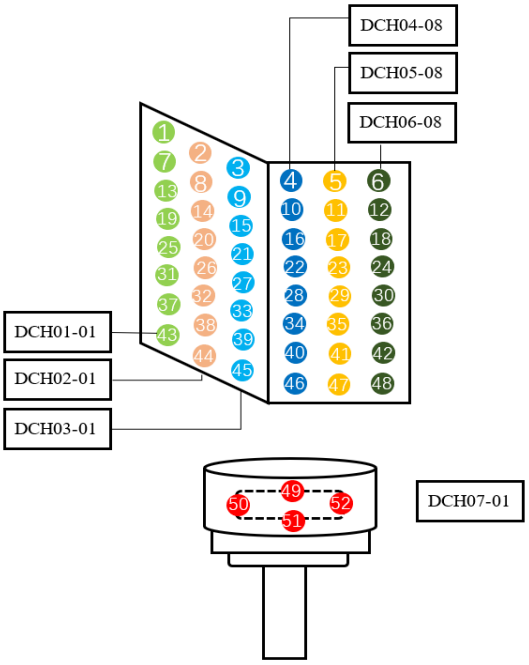


Fig. 10: Thermal sensor deployment on the spindle.

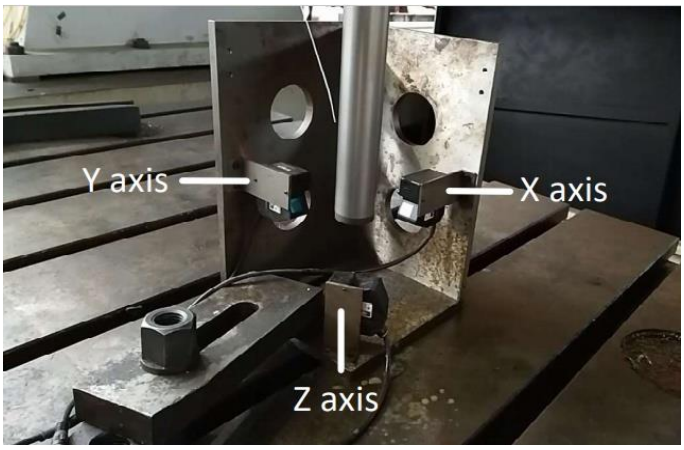
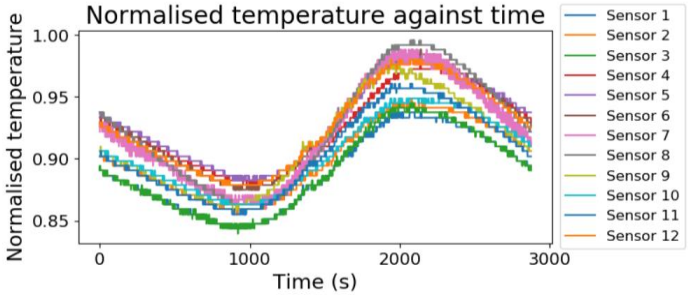
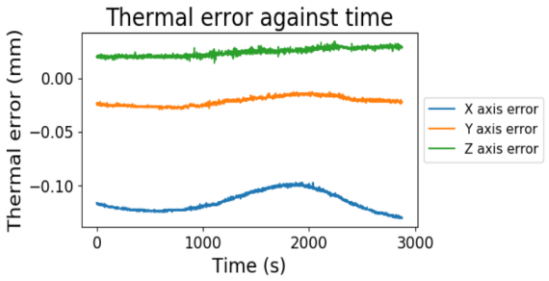


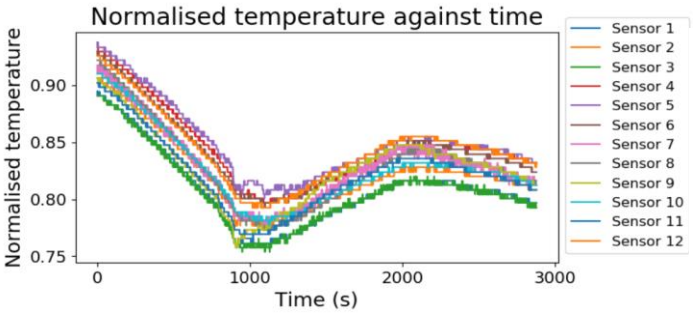
Fig. 11: Deployment of laser displacement sensors for thermal deformation measurement.



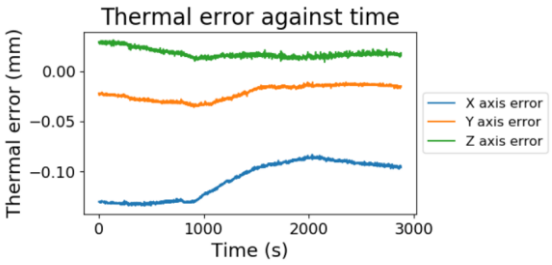
(a) Normalised temperature data (input): example 1



(b) Thermal error data (output): example 1



(c) Normalised temperature data (input): example 2



(d) Thermal error data (output): example 2

Fig. 12: Temperature data and thermal errors.

4.2 Analysis on *i*GRA

For the collected data from the 128 thermal sensors on the terminal layer, they are sent to the fog layer for pre-processing by *i*GRA. Fig. 13 illustrates the total grey relational grade among 1-12 sensors on the spindle according to Equations (4)-(6). Based on statistics, there are 12 pairs, 28 pairs, 53 pairs and 88 pairs of sensor data for more than 0.9, 0.8, 0.7 and 0.6 in terms of total grey relational grade, respectively. It is paramount to determine the value of an appropriate $threshold$ in order to ensure optimal data transmission

and the training accuracy of the LSTM networks. In this case, the training accuracy of the LSTM networks is set as priory. Therefore, the $threshold$ that achieves the highest training accuracy will be selected.

For this case, there are 8,128 grey relational grades calculated in total according to Equation (6). To determine the best $threshold$ for sensor selection, different values of $threshold$ were compared in terms of removed sensor data nodes, saved transmitted data and final training accuracy according to Equation (8). In this application, the final training accuracy is more important, so that $w_1 = 0.1$ and $w_2 = 0.9$ according to Equation (8). The results are shown in Table 6. The details of training accuracy with the designed LSTM networks are shown in Fig. 14, where each dataset is utilised for training for 10 times to ensure the robustness of the approach (10 times was the most used times for training in many applications (Kalamatianos et al., 2018)). It indicates that when $threshold = 0.65$, the fitness is the highest (0.97). That is, the final training accuracy is 93.0% and the volume of 5.00GB can be reduced when data are transmitted to the cloud layer. It represents 52.63% saving in total data transmission. It can be clearly observed that the threshold can achieve optimum accuracy quickly with robustness.

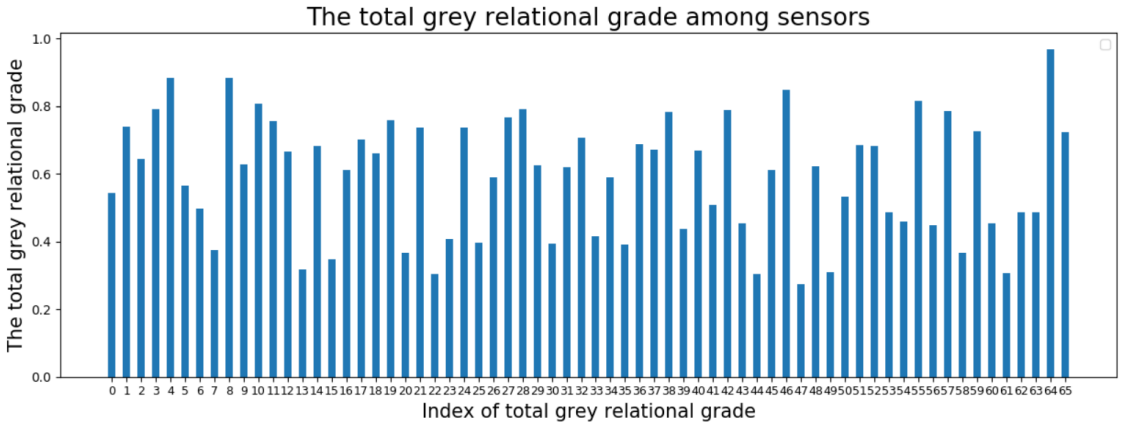


Fig. 13: The total grey relational grade among 1-12 sensors.

Table 6: Results of *i*GRA using various threshold settings.

Threshold	Removed sensor nodes	Saved data transmission (GB)	Final training accuracy (%)	Fitness
0.9	12	0.87	89	0.50
0.85	16	1.16	84	0.01
0.8	28	2.03	92	0.82
0.75	40	2.90	87	0.33
0.7	53	3.84	91	0.75
0.65	**69**	**5.00**	**93**	**0.97**
0.6	88	6.38	91	0.80

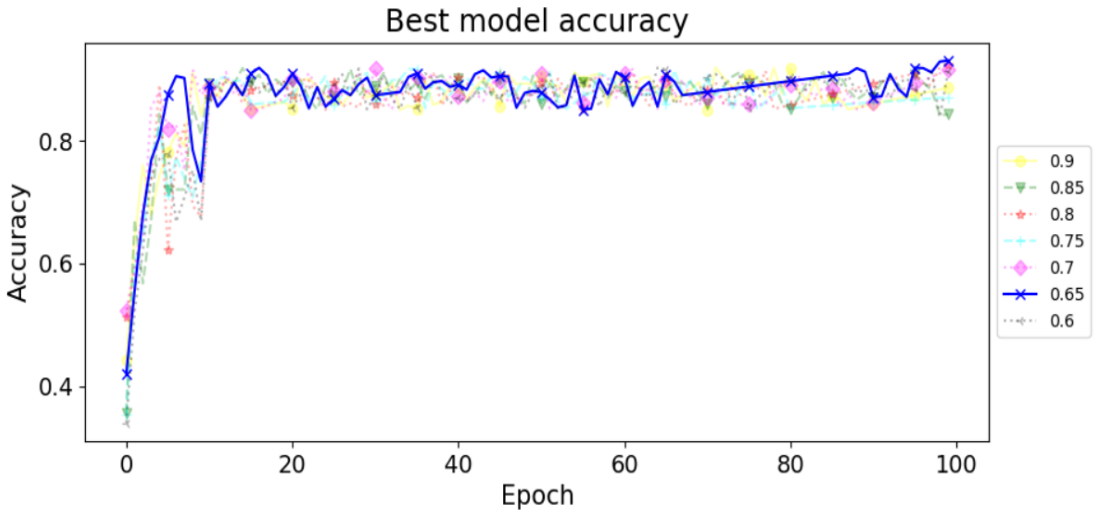


Fig. 14: Training accuracy for different thresholds in the designed LSTM networks.

4.3 Analysis on training the LSTM networks

The designed LSTM networks is trained when $threshold = 0.65$. When the model is required to update, an improved gradient descent is applied to optimise the LSTM networks. To justify the selection of the LSTM design in this research, different algorithms, including RNNs, Artificial Neural Network (ANN), CNNs and Support Vector Machine (SVM), were utilised for benchmarking. Fig. 15 and Table 7 show the training details. It shows that both the best and average training accuracy for a 1-layer LSTM networks with the improved gradient decent can achieve 93.1% and 88.9%, which are the highest among the comparative algorithms. Compared with the 1-layer LSTM networks, a 2-layer LSTM networks is 25.46% lower in the average accuracy and a 3-layer LSTM is 2.90% lower in the average accuracy.

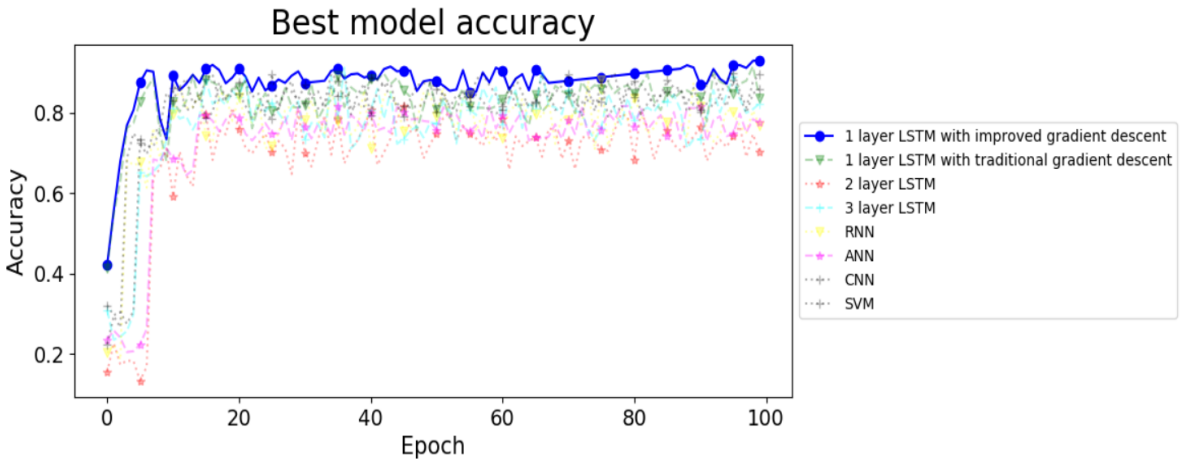


Fig. 15: Training accuracy of different algorithms.

Table 7: Benchmarking for different algorithms.

Algorithms	Best accuracy (%)	Worst accuracy (%)	Average accuracy (%)
1-layer LSTM with an improved gradient descent	**93.1%**	**85.2%**	**88.9%**
1-layer LSTM with a traditional gradient descent	85.0%	81.9%	83.9%
2-layer LSTM	69.4%	67.1%	68.8%
3-layer LSTM	90.4%	87.6%	89.5%
RNN	85.3%	83.1%	84.3%
ANN	77.5%	71.2%	76.7%
CNN	89.6%	81.0%	84.3%
SVM	85.9%	82.0%	83.9%

Experiments were also carried out to select optimal parameters in Equations (19)-(21). Fig. 16 and Table 8 illustrate the training details. It shows that, when training parameters are ε_0=0.1 and k=0.02, the best and average accuracy can achieve 93.1% and 88.9%, which are the highest among the results in different parameter settings. It can be seen that when ε_0=0.1 and k=0.01, the best accuracy is the lowest (73.4%). The proposed parameter has 19.70% higher best training accuracy than that of the lowest best accuracy. Based on the proposed method, they system is deployed in production process, the thermal errors before and after deploying the system ($E_{t\_without}$ and E_{t_with}) have been measured. According to calculation Equation (16), 46.53% improvement in machining accuracy was achieved.

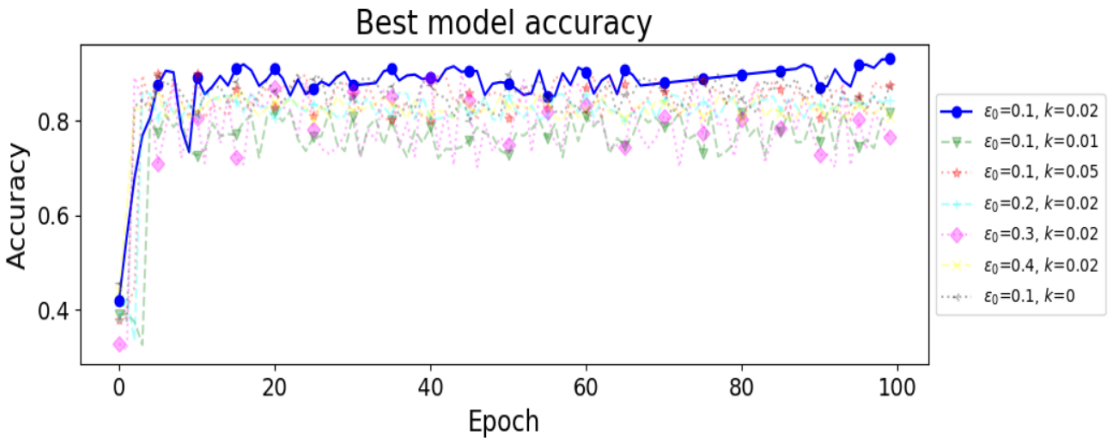


Fig. 16: Training accuracy of different training parameters.

Table 8: Benchmark for different training parameters.

Parameters	Best accuracy	Worst accuracy	Average accuracy
ε_0=0.1, k=0.02	**93.1%**	**85.2%**	**88.9%**
ε_0=0.1, k=0.01	73.4%	68.5%	72.0%
ε_0=0.1, k=0.05	84.2%	75.0%	80.7%
ε_0=0.2, k=0.02	83.8%	69.1%	77.5%
ε_0=0.3, k=0.02	80.4%	73.4%	78.0%
ε_0=0.4, k=0.02	85.1%	72.6%	80.0%
ε_0=0.1, k=0	86.6%	77.8%	83.4%

5. Conclusions

To improve the machining precision of heavy-duty CNC machines, an important research topic is how to design an effective approach for accurately predicting the deformations of machine elements caused by heats generated during machining processes. Approaches of data-based modelling have become increasingly adopted but they exhibit low efficiency in processing large-volume thermal data. To tackle the issue, this paper presents a new system enabled by a LSTM networks and fog-cloud architecture for thermal error prediction of heavy-duty machines. The system has the following characteristics: (1) The system uses physics-based modelling to optimise data-based modelling in the aspect of deciding the optimal number/locations of deployed thermal sensors and minimised measured data; (2) The LSTM networks is integrated with an effective data pre-processor to improve thermal error prediction in terms of prediction accuracy and data processing efficiency; (3) The fog-cloud architecture can improve the system efficiency of data transmission and system responsiveness to compensate thermal errors on the fly. The developed system was validated using an industrial heavy-duty CNC machine. Based on industrial case studies, the volume of transmitted data was minimised by 52.63% and the machining accuracy was improved by 46.53%, in comparison with the processes without using the developed system. It proves that the designed system has a great potential in supporting real-world industrial applications.

In future research, the following aspects will be further investigated to improve the system:

- Computation resource optimisation between both fog layer and cloud layer will be further researched to achieve optimum computation resource distribution among the layers;

- To enhance the machining error prediction, different data source (e.g., vibration, contact force, etc.) could be fused for analysis;
- This research investigates machining inaccuracy from the aspect of thermal errors. There are more factors that influence machining accuracy, such as machining condition, machining shapes, schedules, etc. These will be analysed in the future.

Acknowledgement:

This research was sponsored by the European Commission (Smarter project, No. PEOPLE-2013-IAPP-610675), the National Natural Science Foundation of China (No. 51975444), and an international cooperative project between the Wuhan University of Technology (China) and the Birmingham University (UK). The authors also acknowledge industrial partners and colleagues from the collaborative universities to support the research.

References:

Abdulshahed A., Longstaff A., Fletcher S. and Myers A. (2015). Thermal error modelling of machine tools based on ANFIS with fuzzy c-means clustering using a thermal imaging camera. *Applied Mathematical Modelling*. 41, pp.130-142.

Abdulshahed, A., Longstaff, A., Fletcher, S. and Potdar, A. (2016). Thermal error modelling of a gantry-type 5-axis machine tool using a Grey Neural Network Model. *Journal of Manufacturing Systems*, 41, pp.130-142.

Ahsan M., Ibrahimy M., Khalifa O. and Ullah M. (2012). VHDL modeling of EMG signal classification using artificial neural network. *Journal of Applied Sciences*, 12(3), pp.244-253.

An Q., Tao Z., Xu X., El Mansori M. and Chen M. (2020). A data-driven model for milling tool remaining useful life prediction with convolutional and stacked LSTM network. *Measurement*, 154, p.107461.

Bellavista P., Berrocal J., Corradi A., Das S., Foschini L. and Zanni A. (2019). A survey on fog computing for the Internet of Things. *Pervasive and Mobile Computing*, 52, pp.71-99.

Blaser, P., Pavliček, F., Mori, K., Mayr, J., Weikert, S. and Wegener, K. (2017). Adaptive learning control for thermal error compensation of 5-axis machine tools. *Journal of Manufacturing Systems*, 44, pp.302-309.

Bryan J. (1990). International status of thermal error research. *CIRP Annals*, 39(2), pp. 645-656.

Cheng Q., Qi Z., Zhang G., Zhao Y., Sun B. and Gu P. (2015). Robust modelling and prediction of thermally induced positional error based on grey rough set theory and neural networks. *International Journal of Advanced Manufacturing Technology*, 83(5-8), pp. 753-764.

Chen, L., Xu, G., Zhang, S., Yan, W. and Wu, Q. (2020). Health indicator construction of machinery based on end-to-end trainable convolution recurrent neural networks. *Journal of Manufacturing Systems*, 54, pp.1-11.

Cinar Y., Mirisaee H., Goswami P., Gaussier E. and Aït-Bachir A. (2018). Period-aware content attention RNNs for time series forecasting with missing values. *Neurocomputing*, 312, pp. 177-186.

Ellis M. and Chinde V. (2020). An encoder–decoder LSTM-based EMPC framework applied to a building HVAC system. *Chemical Engineering Research and Design*, 160, pp.508-520.

Fujishima M., Narimatsu K., Irinoa N., Mori M. and Ibaraki S. (2019). Adaptive thermal displacement compensation method based on deep learning. *CIRP Journal of Manufacturing Science and Technology*, 25, pp. 22-25.

Hu P., Dhelim S., Ning H. and Qiu T. (2017). Survey on fog computing: architecture, key technologies, applications and open issues. *Journal of Network and Computer Applications*, 98, pp. 27-42.

Huang J., Zhou Z.D., Liu M., Zhang E., Chen M., Pham D.T. and Ji C. (2015). Real-time measurement of temperature field in heavy-duty machine tools using fiber bragg grating sensors and analysis of thermal shift errors. *Mechatronics*, 31, pp. 16-21.

Huang, Y. and Hoshi, T. (2001). Optimization of fixture design with consideration of thermal deformation in face milling. *Journal of Manufacturing Systems*, 19(5), pp.332-340.

Kalamatianos R., Kermanidis, K., Karydis, I. and Avlonitis, M. (2018). Treating stochasticity of olive-fruit fly's outbreaks via machine learning algorithms. *Neurocomputing*, 280, pp.135-146.

Karakaya B., Kaya T. and Gulten A. (2018). FPGA-based ANN design for detecting epileptic seizure in EEG signal. *Balkan Journal of Electrical and Computer Engineering*, pp.15-19.

Kumar V. and Singh H. (2020). Parametric optimization of rotary ultrasonic drilling using Grey Relational Analysis. *Materials Today: Proceedings*, 22, pp.2676-2695.

Kuo Y., Yang T. and Huang G. (2008). The use of grey relational analysis in solving multiple attribute decision-making problems. *Computers & Industrial Engineering*, 55(1), pp. 80-93.

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

Li F., Li T., Jiang Y., Wang H. and Ehmann K. (2019). Explicit error modeling of dynamic thermal errors of heavy machine tool frames caused by ambient temperature fluctuations. *Journal of Manufacturing Processes*, 48, pp. 320-338.

Li F., Li T., Wang H. and Jiang Y. (2017). A temperature sensor clustering method for thermal error modelling of heavy milling machine tools. *Applied Sciences*, 7(1), p. 82.

Li F., Xiang W., Wang J., Zhou X. and Tang B. (2018). Quantum weighted long short-term memory neural network and its application in state degradation trend prediction of rotating machinery. *Neural Networks*, 106, pp.237-248.

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*, 95, pp. 20-38.

Liang Y.C., Li W.D., Lu X. and Wang S. (2019). Fog computing and convolutional neural network enabled prognosis for machining process optimization. *Journal of Manufacturing Systems*, 52, pp. 32-42.

Liu, C., Vengayil, H., Zhong, R. and Xu, X. (2018). A systematic development method for cyber-physical machine tools. *Journal of Manufacturing Systems*, 48, pp.13-24.

Liu C. and Zhu L. (2020). A two-stage approach for predicting the remaining useful life of tools using bidirectional long short-term memory. *Measurement*, 164, p.108029.

Liu Q., Yan J., Pham D.T., Zhou Z.D., Xu W., Wei Q. and Ji C. (2015). Identification and optimal selection of temperature-sensitive measuring points of thermal error compensation on a heavy-duty machine tool. *International Journal of Advanced Manufacturing Technology*, 85, pp. 345-353.

Ma C., Zhao L., Mei X., Shi H. and Yang J. (2016). Thermal error compensation of high-speed spindle system based on a modified BP neural network. *International Journal of Advanced Manufacturing Technology*, 89, pp. 3071-3085.

Miao E., Liu Y., Liu H., Gao Z. and Li W. (2015). Study on the effects of changes in temperature-sensitive points on thermal error compensation model for CNC machine tool. *International Journal of Machine Tools and Manufacture*, 97, pp. 50-59.

Mohanty S., Lydia E., Elhoseny M., Al Otaibi M. and Shankar K. (2020). Deep learning with LSTM based distributed data mining model for energy efficient wireless sensor networks. *Physical Communication*, 40, p. 101097.

Qu Z., Zheng, S., Wang, X., Song, X., Li, B. and Song, X. (2018). Converged Recommendation System Based on RNN and BP Neural Networks. *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*.

Sarić R., Jokić, D., Beganović, N., Pokvić, L. and Badnjević, A. (2020). FPGA-based real-time epileptic seizure classification using Artificial Neural Network. *Biomedical Signal Processing and Control*, 62, p.102106.

Sood S.K. and Mahajan I. (2018). Fog-cloud based cyber-physical system for distinguishing, detecting and preventing mosquito borne diseases. *Future Generation Computer Systems*, 88, pp. 764-775.

Tao, F., Qi, Q., Liu, A. and Kusiak, A. (2018). Data-driven smart manufacturing. *Journal of Manufacturing Systems*, 48, pp.157-169.

Vlachas P., Pathak J., Hunt B., Sapsis T., Girvan M., Ott E. and Koumoutsakos P. (2020). Backpropagation algorithms and Reservoir Computing in Recurrent Neural Networks for the forecasting of complex spatiotemporal dynamics. *Neural Networks*, 126, pp.191-217.

Vlachas P., Pathak J., Hunt B., Sapsis T., Girvan M., Ott E. and Koumoutsakos P. (2020). Backpropagation algorithms and reservoir computing in Recurrent Neural Networks for the forecasting of complex spatiotemporal dynamics. *Neural Networks*, 126, pp.191-217.

Woyessa G., Pedersen J., Nielsen K. and Bang O. (2020). Enhanced pressure and thermal sensitivity of polymer optical fiber Bragg grating sensors. *Optics & Laser Technology*, 130, p.106357.

Wu D., Liu S., Zhang L., Terpenny J., Gao R., Kurfess T. and Guzzo J. (2018). A fog computing-based framework for process monitoring and prognosis in cyber-manufacturing. *Journal of Manufacturing Systems*, 43, pp. 25-34.

Xia T., Song Y., Zheng Y., Pan E. and Xi L. (2020). An ensemble framework based on convolutional bi-directional LSTM with multiple time windows for remaining useful life estimation. *Computers in Industry*, 115, p.103182.

Xu, W., Cui, J., Li, L., Yao, B., Tian, S. and Zhou, Z. (2020). Digital twin-based industrial cloud robotics: Framework, control approach and implementation. *Journal of Manufacturing Systems* (in press).

Yang J., Guo Y. and Zhao W. (2019). Long short-term memory neural network based fault detection and isolation for electro-mechanical actuators. *Neurocomputing*, 360, pp.85-96.

Yao X., Hu T., Yin G. and Cheng C. (2020). Thermal error modeling and prediction analysis based on OM algorithm for machine tool's spindle. *International Journal of Advanced Manufacturing Technology*, 106, pp. 3345-3356.

Zairi H., Kedir T/M., Meddah K. and Ould S. (2019). FPGA-based system for artificial neural network arrhythmia classification. *Neural Computing and Applications*, 32(8), pp.4105-4120.

Zhang J., Wang P., Yan R. and Gao R. (2018). Long short-term memory for machine remaining life prediction. *Journal of Manufacturing Systems*, 48, pp.78-86.