Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 58-19 (2024) 1258–1263

Using a Minimalist Bi-LSTM for Multi-Faceted Bearing Fault Detection ⋆

Alexandros Noussis * Ahmed Saif * Abdelhakim Khatab ** Claver Diallo *

* Dalhousie University, Industrial Engineering, Halifax, Nova Scotia, Canada (e-mail: {Alex.Noussis,Ahmed.Saif,Claver.Diallo}@dal.ca)
** Laboratory of Industrial Engineering, Production and Maintenance (LGIPM), Lorraine University, Metz, France (e-mail: abdelhakim.khatab@univ-lorraine.fr)

Abstract: A bidirectional long short-term memory (Bi-LSTM) model is developed to predict a multi-faceted bearing health characteristic using time series vibration measurements from the Case Western Reserve University (CWRU) seeded fault test data. A maximal amount of data is applied with minimal preprocessing to test the model's capabilities and several hyperparameters are tuned to maximize model performance. The utility of the CWRU dataset for related problems and the applicability of Bi-LSTM architecture for time series data is highlighted. The proposed model achieved a final test prediction accuracy of 98.42% and had low computation time, making it an interesting candidate for application in bearing fault prognosis.

Keywords: Data-driven machine prognostics, Fault detection, Machine learning, Bi-LSTM

1. INTRODUCTION

Leveraging modern machine learning (ML) techniques for asset health monitoring is a growing trend in both research and practice (Polverino et al., 2023; Wang et al., 2019; Deutsch and He, 2017). As with other applications, it is crucial to select an ML model that captures inherent characteristics of the data to be used. Monitoring sensors in mechanical systems usually produce time series data (TSD) with temporal interdependencies to be accounted for. Recurrent neural networks (RNNs) are oft employed when working with TSD, though a drawback is their limited memory capacity, which causes long-run relationships between points to decay as the amount of data increases. To overcome this limitation when working with large TSD sequences, long short-term memory models (LSTMs) are preferred. In particular, bidirectional LSTMs (Bi-LSTMs), which analyze data sequences in both forward and reverse directions, have proven effective for capturing vast amounts of information about TSD trends over long periods. In data-driven machine prognostics, given the large quantities of TSD produced by various system sensors, application of Bi-LSTMs seems promising.

Accordingly, this paper aims to test the performance of a simplified Bi-LSTM model developed for multi-faceted bearing fault detection, and compare it against other benchmark ML methods proposed in literature, such as in Zhuang et al. (2019) and Han et al. (2021). For this purpose, a popular dataset for testing mechanical system health monitoring methodologies, known as the *Case Western Reserve University seeded fault test data* (or the *CWRU dataset*, for short), is used (Bearing-Data-Center, n.d.). The dataset consists of time series vibration readings from a motor system that was run multiple times with different bearings installed and at different motor loads. Bearings tested were subjected to intentional damage to create distinct types of faults with several characteristics. The CWRU dataset has frequently been used to test the ability of ML methods to detect vibration differences based on bearing health characteristics. However, only a few Bi-LSTMs have been tested on this dataset, with most suffering from serious limitations such as omission of valuable data subsets and extensive preprocessing.

In contrast to previous studies that used Bi-LSTMs for bearing fault detection, a maximal amount of data from the CWRU dataset is used to train the proposed model and data is subjected to minimal preprocessing. Maximization of data ensures variety in the types of faults present and emulates the volume of data an ML model would need to analyze in a real-world scenario. These added complexities are important for testing Bi-LSTM resilience and robustness. Preprocessing was minimized for several reasons. As noted by Engel et al. (2013), preprocessing can result in important data attributes being obscured, insignificant data characteristics being assigned importance by the ML model, and preprocessing taking unreasonably long amounts of time to perform. Furthermore, preprocessing may not be feasible in real-time industrial scenarios where machine health is being constantly monitored.

This paper's model is trained to predict which bearing health characteristics are associated with sequences of data points from vibration tests. Several model hyperparameters are tuned prior to obtaining final results. The results obtained demonstrate the flexibility and robustness of Bi-LSTM architecture for mechanical system health monitor-

⋆ We acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC) and Nova Scotia Graduate Scholarship (NSGS) program.

10.1016/j.ifacol.2024.09.071

ing, which is the main contribution of this paper. Arguably, superior performance of the proposed model is achieved by using a greater diversity and volume of data from the CWRU dataset than benchmark models such as Tong et al. (2021) and Shenfield and Howarth (2020). Furthermore, it seems that restricting the model to rely solely on a Bi-LSTM layer for classifications while using raw data has led to the best outcomes.

The remainder of this paper is organized in four sections. Section 2 reviews the literature on ML models for bearing fault detection, particularly those using the CWRU dataset. Section 3 begins with a description of the dataset and how it was cleaned before detailing the development of the model. Hyperparameter tuning and the results of the finalized model are described in Section 4. Lastly, the outcomes of this paper's model and conclusions are discussed in Section 5.

2. LITERATURE REVIEW

Several papers have reviewed ML applications in bearing fault detection and classification. Mushtaq et al. (2021) reviewed signal processing, classical ML, and deep learning techniques employed in bearing fault diagnosis. A notable characteristic of many papers in this review was the prevalence of the CWRU dataset. Of 52 papers reviewed, 30 employed the CWRU dataset for training and testing. Papers using RNNs were scant in prior reviews, with no papers noted in Mushtaq et al. (2021) and only four noted in Neupane and Seok (2020). A lack of research into RNNs, particularly for the CWRU dataset, is unusual given that the dataset is comprised of TSD, for which RNNs are often considered well-suited.

Among the studies that have employed RNN models when working with the CWRU dataset are those of Liu et al. (2018), Wang et al. (2023), and Shenfield and Howarth (2020), which used the gated recurrent unit architecture. In contrast, Zhang et al. (2021), Yuhai et al. (2018), Jiang et al. (2018), Yu et al. (2019), and Zhuang et al. (2019) used unidirectional LSTM models and architecture, whereas Bi-LSTM models were utilized by Sun and Wang (2023), Zou et al. (2021), Liang et al. (2020), Choudakkanavar and Mangai (2022), Tong et al. (2021), Han et al. (2021), Qi et al. (2023), Xu et al. (2020), and Liu et al. (2021). Notably, not all these papers used the RNN architecture for fault diagnosis, with some employing it instead for vibration signal prediction and modeling. Additionally, some papers used the concept behind RNN architecture, rather than the architecture itself. The performance of past models will be used as a benchmark to evaluate the model proposed herein.

Three major gaps were identified when examining the above literature. First, many authors tended to avoid providing complete details on hyperparameter settings used with their ML models. Common features to mention would be: the number of hidden units in LSTM and Bi-LSTM layers, the optimizer, loss function, learning rate, number of epochs allowed, and epoch batch size. While these details were included in papers like Han et al. (2021) and Qi et al. (2023), other papers like Sun and Wang (2023) and Tong et al. (2021) did not readily provide all of this information. As such, there is a lack of clarity on standard hyperparameter tuning options for the problem. This paper provides a clear description of hyperparameters tuned during model refinement and the final settings for all relevant hyperparameters.

Second, all papers reviewed either heavily preprocessed the CWRU dataset before in their ML model or culled most of the data. Often, only around 10 out of 29 bearing health classes would be considered. Some papers also failed to clarify what parts of the dataset were used. Table 1 highlights how the CWRU dataset was used in articles. Given these decisions in data usage, flexibility and robustness of Bi-LSTM-based models has not been properly examined in the context of the CWRU dataset. This paper minimizes preprocessing efforts for the dataset and retains all data containing the standard formatting traits and attributes. As will be outlined in Section 3, 27 bearing health classes were considered.

The third, and final, gap is the lack of understanding of how Bi-LSTM architecture impacts model performance. Most papers reviewed either augmented their model performance by including additional types of architecture or preprocessed the data with unique algorithms. As such, advantages and disadvantages of a "pure" Bi-LSTM architecture, when not augmented with additional architecture or heavily modified input data, are not well documented. This paper restricts preprocessing to only the addition of a bearing class attribute and the removal of undesirable data while also restricting model architecture to only contain a singular Bi-LSTM layer.

3. METHODOLOGY

The current section consists of two parts. In subsection 3.1, the CWRU dataset is described, along with the work performed to clean and consolidate the data. Subsection 3.2 briefly describes data preprocessing, then outlines the basic architecture of the Bi-LSTM and how its hyperparameters are tuned.

3.1 Dataset Description and Cleaning

Dataset Description The dataset selected for this project is the CWRU dataset (Bearing-Data-Center, n.d.). Ball bearings were placed on the drive and fan ends of a test stand motor shaft. The test stand system, depicted in Figure 1, contained a two-horsepower motor, torque transducer, dynamometer, and control electronics. For each test, the system was run at a specified motor load, and accelerometers connected to the drive end, fan end, and base motor housing recorded vibrations as time series data. Data was recorded at rates of 12,000 and 48,000 samples per second (Bearing-Data-Center, n.d.).

Bearings tested in this system contained faults created by an electro-discharge machine. For each bearing, a fault was created either in the inner raceway, ball, or outer raceway. The fault diameter for each bearing was between 7 and 28 mils. Once a fault was created, the bearing was placed in the test system, either on the drive or fan end, and the motor was run. Bearings with an outer raceway fault also had the fault's positioning on the motor shaft oriented to either the 6:00, 3:00, or 12:00 positions relative to the load

Table 1. CWRU dataset use in academic works

References	**Denoising, Cleansing**	**Excluded Data**	**Unused Loads (HP)**	**Unused Damage Types**	**Unused Fault Sizes (mils)**
Zhang et al. (2021)	Yes	Und.	1,2,3	O3, O12	28
Wang et al. (2023)	Yes	48k-DE, Uncl.	Uncl.	Uncl.	28
Shenfield and Howarth (2020)	No	12k-DE, 12k-FE	0	Uncl.	28
Yuhai et al. (2018)	Yes	Uncl.	Uncl.	Uncl.	Uncl.
Jiang et al. (2018)	Yes	Uncl.	1,3	O3, O12	None
Yu et al. (2019)	No	48k-DE, Uncl.	0,2,3	None	28
Zhuang et al. (2019)	No	48k-DE, Uncl.	None	Uncl.	28
Sun and Wang (2023)	Yes	48k-DE, 12k-FE	1,2,3	Uncl.	28
Zou et al. (2021)	Yes	48k-DE, 12k-FE	1,2,3	Uncl.	28
Choudakkanavar and Mangai (2022)	No	48k-DE, 12k-FE	0,2,3	O3	None
Liang et al. (2020)	Yes	Und., 48k-DE, 12k-FE	1,2,3	Inner, Ball, O3, O12	14, 21, 28
Tong et al. (2021)	No	48k-DE, 12k-FE	1,2,3	O3, O12	28
Han et al. (2021)	No	48k-DE, 12k-FE	None	Uncl.	28
Qi et al. (2023)	Yes	48k-DE, 12k-FE	None	Uncl.	28
Xu et al. (2020)	Yes	48k-DE	0,1,2	Uncl.	Uncl.
Liu et al. (2021)	Yes	48k-DE, 12k-FE	None	Uncl.	14, 21, 28
This Paper	No	48k-DE	None	None	28

Uncl.: Unclear, **Und.**: Undamaged, **12k-DE**: 12k Drive End, **48k-DE**: 48k Drive End, **12k-FE**: 12k Fan End, **O3**: Outer 3:00, **O12**: Outer 12:00

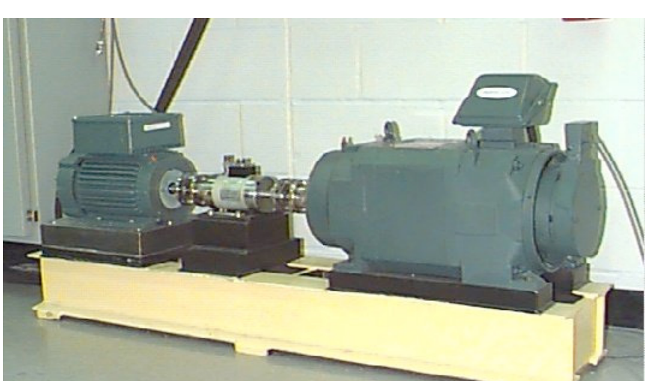

Fig. 1. CWRU system setup (Bearing-Data-Center, n.d.)

zone. Undamaged bearings were also tested to produce baseline data (Bearing-Data-Center, n.d.).

Dataset Cleaning All data was gathered from the CWRU dataset website with some contained in MATLAB files and some on the website itself. The splitting of data across different formats resulted in a lack of consolidation. As such, all vibration and bearing health data was transferred into MS Excel spreadsheets. Inconsistencies in certain features and tests led to the following data being omitted:

- Vibration data collected from the base accelerometer.
- Data collected at a frequency of 48,000 samples/sec.
- Data collected for bearings with faults of 28 mils.

Each test's bearing health characteristics were consolidated into a singular qualitative feature known as bearing class. A sample of the legend for the bearing classes is provided in Table 2. Each combination of health characteristics corresponds to a different bearing class the Bi-LSTM had to consider in predictions.

Table 2. Sample of bearing health class list

Bearing Class Number	**Damaged Bearing Location**	**Damaged Component**	**Fault Position (if OR)**	**Fault Size (mils)**
1	UND	UND	None	None
2	DE	IR	None	7
3	DE	IR	None	14
...	...	...	...	...
25	FE	OR	3:00	14
26	FE	OR	3:00	21
27	FE	OR	12:00	7

UND: Undamaged, **DE**: Drive End, **FE**: Fan End, **IR**: Inner Raceway, **OR**: Outer Raceway

When all cleaning and consolidation was complete, each data point consisted of four features: the drive end vibration measurement, the fan end vibration measurement, the test's motor load, and the bearing class attribute to be predicted by the Bi-LSTM model. In total, 27 of the 29 bearing health classes were employed for testing in this paper. The final version of the CWRU dataset used was spread across 7 Excel files and had a total file size of 443.96 MB. Across the files, 101 sheets contained a total of approximately 13.4 million rows, each with four columns.

3.2 Bi-LSTM Model Development

Data Preparation The only data preparation carried out was categorical feature encoding for the bearing class attribute and sampling of data points. The bearing class attribute was encoded in the one-hot format. For sampling, data had to be grouped in sequences of consecutive data points to retain their temporal relationships. Moreover, given the large volume of data available, data was sampled from across the entire timespan of each bearing trial. This was done to enable the best chance of capturing signal characteristics unique to a bearing class. Two decisions had to be made when sampling data: how many data points (time steps) to include in each sample sequence and what proportion of the data to sample in total.

Model Architecture. Given the large volume of data being used and the high likelihood of large data sequences being sampled, the Bi-LSTM architecture was a sensible and reasonable choice given its ability to retain long-run relationships between data points and its capturing of data sequence information in both directions. The final model architecture consisted of three layers:

(1) An input layer of the shape $(3 \times t)$ where t is the number of data points in each sample sequence. Each point in a sequence contained three features.
(2) A Bi-LSTM layer with an adjustable number of hidden units.
(3) A dense output layer with as many units as there were bearing health classes to consider (27 units).

The following hyperparameters were selected at the outset with no additional fine tuning:

(1) The number of Bi-LSTM layers was one.
(2) The standard tanh and sigmoid functions were used in the Bi-LSTM layer.
(3) The output layer's activation function was softmax.
(4) The loss function was categorical cross-entropy.
(5) The maximum number of epochs allowed during training was 100.

Early stopping was also implemented for model training to ensure the model did not overfit to training data nor waste time producing insignificant performance improvements. The validation dataset loss function value (categorical cross-entropy) was used as the monitored function, patience was set to 5 epochs, and the model was permitted to restore the best weights found during training.

4. MODEL TUNING AND EVALUATION

This section presents the tuning of the model's hyperparameters. Then, the final model's performance is highlighted and compared against the performance of models from prior literature. The primary metrics are model accuracy (AC) and computation time (CPUt), though precision and recall are also considered. All tests were conducted using an HP Omen laptop with an 11th Generation Intel Core i7-11800H @ 2.30 GHz processor, 16.0 GB of RAM, and a NVIDIA GeForce RTX 3060 Laptop GPU running a 64-bit version of Windows 11 Home.

Model accuracy, in this case, represents the proportion of instances where the ML model correctly predicted all bearing health characteristics (i.e., correctly assigned the bearing health class) for a bearing based on a sample of time series data points. Incorrectly predicting a single health characteristic is classified as a failure. Precision and recall are also based on whether the bearing class representing all health characteristics was correctly predicted, rather than based on the accuracy relating to individual health characteristics.

4.1 Hyperparameter Tuning

Seven hyperparameters were selected to be tuned with the goal of maximizing the performance of the Bi-LSTM model. The settings and values tested for each hyperparameter are as follows:

(1) Optimizer algorithm : stochastic gradient descent, RMSprop, Adam, Adadelta, Adagrad, Adamax, and Nadam optimizers with their standard hyperparameters.
(2) Learning rate: 0.000001, 0.00001, 0.0001, 0.001, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.011, 0.012, 0.013, 0.014, 0.015, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, and 1.
(3) Number of hidden units in Bi-LSTM layer: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 250, and 500.
(4) Batch size per epoch: 8, 16, 32, 64, 128, and 256.
(5) Early stopping minimum improvement amount (minimum delta): 0.000001, 0.000003, 0.000005, 0.000008, 0.00001, 0.000025, 0.00005, 0.000075, 0.0001, 0.001, 0.01, 0.1, and 1.
(6) Number of time steps (consecutive data points) captured in each sample sequence: 5, 10, 25, 50, 100, 150, 200, 250, 300, 500, and 1000.
(7) Proportion of available data sampled: 0.05, 0.1, 0.25, 0.35, and 0.5.

Tuning was performed manually and was approached similar to a greedy heuristic with each hyperparameter being individually tuned to maximize model performance. After an ideal hyperparameter value was determined, it was used in the tuning that followed for other hyperparameters. When maximizing model performance, the metrics considered were the categorical cross-entropy (loss function) value and accuracy for the test dataset, along with the computation time to train and test the model. Computation time was only considered if the most accurate hyperparameter setting resulted in significantly larger times. The hyperparameters and decisions selected for the final version of the model were as follows:

- Data sampling hyperparameters and decisions:
  - Time steps in sample sequence = 200
  - Proportion of available data sampled = 0.5
  - Training/Test/Validation sample split: 70/20/10
  - Shuffling and stratification implemented in data splitting
- Model architecture hyperparameters and decisions:
  - Model input shape: 200 by 3
  - Number of Bi-LSTM layers = 1
  - Number of hidden units in Bi-LSTM layer = 60
  - Bi-LSTM activation functions: standard tanh and sigmoid
  - Number of nodes in dense output layer = 27
  - Dense output layer activation function: standard softmax
- Model training hyperparameters and decisions:
  - Optimizer selected: RMSprop
  - Learning rate = 0.01
  - Loss function: categorical cross-entropy
  - Batch size = 64
  - Maximum number of epochs = 100
  - Early stopping implemented with restoration of best weights
  - Early stopping monitoring function: categorical cross-entropy for validation data
  - Early stopping minimum delta = 0.00005
  - Early stopping patience = 5 epochs

4.2 Model Evaluation

Following hyperparameter tuning, the finalized model was trained and tested once more to visualize model progress and determine its performance. The model progress across its training epochs can be viewed in Figure 2.

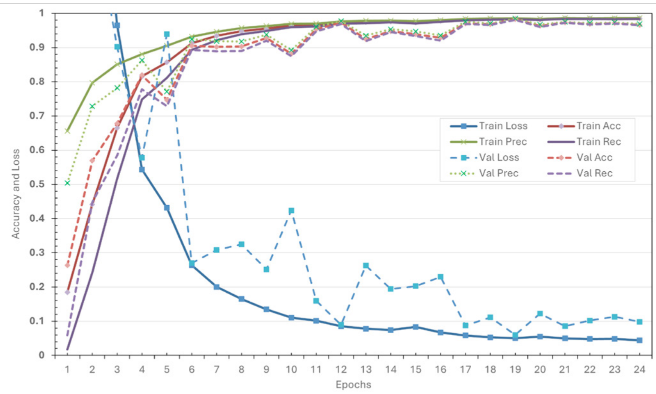

Fig. 2. Model training progress

The finalized model's performance with test data can be summarized as follows:

- A minimum categorical cross-entropy value of 0.0546 was achieved.
- A maximum accuracy of 0.9842 was achieved.
- A maximum precision of 0.9859 was achieved.
- A maximum recall of 0.9830 was achieved.
- The model took 485.6 seconds to output its prediction results for the test data.

Most results produced in this final test are in line with those observed during hyperparameter tuning. The following additional observations can be made on the model's performance:

- The majority of model performance improvements occurred in the first five training epochs with diminishing returns beyond that point.
- Model training and validation metrics closely mirrored each other during training, indicating that overfitting did not occur.
- The model test accuracy of 0.9842 indicates that underfitting did not occur.
- Precision, recall, and the F1 score for each bearing health class was above 0.9.
- The model did not incorrectly label any damaged bearings as being undamaged.

Table 3 presents the performance of all models that did not employ denoising or cleansing when using the CWRU data. As can be observed, the proposed model yielded among the highest accuracy values (AC%) and had the second-best computation time (CPUt) with a "minimalist" Bi-LSTM model, while using a less preprocessed and more complete dataset than the other models. Thus, the proposed model offers good trade-offs for industrial applications.

Table 3. Performance of academic models using the CWRU dataset

Method	AC (%)	CPUt(s)
RNN-WDCNN (Shenfield and Howarth, 2020)	96.2	3190 – 6389
Stacked LSTM (Yu et al., 2019)	90–100	–
SRDCNN (Zhuang et al., 2019)	**99.4**	–
1D-CNN-Bi-LSTM (Choudakkanavar and Mangai, 2022)	98.2	–
DRSN+Bi-LSTM (Tong et al., 2021)	**99.6**	577
BLC-CNN (Han et al., 2021)	85–100	**312**
Bi-LSTM (This Paper)	**98.4**	**486**

AC: Accuracy, **CPUt**: Computation Time

5. DISCUSSION AND CONCLUSION

This paper aimed to demonstrate the applicability of Bi-LSTM models in mechanical system health monitoring by using the CWRU Seeded Fault Test Data. The literature review highlighted both the under-utilization of the CWRU dataset and the obfuscation of Bi-LSTM architecture performance in prior academic work. A minimalist Bi-LSTM model was developed to take in sequences of consecutive data points from the CWRU dataset containing vibration and motor load data with the goal of predicting bearing fault characteristics. Data preprocessing was restricted to only the introduction of an additional attribute for class labelling and the removal of inconsistent data. Furthermore, a maximal amount of data from the CWRU dataset was employed with 27 combinations of bearing health characteristics used as classes and approximately 13.4 million rows of data being available for analysis. Following hyperparameter tuning, the model produced an accuracy of 98.4% and a computation time of 485.7 seconds. Accuracy was in line with the performance of models examined in the literature and computation time was significantly shorter than that of most other models.

As noted in Section 2, most academic work either performed feature modification and extraction or culled the dataset to contain only data from a small subset of test scenarios. Past models heavily simplified the problem and omitted realistic challenges present in the data. The proposed model was required to simultaneously predict whether a bearing was damaged, the location of the damaged bearing, type of damage present, size of the fault, and (if applicable) position of the fault relative to the motor shaft. This paper's model achieved such performance levels while heavily limiting preprocessing, restricting architecture to a single Bi-LSTM layer, and greatly increasing the amount of data and number of classes under consideration. As such, the model can be considered an improvement to prior academic models and could be further enhanced with more thorough hyperparameter tuning and consideration of uncertainty around the predictions. Such strong performance can likely be attributed to the use of a large volume of training data and the choice to have large data sequences used for sampling. This paper's model provides the novel contributions of demonstrating the strength and flexibility of Bi-LSTM architecture when handling complex TSD and the lack of necessity for major data preprocessing on the CWRU dataset. Notable limitations of this work are:

(1) The model was trained and tested on a laptop not inherently designed for ML work. Better computation times may be possible with a more suitable setup.
(2) Hyperparameter tuning was constrained by time. Average model performance for each hyperparameter tuning option was only based on five trials. Furthermore, only a small set of values were tested for

each hyperparameter. A thorough testing of values for a wider array of hyperparameters may yield better accuracy and computation times.

(3) Hyperparameter impacts to computation times appeared to have been distorted due to system difficulties during tuning. Clearing of caches and system cooldowns between trials may help mitigate the issue.

Future research avenues include:

- further exploration of hyperparameter impacts on the proposed model via additional tuning,
- examination of ways to improve model performance through alternative optimization methods and additional architectures, and
- implementation of Monte Carlo Dropout to build probabilistic distributions around predictions and allow more realistic decision-making.

ACKNOWLEDGEMENTS

The authors thank Dr. Issam Hammad for his assistance with methodology refinements during the coding phase.

REFERENCES

Bearing-Data-Center (n.d.). Seeded fault test data. `https://engineering.case.edu/bearingdatacenter`. Online; accessed February 12, 2024.

Choudakkanavar, G. and Mangai, J.A. (2022). A hybrid 1D-CNN-Bi-LSTM based model with spatial dropout for multiple fault diagnosis of roller bearing. *International Journal of Advanced Computer Science and Applications*, 13(8).

Deutsch, J. and He, D. (2017). Using deep learning-based approach to predict remaining useful life of rotating components. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(1), 11–20.

Engel, J., Gerretzen, J., Szymańska, E., Jansen, J.J., Downey, G., Blanchet, L., and Buydens, L.M. (2013). Breaking with trends in pre-processing? *TrAC Trends in Analytical Chemistry*, 50, 96–106.

Han, T., Ma, R., and Zheng, J. (2021). Combination bidirectional long short-term memory and capsule network for rotating machinery fault diagnosis. *Measurement*, 176, 109208.

Jiang, H., Li, X., Shao, H., and Zhao, K. (2018). Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network. *Measurement Science and Technology*, 29(6), 065107.

Liang, J., Wang, L., Wu, J., Liu, Z., and Yu, G. (2020). Elimination of end effects in LMD by Bi-LSTM regression network and applications for rolling element bearings characteristic extraction under different loading conditions. *Digital Signal Processing*, 107, 102881.

Liu, H., Zhou, J., Zheng, Y., Jiang, W., and Zhang, Y. (2018). Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA transactions*, 77, 167–178.

Liu, H., Zhao, H., Wang, J., Yuan, S., and Feng, W. (2021). LSTM-GAN-AE: A promising approach for fault diagnosis in machine health monitoring. *IEEE Transactions on Instrumentation and Measurement*, 71, 1–13.

Mushtaq, S., Islam, M.M., and Sohaib, M. (2021). Deep learning aided data-driven fault diagnosis of rotatory machine: A comprehensive review. *Energies*, 14(16), 5150.

Neupane, D. and Seok, J. (2020). Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. *IEEE Access*, 8, 93155–93178.

Polverino, L., Abbate, R., Manco, P., Perfetto, D., Caputo, F., Macchiaroli, R., and Caterino, M. (2023). Machine learning for prognostics and health management of industrial mechanical systems and equipment: A systematic literature review. *International Journal of Engineering Business Management*, 15, 18479790231186848.

Qi, B., Li, Y., Yao, W., and Li, Z. (2023). Application of EMD combined with deep learning and knowledge graph in bearing fault. *Journal of Signal Processing Systems*, 1–20.

Shenfield, A. and Howarth, M. (2020). A novel deep learning model for the detection and identification of rolling element-bearing faults. *Sensors*, 20(18), 5112.

Sun, Y. and Wang, H. (2023). Study of diagnosis model for rotating machinery in advanced nuclear reactor based on deep learning model. *Frontiers in Energy Research*, 11, 1210703.

Tong, Y., Wu, P., He, J., Zhang, X., and Zhao, X. (2021). Bearing fault diagnosis by combining a deep residual shrinkage network and bidirectional LSTM. *Measurement Science and Technology*, 33(3), 034001.

Wang, H., Guo, Y., Liu, X., Yang, J., Zhang, X., and Shi, L. (2023). Fault diagnosis method for imbalanced data of rotating machinery based on time domain signal prediction and sc-resnest. *IEEE Access*, 11, 38875–38893. doi:10.1109/access.2023.3264636.

Wang, J., Mo, Z., Zhang, H., and Miao, Q. (2019). A deep learning method for bearing fault diagnosis based on time-frequency image. *IEEE Access*, 7, 42373–42383.

Xu, X., Tao, Z., Ming, W., An, Q., and Chen, M. (2020). Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion. *Measurement*, 165, 108086.

Yu, L., Qu, J., Gao, F., Tian, Y., et al. (2019). A novel hierarchical algorithm for bearing fault diagnosis based on stacked LSTM. *Shock and Vibration*, 2019.

Yuhai, G., Shuo, L., Linfeng, H., et al. (2018). Research on failure prediction using dbn and lstm neural network. In *2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, 1705–1709. IEEE.

Zhang, X., Zhao, B., and Lin, Y. (2021). Machine learning based bearing fault diagnosis using the case western reserve university data: a review. *IEEE Access*, 9, 155598–155608.

Zhuang, Z., Lv, H., Xu, J., Huang, Z., and Qin, W. (2019). A deep learning method for bearing fault diagnosis through stacked residual dilated convolutions. *Applied Sciences*, 9(9), 1823.

Zou, F., Zhang, H., Sang, S., Li, X., He, W., and Liu, X. (2021). Bearing fault diagnosis based on combined multi-scale weighted entropy morphological filtering and bi-LSTM. *Applied Intelligence*, 1–18.