2022 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE) | 978-1-6654-8432-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/CCECE49351.2022.9918222

Deep Neural Network Modeling for Accurate Electric Motor Temperature Prediction

Siavash Hosseini
Department of Software Engineering
Lakehead University
Thunder Bay, ON P7B 5E1, Canada
shossei4@lakeheadu.ca

Amirmohammad Shahbandegan
Department of Computer Science
Lakehead University
Thunder Bay, ON P7B 5E1, Canada
ashahban@lakeheadu.ca

Thangarajah Akilan
Department of Software Engineering
Lakehead University
Thunder Bay, ON P7B 5E1, Canada
takilan@lakeheadu.ca

***Abstract*—Electric motors are becoming widely used in many different applications, such as electric cars, and turbines. Measuring the temperature of internal components of an electric motor, like permanent magnet synchronous motor (PMSM) is vital to maintain its safe operation. However, measuring the temperature of the permanent magnet and stator directly comes at the expense of higher cost and additional hardware requirement, for instance, a sensor network. To overcome these limitations, machine learning (ML) techniques can be employed to model the mentioned parameters without the need of specialized sensors and design ideas for housing them inside the motors. Classical methods, like lumped-parameter thermal networks (LPTNs) are capable of calculating the temperature of internal elements of PMSMs. But, these methods require expertise and may lack an acceptable accuracy. In this study, two deep neural networks (DNNs) were modeled using convolutional neural network (CNN) and long short-term memory (LSTM) units to predict the temperature of four target values of PMSMs: stator tooth, stator yoke, stator winding, and permanent magnet. For attribute conditioning, exponentially weighted moving average (EWMA) and exponentially weighted moving standard deviation (EWMS) were applied. A thorough ablation analysis shows that the CNN-based model predicts the targets better than the LSTM model with an average mean squared error (MSE) of 2.64 $^{\circ}C^2$ and an average R^2 of 0.9924. It is also found that the proposed CNN-based model achieves a 13% mean average performance (mAP) improvement compared to the existing state-of-the-art solution.**

***Index Terms*—deep learning, permanent magnet synchronous motor, temperature prediction, time series analysis**

I. Introduction

The main part of the heat loss comes from iron loss, copper loss as well as mechanical loss. Stator's voltage has direct effect on iron loss and motor speed affect mechanical loss. The copper loss influence the heating degree of stator winding. In order to guarantee the lifespan and proper functionality of the PMSMs, there is a focus in the research community to measure temperature of PMSM's internal components (cf. Fig. 1) using from mathematical modeling to modern data-driven approaches. For instance, Wallscheid *et al.* [1] utilized precise flux observer to measure magnet's temperature indirectly in the fundamental wave domain without any signal injectors and sensors. However, the methods that have been used in this case comes with huge cost and preparation of apparatus. This is why researchers have started applying data-driven models for electric motors' temperature estimation after

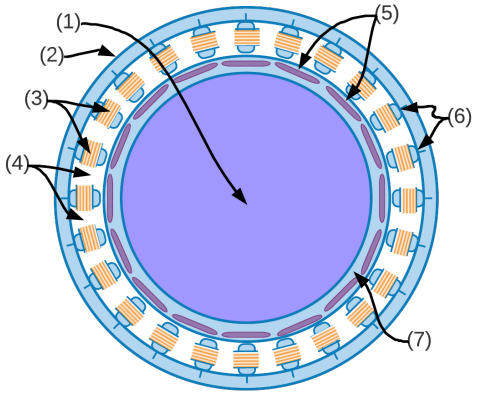

Fig. 1. PMSM's cross section: (1) rotor shaft, (2) stator yoke, (3) stator winding, (4) stator slot, (5) permanent magnet, (6) stator tooth, (7) rotor ion.

gathering enough data. In this line, the traditional methods deal with finite element method for temperature prediction [2]. It is mainly derived by simulating transient temperature by utilizing finite element, establish the temperature field, and finally estimating the motor temperature. This technique is capable to measure the current linear data and it is not applicable for large scale and nonlinear historical data. On the other hand, ML techniques can efficiently solve the aforementioned problem and improve the prediction process in comparison with former methods. For instance, Hyeontae *et al.* [3] used a ML-based approach, particularly, the decision tree, for temperature estimation of Linz–Donawiz converter for improving the quality of melted pig iron. Similarly, Wang [4] built a random forest model to predict the temperature of a ladle furnace. Su *et al.* [5] proposed an extreme learning machine-based implementation to estimate iron temperature in a blast furnace. Hence, Zhukov *et al.* [6] studied different methods based on random forest for assessing the security level in power systems. The feature extraction, in most of the ML techniques, rely on human involvement [7]–[9]. For elementary tasks, this way of feature extraction is efficient and simple but for complex problem, manual feature extraction is a challenging procedure [10]–[13]. Meanwhile, the DNNs are capable of handling the feature extraction automatically within itself via an end-to-end learning procedure [14], [15].

In recent decade, DNNs have been widely adopted for solving practical industrial problems [16], [17]. Thus, by exploiting the advantage of DNNs, this work proposes two models based on CNNs and LSTM modules. The rest of this paper is organized as follows. Section II reviews important relevant works. Section III elaborates the proposed methodology, and Section IV discuss the experimental findings. Finally, Section V concludes the paper with future direction.

II. Literature review

Kirchgassner *et al.* [18] studied multiple machine learning models for accurate estimation of motors' thermal stress. For this, they analyzed the applicability of different approaches, such as support vector machine (SVM), K-nearest neighbors (KNN), randomized tree, and neural networks (NN). They argued that classical supervised learning models can predict the temperature profiles with an acceptable accuracy even in high dynamic drive cycles. At the same time, some researchers attempted to exploit the capability of DNNs in handling sequential data, in this case, temperature of PMSM's internal elements. For instance, Kasburg *et al.* [19] utilized LSTM neural network for photovoltaic power generation estimation that achieved a significant improvement over former studies. Gui *et al.* [20] implemented a DNN-based model, for multi-step feature selection, and a genetic algorithm (GA)-based solution, to predict the temperature of a reheater system. Similarly, Wallscheid *et al.* [21] investigated the application of recurrent neural networks (RNNs) to accomplish precise temperature prediction of PMSMs. They used particle swarm optimization for finding suitable hyper-parameters, for example, the number of hidden layers and neurons of their model.

Kirchgässner *et al.* [22] employed temporal convolutional network (TCN) for electric motors' temperature prediction. Their model achieves a mean average MSE (MA-MSE) of 3.04 $°C^2$. Similar to the canonical RNN, the TCN is a causal system such that a response at time stamp t is the result of 1D-kernels convolved only with input quantities from time stamp t and before in the earlier time stamps. Thus, the TCN can be viewed as 1-D fully convolutional networks (FCN) with causal convolutions. It is found that the TCN outperforms generic RNNs in certain applications, like polyphonic music modeling, character/word-level language modeling, and IoT anomaly detection [23]. On the other hand, Lee *et al.* [24] proposed a feed-forward neural network (FNN) based on nonlinear auto-regressive exogenous (NARX) to estimate the temperature of the permanent magnet and the stator winding in PMSMs. The authors indicated that their NARX structure would produce better results than the RNN and LSTM-based models if the relationship between the input and the output is known. Although there are some works that focus on the application of DNNs for PMSMs' internal components' temperature prediction, there is still a huge scope for further investigation and development of such models to be entrusted by the industrial community. Thus, in this work, two different DNNs have been introduced aiming to enhance the precision of temperature prediction of permanent magnet, stator tooth, stator winding, and stator yoke of PMSMs, simultaneously.

TABLE I
Architecture Detail of the 1-D CNN-based Model

Layer ID	Layer Type	Output Dimension
Input	Input layer	(1024, 128, 108)
L1	Conv 1D	(1024, 122, 64)
L2	Conv 1D	(1024, 116, 64)
L3	MaxPooling 1D	(1024, 58, 64)
L4	Flatten	(1024, 3712)
L5	Dense	(1024, 32)
L6	Dense	(1024, 4)
Total number of trainable parameters: 204,196		
Number of filters in L1 and L2: 64		
Kernel size in L1 and L2: 7, and in L3: 2		
Activation function: L1-L5: ReLu; L6: linear		

III. Methodology

A. The Proposed Models

DNNs have been proved to be strong tools for dealing with regression, classification, and optimization problems in industrial environments [9], [10], [25]. Thus, this work extends upon the existing DNN technologies, such as 1-D convolution and LSTM cells to improve the temperature prediction accuracy of PMSMs. The following subsections highlights the key details of the two proposed models.

1) 1-D CNN-based Model: CNNs are widely used in visual recognition-based applications, viz. image classification, object detection, tracking, and segmentation [8], [15], [26]. One of the outstanding characteristics of CNNs is the spatio-local connectivity, through which layers share the parameters, making them efficient learning models. It has been realized that not only do these networks have superior performance, but also have dominant performance in sequential data analytical problems. In CNNs, the convolution (Conv) layer plays a vital role in feature extraction. The data that pass through the Conv layers convolve with the respective kernels in each layer. The Conv operation (i.e., dot product) between the whole input data and the kernels will generate a volume of feature maps. In this case, as the data is a 1-D sequential data, each Conv layer receives a 1-D input data, $x(n)$. Then, a 1-D kernel $w(n)$, convolve with the input will generate a feature map, $z(n)$ as defined in (1) [27].

$$z(n) = x(n) * w(n) = \sum_{m=-l}^{l} x(m) \cdot w(n-m), \qquad (1)$$

where l stands for the kernel size.

The proposed 1-D CNN-based model consists of six layers subsuming a total of 204, 196 trainable parameters. Table I summarizes the proposed CNN-based model's connectivity pattern with respective layer details.

2) LSTM-based Model: LSTM networks have shown great promise in sequential data analysis, such as moving object detection in videos [26], and speech recognition [28], and stock market prediction [29]. Due to their recurrent connections and

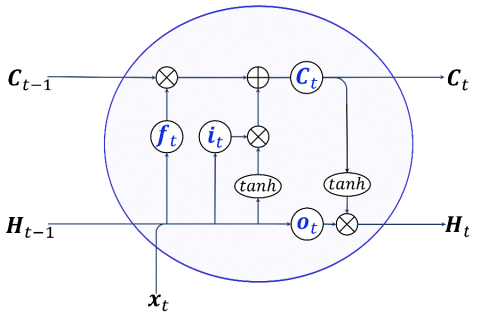

Fig. 2. Illustration of a standard LSTM cell with three gates that control information flow from a time-series data, where $\mathbf{X}_t$, $\mathbf{C}_t$, $\mathbf{H}_t$, $\mathbf{i}_t$, $\mathbf{f}_t$ and $\mathbf{o}_t$ are the input quantity from a time-series data, cell state, hidden state, input gate, forget gate, and output gate, respectively, at timestamp, t.

memory mechanism, the LSTM networks can preserve valuable past observations. Through three gating functions: input gate, output gate, and forget gate defined by (2) - (6), these networks are capable of adding and removing information to the memory. Fig. 2 illustrates the interconnection of these gates in a standard LSTM unit.

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + b_i), \tag{2}$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + b_f), \tag{3}$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + b_o), \tag{4}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c), \tag{5}$$

$$H_t = o_t \circ \tanh(C_t). \tag{6}$$

where X_t is an input quantity from a time-series data, C_t is the cell state, H_t is the hidden state, and i_t, f_t, and o_t are the gates of the LSTM block at timestamp t. Hence, W, '$*$', and '$\circ$' denote conv kernels specific to the gates and internal states, the conv operator, and Hadamard product. The σ is a `hard sigmoid` function. The proposed LSTM-based model consists of five layers subsuming a total of 936, 228 trainable parameters. Table II summarizes the proposed LSTM-based model's connectivity pattern with respective layer details.

TABLE II
ARCHITECTURE DETAIL OF THE LSTM-BASED MODEL

Layer ID	Layer Type	Output Dimension
Input	Input layer	(1024, 128, 108)
L1	LSTM	(1024, 128, 256)
L2	LSTM	(1024, 256)
L3	Dense	(1024, 128)
L4	Dense	(1024, 32)
L5	Dense	(1024, 4)
Total number of trainable parameters: 936,228		
Number of hidden units in L1 and L2: 256		
Activation function: L1, L2: tanh; L3, L4: Relu; L5: linear		

B. Data Pre-processing

1) Data Scaling: Data scaling is an essential pre-processing step in time-series prediction problems and it has shown great performance gain in DNNs [30]. In this work, the standard scalar was used to normalize the values by subtracting the mean and dividing by the standard deviation of the values considering all the samples as given in (7). It is worth mentioning that during inference time on the test set, the predicted target values were re-scaled back to the original range.

$$x_{scaled} = \frac{x - \mu(x)}{\sigma(x)}, \tag{7}$$

where x is the raw input, and its sample mean and standard deviation are denoted by $\mu(x)$ and $\sigma(x)$, respectively.

TABLE III
INPUT AND TARGET VARIABLES [22]

Variable name	Symbol
Measured Input Variables	
Ambient temperature	θ_a
Liquid coolant temperature	θ_c
Actual voltage d-axis component	u_d
Actual voltage q-axis component	u_q
Actual current d-axis component	i_d
Actual current q-axis component	i_q
Motor speed	n_{mech}
Torque	τ
Derived Input Features	
Voltage magnitude: $\sqrt{u_d^2 + u_q^2}$	u_s
Current magnitude: $\sqrt{i_d^2 + i_q^2}$	i_s
Electric apparent power: $1.5 * u_s * i_s$	S_{el}
Joint Interaction-A: $i_s * n_{mech}$	J_1
Joint Interaction-B: $S_{el} * n_{mech}$	J_2
Target Variables	
Permanent magnet temperature	θ_{PM}
Stator teeth temperature	θ_{ST}
Stator winding temperature	θ_{SW}
Stator yoke temperature	θ_{SY}

2) Derived Attributes: According to the physical representations of quantities, like the total current, the total voltage and the apparent power, and joint interaction between two inputs, a total of five features are derived as tabulated in Table III.

3) EWMA and EWMS: New features are computed based using EWMA and EWMS with different span values defeined by (8) and (9), where $\alpha = 2/(s+1)$ and s is an arbitrary span or a lookback period defined by the user. As a result, at each time step t the EWMA and EWMS of each input (cf. Table III) with four different spans are fed to the proposed models as additional inputs. As the input series is extremely dynamic with high fluctuation, the exponentially weighted features can capture some historical information of the inputs and boost the models' performances. A sample feature and its EWMA and EWMS with different spans are depicted in Fig. 3 and Fig. 4.

$$\mu_x = \frac{\sum_{i=0}^{t}(1-\alpha)^i x_{t-i}}{\sum_{i=0}^{t}(1-\alpha)^i}, \tag{8}$$

$$\sigma_x = \frac{\sum_{i=0}^{t}(1-\alpha)^i (x_i - \mu_t)}{\sum_{i=0}^{t}(1-\alpha)^i}. \tag{9}$$

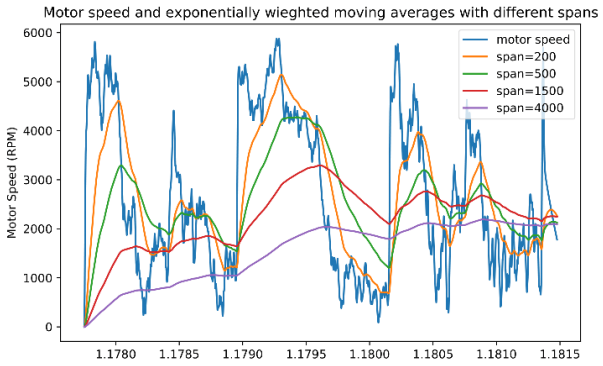

Fig. 3. Raw motor speed and its EWMA features with different spans.

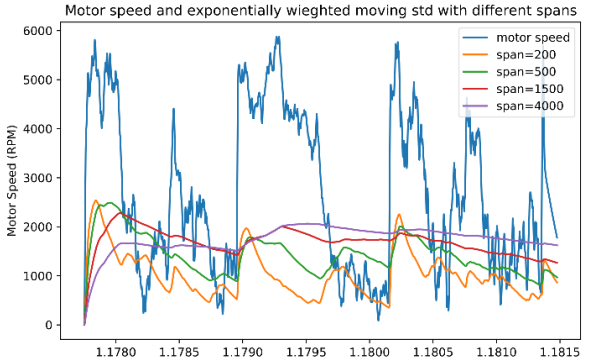

Fig. 4. Raw motor speed and its EWMS with different spans.

C. Model Implementation and Training Strategy

Before finalizing the models described in Tables I and II, several CNN and LSTM networks with different number of layers and various architectures were implemented and their performances were recorded in the test set. Those primitive results are omitted due to space limitation. The finalized models are trained using `Adam` optimizer with an initial learning rate of 0.001, and an input batch size of 1024. Moreover, early stopping technique was used to handle overfitting issues and both models reached to the optimal condition between 8 and 17 epochs.

IV. Experimental Analysis

A. Data set

The data set comprises 185 hours of temperature recording of a PMSM sampled at 2 Hz frequency [22]. The variables of the data set along with the derived attributes inspired by [22] and target values are listed in Table III. It consists of sixty-nine profiles under different load conditions with the duration of each profile ranging from 0.3 to 6.1 hours. For a fair comparison, following Kirchgässner *et al.* [22], the profile ID 65 is chosen as the testing set. The rest is used for training, except for one profile (about 4.6 hours), which is used for model validation. A sample from the training set, profile ID 52, is visualized through a trend plot in Fig. 6.

Fig. 5. Heatmap of linear correlations between different attributes of the data.

To better understand the nature of the data, correlation between each of the attribute is calculated and visualized via a heat-map as in Fig. 5. It is noteworthy that the correlation between two of the attributes, namely, `i_q` and `torque`, is equal to 1. Thus, it is safe to remove `torque` from the input attribute list of the proposed models. In addition, as a result of some elements' close proximity in PMSM' internal structure, they experience similar thermal stress, so they have high correlation values. For example, the correlation between θ_{ST}, θ_{SY} and θ_{SW} is close to 1, because they are close to each other (cf. Fig. 1) and receive approximately the same amount of heat compared to center element – the rotor shaft.

B. Evaluation Metrics

Since this work handles a regression problems (i.e., real value prediction), the mean squared error defined in (10) is used as the primary evaluation metric. The smaller MSE value indicates better performance of the model under study.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, i = 1, 2, \cdots n, \qquad (10)$$

where n, y_i, $\hat{y}_i$, and i are the total number of data points predicted, true value, the predicted value, and sample index, respectively.

C. Overall Analysis

The data set used in this work is an emerging one; hence, there is not many works that have used the data set for experimental studies. Thus, this work is able to perform a comparative analysis with only one recent work - Kirchgässner *et al.* [22] in Table IV. The analysis reveals that the 1-D CNN-based model has performed better than the LSTM-based model and the existing solution. Compared to the LSTM-based model, the CNN-based model records $> 35\%$ better overall performance. The intuition for this improvement is the

TABLE IV
COMPARISON BETWEEN DIFFERENT MODELS WRT MSE: MA-MSE - MEAN AVERAGE MSE ACROSS ALL FOR TARGETS, MAP - MEAN AVERAGE PERFORMANCE COMPARED TO THE BASELINE MODEL, TCN [22]. NOTE: THE BEST PERFORMANCES ARE INKED IN GREEN

Model	θ_{ST}	θ_{SW}	θ_{SY}	θ_{PM}	MA-MSE	mAP Improvement
Proposed 1-D CNN Model	2.21	3.34	1.52	3.50	2.64	+13%
Proposed LSTM Model	3.75	5.62	3.98	3.14	4.12	-35%
TCN-Kirchgässner *et al.* [22]	2.84	6.90	1.80	0.65	3.04	Baseline
LTPN-Gedlu *et al.* [31]	-	-	-	-	5.73	-88%
RNN-Kirchgässner *et al.* [22]	-	-	-	-	8.70	-186%

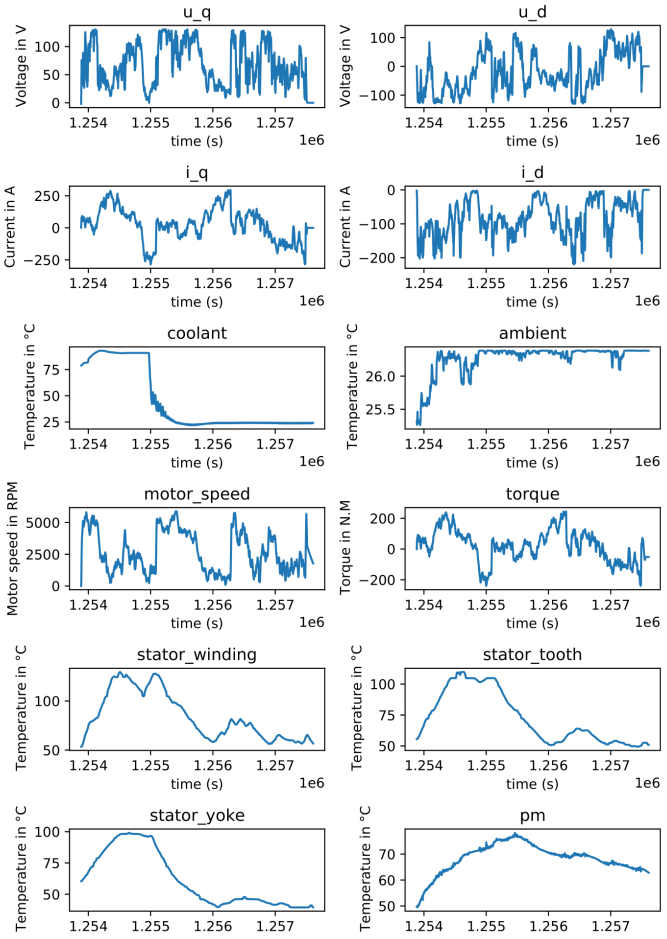

Fig. 6. Profile ID 52. A sample load profile from the training set.

excellent feature learning capability of the CNNs. As EWMA and EWMS capture the historical information, the CNN model unearths some intricate features that can not be learnt by pure long short-term memory-based systems. At the same time, compared to the existing work, the proposed CNN-based model also achieves a 13% enhancement in average overall performance across all the target variables. Hence, the model's R^2 score for each target variable: θ_{ST}, θ_{SW}, θ_{SY}, and θ_{PM} is found to be 0.9950, 0.9954, 0.9948 and 0.9846, respectively. However, the R^2 scores are omitted in the Table IV, since non of the existing works present this measure. The performances of the proposed models on the test profile are compared to the actual target values in Fig. 7 and 8.

V. CONCLUSION

This work proposed 1-D CNN-based model and an LSTM-based architectures to undertake temperature prediction of the internal elements of permanent magnet synchronous motors. In the initial stage, EWMA and EWMS were utilized for attribute conditioning and generating sufficient features. Experimental results show that CNN model can predict the desired target values with high precision and an average MSE within 2.64 $°C^2$. It is a significant improvement in comparison with former studies. The introduced data-driven temperature estimation method shows great promise that can be a viable replacement for classical methods. From an application prospective, the proposed models can be generalized to other type of electric motors and even other components of a system, where the thermal stress prediction is a paramount safety concern. However, this paper has not investigated the generalizability of the model and it is left for future work. It is also worth analyzing the effect of individual predictors and their respective exponentially weighted history on the performance of the proposed models.

REFERENCES

[1] O. Wallscheid, A. Specht, and J. Böcker, "Observing the permanent-magnet temperature of synchronous motors based on electrical fundamental wave model quantities," *IEEE Trans. Industrial Electronics*, vol. 64, no. 5, pp. 3921–3929, 2017.

[2] A. J. Grobler, S. R. Holm, and G. van Schoor, "A two-dimensional analytic thermal model for a high-speed pmsm magnet," *IEEE Trans. Industrial Electronics*, vol. 62, no. 11, pp. 6756–6764, 2015.

[3] H. Jo, H. J. Hwang, D. Phan, Y. Lee, and H. Jang, "Endpoint temperature prediction model for ld converters using machine-learning techniques," in *2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA)*, pp. 22–26, IEEE, 2019.

[4] X. Wang, "Ladle furnace temperature prediction model based on large-scale data with random forest," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 770–774, 2016.

[5] X. Su, S. Zhang, Y. Yin, and W. Xiao, "Prediction model of hot metal temperature for blast furnace based on improved multi-layer extreme learning machine," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 10, pp. 2739–2752, 2019.

[6] A. Zhukov, N. Tomin, V. Kurbatsky, D. Sidorov, D. Panasetsky, and A. Foley, "Ensemble methods of classification for power systems security assessment," *App. comp. & info.*, vol. 15, no. 1, pp. 45–53, 2019.

[7] S. Hosseini, G. Roshani, and S. Setayeshi, "Precise gamma based two-phase flow meter using frequency feature extraction and only one detector," *Flow Measurement and Instrumenta.*, vol. 72, p. 101693, 2020.

[8] T. Akilan, Q. J. Wu, Y. Yang, and A. Safaei, "Fusion of transfer learning features and its application in image classification," in *2017 IEEE 30th Canadian Conf. on Elec. and Compu. Engineering*, pp. 1–5, 2017.

[9] S. Hosseini, O. Taylan, M. Abusurrah, T. Akilan, E. Nazemi, E. Eftekhari-Zadeh, F. Bano, and G. H. Roshani, "Application of wavelet feature extraction and artificial neural networks for improving the performance of gas–liquid two-phase flow meters used in oil and petrochemical industries," *Polymers*, vol. 13, no. 21, p. 3647, 2021.

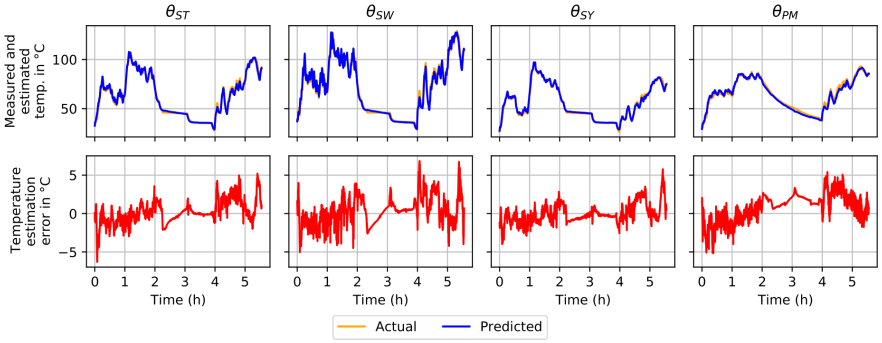

Fig. 7. Performance of the 1-D CNN model on Profile ID: 65.

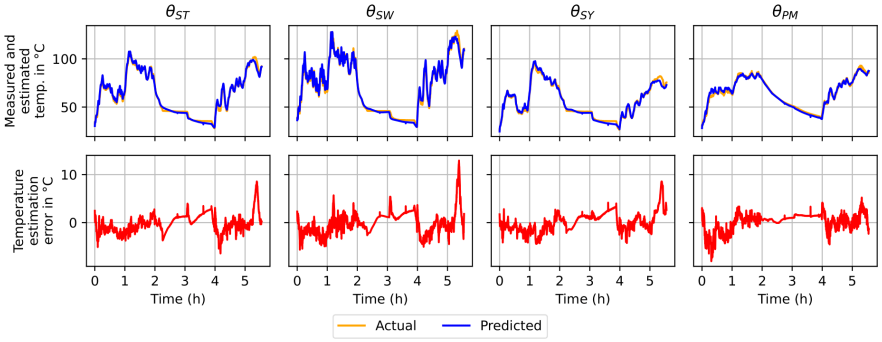

Fig. 8. Performance of the LSTM model on Profile ID: 65.

[10] M. Bahiraei, L. K. Foong, S. Hosseini, and N. Mazaheri, "Predicting heat transfer rate of a ribbed triple-tube heat exchanger working with nanofluid using neural network enhanced by advanced optimization algorithms," *Powder Technology*, vol. 381, pp. 459–476, 2021.

[11] M. Bahiraei, N. Mazaheri, and S. Hosseini, "Neural network modeling of thermo-hydraulic attributes and entropy generation of an ecofriendly nanofluid flow inside tubes equipped with novel rotary coaxial double-twisted tape," *Powder Technology*, vol. 369, pp. 162–175, 2020.

[12] S. Hosseini, S. Setayeshi, G. Roshani, A. Zahedi, and F. Shama, "Increasing efficiency of two-phase flowmeters using frequency-domain feature extraction and neural network in the detector output spectrum," *Journal of Modeling in Engineering*, vol. 19, no. 67, pp. 47–57, 2021.

[13] S. Hosseini, A. M. Iliyasu, T. Akilan, A. S. Salama, E. Eftekhari-Zadeh, and K. Hirota, "Accurate flow regime classification and void fraction measurement in two-phase flowmeters using frequency-domain feature extraction and neural networks," *Separations*, vol. 9, no. 7, p. 160, 2022.

[14] H. Li, "Deep learning for natural language processing: advantages and challenges," *National Science Review*, 2017.

[15] Y. Yang, Q. M. J. Wu, X. Feng, and T. Akilan, "Recomputation of the dense layers for performance improvement of dcnn," *IEEE Trans. on Pattern Analy. and Mach. Intelli.*, vol. 42, no. 11, pp. 2912–2925, 2020.

[16] M. Y. Rad and S. Shahbandegan, "An intelligent algorithm for mapping of applications on parallel reconfigurable systems," in *2020 6th Iranian Conf. on Sig. Process. and Intelli. Sys. (ICSPIS)*, pp. 1–6, 2020.

[17] M. Bahiraei, L. K. Foong, S. Hosseini, and N. Mazaheri, "Neural network combined with nature-inspired algorithms to estimate overall heat transfer coefficient of a ribbed triple-tube heat exchanger operating with a hybrid nanofluid," *Measurement*, vol. 174, p. 108967, 2021.

[18] W. Kirchgässner, O. Wallscheid, and J. Böcker, "Data-driven permanent magnet temperature estimation in synchronous motors with supervised machine learning: A benchmark," *IEEE Trans. Energy Conversion*, vol. 36, no. 3, pp. 2059–2067, 2021.

[19] C. Kasburg and S. F. Stefenon, "Deep learning for photovoltaic generation forecast in active solar trackers," *IEEE Latin America Transactions*, vol. 17, no. 12, pp. 2013–2019, 2019.

[20] N. Gui, J. Lou, Z. Qiu, and W. Gui, "Temporal feature selection for multi-step ahead reheater temperature prediction," *Processes*, vol. 7, no. 7, p. 473, 2019.

[21] O. Wallscheid, W. Kirchgässner, and J. Böcker, "Investigation of long short-term memory networks to temperature prediction for permanent magnet synchronous motors," in *Intl. joint conf. on neural networks*, pp. 1940–1947, IEEE, 2017.

[22] W. Kirchgässner, O. Wallscheid, and J. Böcker, "Estimating electric motor temperatures with deep residual machine learning," *IEEE Trans. Power Electronics*, vol. 36, no. 7, pp. 7480–7488, 2020.

[23] Z. Wang, J. Tian, H. Fang, L. Chen, and J. Qin, "Lightlog: A lightweight temporal convolutional network for log anomaly detection on the edge," *Computer Networks*, p. 108616, 2021.

[24] J. Lee and J.-I. Ha, "Temperature estimation of pmsm using a difference-estimating feedforward neural network," *IEEE Access*, vol. 8, pp. 130855–130865, 2020.

[25] J. Li and T. Akilan, "Global attention-based encoder-decoder lstm model for temperature prediction of permanent magnet synchronous motors," *https://arxiv.org/abs/2208.00293*, 2022.

[26] T. Akilan, Q. J. Wu, A. Safaei, J. Huo, and Y. Yang, "A 3d cnn-lstm-based image-to-image foreground segmentation," *IEEE Trans. Intelligent Transportation Systems*, vol. 21, no. 3, pp. 959–971, 2020.

[27] J. Zhao, X. Mao, and L. Chen, "Speech emotion recognition using deep 1d & 2d cnn lstm networks," *Biomed. sig. process. and cont.*, vol. 47, pp. 312–323, 2019.

[28] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm networks," in *Proceedings. 2005 IEEE Intl, Joint Conf. on Neural Networks, 2005.*, vol. 4, pp. 2047–2052, IEEE, 2005.

[29] K. Chen, Y. Zhou, and F. Dai, "A lstm-based method for stock returns prediction: A case study of china stock market," in *2015 IEEE international conference on big data (big data)*, pp. 2823–2824, IEEE, 2015.

[30] S. Bhanja and A. Das, "Impact of data normalization on deep neural network for time series forecasting," *arXiv:1812.05519*, 2018.

[31] E. G. Gedlu, O. Wallscheid, and J. Böcker, "Permanent magnet synchronous machine temperature estimation using low-order lumped-parameter thermal network with extended iron loss model," in *The 10th International Conference on Power Electronics, Machines and Drives*, vol. 2020, pp. 937–942, IET, 2020.