ORIGINAL ARTICLE

Open Access

A deep learning-based approach for electrical equipment remaining useful life prediction



Huibin Fu¹ and Ying Liu^{1*}

Abstract

Electrical equipment maintenance is of vital importance to management companies. Efficient maintenance can significantly reduce business costs and avoid safety accidents caused by catastrophic equipment failures. In the current context, predictive maintenance (PdM) is becoming increasingly popular based on machine learning approaches, while its research on electrical equipment such as low-voltage contactors is in its infancy. The failure modes are mainly fusion welding and explosion, and a few are unable to switch on. In this study, a data-driven approach is proposed to predict the remaining useful life (RUL) of the low-voltage contactor. Firstly, the three-phase alternating voltage and current records the life of electrical equipment by tracking the number of times it has been operated. Secondly, the failure-relevant features are extracted by using the time domain, frequency domain, and wavelet methods. Then, a CNN-LSTM network is designed and used to train an electrical equipment RUL prediction model based on the extracted features. An experimental study based on ten datasets collected from low-voltage AC contactors reveals that the proposed method shows merits in comparison with the prevailing deep learning algorithms in terms of MAE and RMSE.

Keywords: Predictive maintenance, Electrical equipment, Remaining useful life, Machine learning, Deep learning

1 Introduction

Equipment maintenance is vital if it is directly related to the lifespan of the equipment [1]. Nowadays, maintenance has been widely applied in different industries such as aircraft, smart manufacturing, assist robots, etc. A large number of equipment-related data are collected, stored, and combined with statistical methods and professional knowledge to maintain the equipment to help the enterprise make decisions [2, 3]. If equipment can be repaired in advance through appropriate methods, it can bring certain benefits [4, 5]. Therefore, improving the accuracy of prediction can further reduce the damage and enhance the tangible benefits to the companies [6].

*Correspondence: LiuY81@Cardiff.ac.uk

Recently, data-driven methods have been widely researched in PdM [7]. Among these researches, the attention to remaining useful life (RUL) prediction is continuously increasing [8]. Traditionally, RUL prediction reveals the degradation of equipment through the analysis of historical data. It studies the degradation trend through statistical methods and establishes the degradation curve and confidence interval. However, this method has some limitations. Due to the large sample size, it takes a long time to research the degradation trend manually. It may also record some key degradation information incorrectly, resulting in the low accuracy of the model. Machine learning as a prevailing tool has been widely investigated in different fields such as fault diagnosis [9], defect visual recognition [10], etc. At the same time, using machine learning to predict RUL also has different degrees of applications in industry, such as the application of bearings and turbine engines [11, 12]. By collecting the vibration signal of the



© The Author(s) 2022. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

¹Department of Mechanical Engineering, School of Engineering, Cardiff University, Cardiff CF24 3AA, UK

bearing, the model of health and sub-health is established and early warning is given to help companies' decisionmaking and maintain in advance [13, 14]. However, electrical equipment such as the low-voltage contactor is still needed to study. Due to a large number of components of electronic products and many influencing factors, the amount of monitoring that needs to be collected is also more than that of bearings, therefore, it is difficult to analyse, and in principle, the electrical equipment, such as the arc phase angle in the opening and closing process, has great randomness, which makes it difficult to analyse the degradation trend, and it is difficult to ensure the accuracy of the RUL model by using machine learning technique. Hence, it is worthwhile to explore deep learning techniques, such as convolutional neural network (CNN), and long short-term memory neural network (LSTM), to automatically learn the life degradation trend of electrical equipment and achieve better maintenance management. In addition, as the electrical equipment is accompanied by a large amount of data collection, the deep learning model needs to introduce some hidden layers, such as the TimeDistributed layer, to help reduce the data dimension and improve the model performance. On the other hand, due to the certain fluctuation of the waveform from arc starting to arc extinguishing from the data collected from several groups of different electrical equipment, the sliding window is used to smooth the relevant data and improve the accuracy of the model. The main contributions of the paper are: 1) The interval of feature extraction has a significant impact on the accuracy of the model. To improve the accuracy of the model, the data feature extraction is carried out for the arcing interval and closing interval; 2) Compared with the traditional mechanical equipment, the sample data of electrical equipment is more different, so a larger training set is needed to train the data to obtain the generalized model; 3) Due to the large amount of data collected by electrical equipment, machine learning is usually difficult to meet the requirements of fitting, so choosing to introduce the CNN-LSTM model can have higher accuracy in predicting the RUL of electrical equipment. The rest of the paper is organized as follows: the machine learning application, sliding window, and deep learning in RUL are reviewed in Sect. 2, Sect. 3 introduces the CNN-LSTM network, an experimental study is demonstrated in Sect. 4 and its results are demonstrated in Sect. 5. Finally, Sect. 6 discusses the results, and Sect. 7 concludes.

2 Literature review

There are many case studies for predicting the RUL of electrical equipment, including statistical models and machine learning models. In the statistical model, the traditional mechanism features, time domain, and frequency domain are widely used in predictive maintenance [15]. Meanwhile, with the rise of artificial intelligence, an increasing number of researchers have explored machine learning and deep learning to help companies maintain in advance the study of the RUL of electrical equipment [16].

2.1 Statistical approaches for predictive maintenance

Statistics is used to extract the mechanical feature of electrical equipment, including arcing time, arcing energy, arcing power, arcing power, etc. Electrical equipment can be identified mainly through the fluctuation of characteristic waveforms or meaningful intervals, such as arcing intervals and closing times. The feature interval of arcing interval and closing time is extracted, so the continuous arcing energy value on the equipment can be understood. In recent research, the arcing capacity has a positive influence on the RUL, the value of arcing energy is calculated through the continuous opening and closing when the released arcing energy is higher, and there is a trend of accelerated degradation of the RUL of the equipment at some time [17].

Time-domain and frequency-domain are the main statistical feature extraction methods of digital manufacturing technology, because they can analyse the results in a variety of ways and from different angles, and can directly reflect the relationship and influence of features, for example, sine wave as a functional form in the frequency domain has its special place. If sine waves are used, some problems related to the electrical areas of interconnects may become easier to understand and solve. On the other hand, wavelet analysis is essential for statistical extraction. The main advantage of wavelet analysis is that the signal can be analysed locally in any time or space domain. Wavelet analysis can discover the information of structural characteristics hidden in the data that cannot be recognized by other signal analysis methods, and these characteristics are particularly important for the identification of mechanical faults and material damage [18]. At present, there is no theoretical standard for selecting the wavelet basis function. There are 15 commonly used wavelet functions, including Haar, Daubechies, Morlet, Meryer, Symlet, Coiflet, Biorthogonal wavelet, etc. [14]. Due to there are many wavelet functions that can be selected and the purposes of wavelet transform are different, there is no general standard at present. According to practical application experience, the Morlet wavelet can be applied to a wide range of applications, including signal representation and classification, image recognition, and feature extraction; the Mexican wavelet is used for system recognition; the Spline wavelet is used for material flaw detection; and the Shannon orthogonal basis for solving differential equations. Hence, choosing wavelet analysis to extract the characteristics of current and voltage of electrical equipment can effectively discover the trend of life degradation. At present, the residual life prediction based on the statistical method needs to study the trend of equipment life degradation manually, which takes time and may not achieve high accuracy.



Another type of statistical approach is a sliding window, the collected data waveform fluctuates continuously in some intervals, selecting the sliding window algorithm can smooth the curve and improve the accuracy of the model. The sliding window can smooth the data curve by setting the window size and the mathematical function in the window. As illustrated in Fig. 1, assuming that the window is set to 4, each calculation will slide from left to right to form a new window, and the window setting function calculation results can effectively reflect the characteristics of the raw data.

2.2 Machine learning for predictive maintenance

Machine learning methods are currently one of the most effective methods in the research of PdM applications [19]. Machine learning is a part of artificial intelligence and is widely used in data analysis, data mining, and predictive analysis. With the development of technology, deep learning has become an increasingly important part of data mining and predictive analysis research [20].

A recurrent neural network (RNN) is popular for processing time-series data. Malhi et al. [21] proposed a method based on competitive learning to predict longterm machine health status. An LSTM network is an RNN, which is an improved RNN and also a deep learning model that can solve the problem of long-term memory that RNN cannot solve. For instance, by installing different sensors to collect signals, the health index data fusion technology based on the health index (HI) helps to understand the degradation process of the unit and predict its RUL. Although it is currently a hot research topic, the resume of degenerate models may be limited by making restrictive assumptions, such as fusing multi-sensor linear or kernel-based function model signals, or by preselecting basis functions. This assumption is generally invalid in industrial production. In practice, it may not be possible to accurately describe the degradation process of the complex relationship between multiple sensor signals and basic signals. By combining multiple sensor signals, a deep neural network (DNN) and LSTM models can better characterize degradation [22]. RUL prediction is also widely used in many intelligent systems. Establishing a deep LSTM-GAN model results in better model performance. Furthermore, the LSTM network can reduce gradient disappearance and mode collapse in the confrontation network, prevent the mode collapse of the generative adversarial network (GAN), and realize self-detection of abnormal data [23, 24]. In addition, a new model, called LSTM and statistical process analysis network (SPAN), is used to predict the residual of key components of the aero-engine bearing transmission system. To predict the multistage performance degradation of aero-engine bearings, it combines the advantages of LSTM. An algorithm based on this model is proposed. Through time and statistics, the characteristics of bearing vibration signals are extracted and divided into multiple stages. Then, multi-level signals are input into LSS for prediction. NASA and FEMTO-ST Research Institute have verified the effectiveness of this method. It has been demonstrated that the proposed method is feasible and has a higher prediction accuracy than recurrent neural network (RNN) and support vector regression (SVR) [25]. Based on the historical maintenance data and GIS data, Chen et al. [26] proposed a merged-LSTM network for the RUL prediction of the automobile. The RUL can be predicted from the service times of opening and closing switches collected by electrical equipment. Learning the hidden mode of these data through the LSTM network can improve the performance of the model and predict the RUL of the equipment [27].

On the other hand, a CNN is another classical deep learning model, it is found that their unique network structure can effectively reduce the complexity of a feedback

neural network, and then proposed a CNN. Now, CNN has become one of the research hotspots in many scientific fields, especially in the field of pattern classification. Because the network avoids the complex pre-processing of images and can directly input the original images, it has been more widely used. In essence, CNN is a multi-layer structure constructed by imitating the processing process of cellular visual information. CNN has many similarities with the ordinary neural network. They all imitate the structure of human nerves and are composed of neurons with learnable weights and bias constants. Each neuron can receive the input signal and output the score of each classification after calculation. However, the input of CNN is generally image. Convolution network successfully reduces the dimension of image recognition problem with a large amount of data through a series of methods, so that it can be trained finally [12]. In actual industrial scenarios, as one of the most significant functional components, the running performance of rolling bearings in the rotation process directly affects the equipment's reliability and safety. A robust structure based on a convolutional neural network (CNN) is proposed for the condition monitoring of rolling bearings because of the harsh and unpredictable working conditions. The normalised CNN model is proposed, which is derived from complete life cycle data and contains different state information and the training model is directly used for online monitoring of other rolling bearings [28]. A new intelligent residual algorithm called the deep learning-based useful life (RUL) prediction method is proposed by experiments on a popular rolling bearing data set prepared on the PRONOSTIA platform. The timefrequency domain information is explored, and the convolution neural network is used to realize multi-scale feature extraction. To prove its effectiveness, it is compared with traditional methods to prove its superiority. Generally speaking, RUL prediction has achieved high accuracy, and the proposed method is feasible and has broad industrial application prospects [29]. A deep separable convolutional network (DSCN) mechanical prediction for RUL. In the proposed digital signal processing network, the monitoring data collected by different sensors are directly used as the input of the prediction network. Then, based on separable convolution, compression and excitation operations, separable convolution building blocks with residual connections are constructed. Through stacking multiple separable convolution blocks, the high-level representation of the input data is automatically learned, and the RUL is estimated [30]. As compared with the prevailing deep time series models, when using CNN to predict RUL, the input data cannot be converted well into time series data due to the influence of the industrial environment. Compared with some deep learning time series models, the training speed and model accuracy cannot be well guaranteed [31].

3 Method: the CNN-LSTM network

A deep learning-based approach called CNN-LSTM neural network is proposed to establish the RUL prediction model based on collected data from electrical equipment. The approach consists of three stages. Firstly, by installing sensors and industrial computers, and then collecting data sets and pre-processing the data, it is necessary to extract the features of the data, including time domain, frequency domain, and wavelet packet. According to the research on the mechanical rationality of the equipment, the normal operation of the equipment has the start, end time of arcing, and opening and closing time. Therefore, the characteristics of the arcing section and the closing time are extracted. Because there are certain sudden changes in the arcing end voltage of different equipment, the time window function is set to extract various sudden changes, because there is a certain fluctuation in the local part of the data waveform, the sliding window algorithm is selected to smooth the curve. The detail of the sliding window is introduced in Sect. 3.1. Secondly, it is found from the data collected in the working environment that the difference in data is greatly affected by the industrial environment. Classify equipment in the same batch to improve the RUL model's prediction accuracy, and to classify and predict equipment in different environments. Finally, after classifying the equipment data of different batches according to different environments, select different data sets of each batch as the test set, and use the CNN-LSTM model to verify the robustness of the model and the differences between equipment, the flow chart of the proposed method is shown in Fig. 2.

3.1 Sliding window for predictive maintenance

To reduce the mutation rate of time series data, this method introduces the sliding window algorithm. Its purpose is to use the window function to smooth the data curve and filter the local abnormal points, to improve the accuracy of the model. In the sliding window strategy, the data sets with the different total life of electrical equipment are extracted by feature extraction, and the two-dimensional matrix containing **R** is used to select the appropriate window length for the features. Slide back through m instances. In addition, according to the research, the RUL of the equipment may have a certain relationship with cumulative wear [32]. Using the sliding window algorithm to set the window function is more advantageous and conveys more information than the single time point.

3.2 LSTM network

The most significant difference between the LSTM network and the RNN is that the hidden layer of the RNN has only one state h, while the hidden layer of the LSTM network has an extra long-term memory cell (denoted as





c) to store the long-term state [33]. Moreover, in order to control the long-term state **c**, the LSTM network designs three gates: the input gate, the forget gate, and the output gate. The schematic diagram of an LSTM cell is shown in Fig. 3. The three gates that control the state of each long-term memory cell are a fully connected layer. The input is a vector, and the output is a real vector between 0 and 1. It can be expressed as:

$$\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{a}(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}), \tag{1}$$

where **a** is the Sigmoid activation function.

 c_t is the state of the long-term memory cell, which can control the transfer of information to the next moment; h_t represents working memory or short-term memory [33]. The three yellow circles in Fig. 2 represent the forget gate, input gate, and output gate, respectively. Regarding the feed-forward process of the LSTM network, firstly, the long-term memory cell c_{t-1} discards some information through the forgetting gate f_t . f_t is controlled by the external input x_t at the current moment, the output h_{t-1} at the last moment, and the long-term memory state c_{t-1} at the last moment. Then the information \hat{c}_t at the current moment is calculated from x_t and h_{t-1} . Secondly, through the input gate i_t control, the part of the new information \hat{c}_t at the current moment is entered into the long-term memory cell to generate a new long-term memory c_t . Among them, i_t is controlled by x_t , h_{t-1} and c_{t-1} . Finally, activate the long-term memory cell c_t , which is controlled by the output gate o_t , and select some relevant information from the accumulated memory c_t to generate the memory h_t to be paid attention to at the current moment. Part of the memory is output to the next LSTM cell. The output gate o_t is controlled by x_t , h_{t-1} and the long-term memory cell c_t .

In the collected data, not only can the next RUL be predicted based on the current data, but can help improve the ability of decision-making [34].

3.3 CNN-LSTM network

Long-short-term memory network is the type of deep learning model which is well-known for processing the time series data, Convolutional neural network also is a popular deep learning model which can handle the 2D matrix data set. This experiment proposed a merge neural network called the CNN-LSTM network model to improve the accuracy of the model. The CNN-LSTM architecture involves using CNN layers for feature extraction on input data combined with LSTM to support sequence prediction. It needs to define a 2D convolutional network as comprised of Conv2D and MaxPooling2D layers ordered into a stack of the required depth. The Conv2D will interpret the matrix and the polling layers will consolidate or abstract the interpretation. However, the CNN model above is only capable of handling a single matrix and it may not handle the time series data set, a proposed method is that it need define the CNN model first, then add it to the LSTM model by wrapping the entire sequence of CNN layers in a TimeDistributed layer. For the collected data on electrical equipment. It needs to read an $\mathbf{R} \times \mathbf{R}$ pixel matrix, with one channel. Conv2D will read the matrix in the $\mathbf{M} \times \mathbf{M}$ and output a new $\mathbf{M} \times \mathbf{M}$ matrix interpretation. MaxPoolig2D merges the interpretations into $\mathbf{N} \times \mathbf{N}$ blocks, reducing the output to $\mathbf{r} \times \mathbf{r}$ bins. The flatten layer will convert an $\mathbf{r} \times \mathbf{r}$ map into a **K** element vector for other layers to process. Secondly, it needs to add a TimeDistributed layer, which can wrap each layer in the CNN model in a TimeDistributed layer when adding it to the main model. Thirdly, the LSTM layer needs to be added, there are three gates, which are the forgotten gate, the input gate, and the output gate, used to control memory in each cell. When the input a_i is input into the LSTM, the LSTM will process the input a_i , which is divided into important information and unimportant information. The unimportant information is forgotten, multiplied by the activation function \boldsymbol{b}_{ω} , and set different hidden layers to the output \boldsymbol{b}_i . This process depends on three factors, including a forgotten gate, input gate and output gate. In a CNN-LSTM network, the parameters may depend on different environments because different parameters such as material, and the environment may change, to some extent, and the results may also be different.

4 Experimental setup

The historical data are collected from different batches of equipment through experimental simulation of different real industrial environments. An accurate prediction model of RUL can offer insights into companies' decisionmaking.

Firstly, it is worthwhile to introduce the background of collected data, which is collected in a sensor base on electrical equipment. The diagram is shown in Fig. 4. The equipment is an AC contactor, which is mainly composed of four parts, including main contact, arc extinguishing device, anti-remanence air gap and auxiliary contact. It is an electrical appliance used to connect and disconnect AC and DC main circuits and high-capacity control circuits from a long distance and frequently. According to the experiment of AC contact, the collected failure modes are mainly fusion welding and explosion, and a few are unable to switch on. Its main control object is the motor, and it can control other power loads, such as electric heating, lighting lamp, electric welding machine, capacitor bank, etc. The equipment continuously records the opening and closing operation and records the data of the three-phase current and three-phase voltage of each switch opening and closing of each batch of different equipment.

Secondly, data pre-processing is the main step. The collected data includes the extraction of data mechanical features, time-domain features, frequency domain features, and wavelet packet features. The data pre-processing is introduced in Sect. 4.2. Finally, the metric used to reveal performance and different results need to be compared, there are two main verification methods in this experiment. The first method is to randomly take one piece of equipment from the same batch of data as the test set, and the rest as the training set. The other way is that the data from the same batch is the training set, and the data from other batches are the test set.

4.1 Data

This data includes the data of three-phase current and voltage collected by the sensor in the experiment, which is opened and closed by simulating the real environment used each time, and the data of current and voltage in the process are recorded. Each opening and closing is an operation, which includes the arc starting time, arc extinguishing time, and closing time of the equipment. There are about 6000-8000 points of continuous waveform data in each operation data set, and the operation times are continuously recorded until the equipment fails. The value of current and voltage is within plus or minus 350. Through

data research, it is concluded that the start of arcing takes a certain time. Therefore, if it is found that t_i consecutive points are less than the threshold before the end of arcing, it is judged as the start time of arcing. When t_i consecutive points are less than the threshold after the end of arcing, it is considered the end time of arcing. The closing judgment principle is the same, and the interval value is judged by setting the threshold value. Because the data is really from the data collected in the industrial environment, some data quality may have some problems and some noise may affect the feature extraction. Therefore, the noise data is replaced before and after according to the machine rational analysis. On the other hand, because the RUL needs to be predicted, the cumulative number of operations per time is increased, which means that when the last failure occurs, the RUL is 1, and when the equipment is operated for the first time, it is the overall service life of the equipment, to establish the output of the RUL of the equipment as training. The input features are shown in Table 1.

4.2 Model setup

In the modelling stage, there are four machine learning algorithms including the LSTM network, DCNN, SVM, and CNN-LSTM network. All the above models will train the RUL of the equipment based on the data extracted from the raw data by using the time domain, frequency domain, and wavelet packet. For the LSTM network, several factors need to be considered, such as the type of layer, the number of layers, and the number of nodes. According to the research, more layers of in-depth learning have a positive impact on the training results, but this may have a

Numeric feature	Note
IA	Phase A current of electrical equipment
IB	Phase B current of electrical equipment
IC	Phase C current of electrical equipment
UAB	The difference between point A and point B
UBC	The difference between point B and point C
UAC	The difference between point A and point C
UA	Phase A voltage of electrical equipment
UB	Phase B voltage of electrical equipment
UC	Phase C voltage of electrical equipment

large burden on the computer, so it is necessary to balance the computational power of the calculation and the accuracy requirements of the model [35]. Therefore, the LSTM is mainly set as three layers: input layer, hidden layer, and output layer. In order to have a certain adjustment space for the model, the fully collected layers are set as the hidden layer, and the Dropout is set to prevent over-fitting. The optimizer selects RMSprop. To make the LSTM positive training data, the learning rate is set as 0.001, and the batch size is set as 100, which means that there are 100 batch data for model training each time. For the CNN network, it needs to reshape the data format, which is $N \times M \times 1$ as the first layer, next, in hidden layers, the Conv2D, the MaxPooling, and activation function needs to be set. the MaxPooling is set 2×1 , which helps to reduce the time of the model and prevent over-fitting, the optimizer also selects RMSprop, the output layer of the dense set as 1 to predict the results. The model call the CNN-LSTM network, which is combined the CNN model and LSTM model, in order to verify that the accuracy of the merged model can be improved compared with that of a single deep learning model, the structure of the CNN and LSTM network is the same as that of the previous structure, but the output layer of CNN is deleted and layer TimeDistributed is added between the structure of CNN and LSTM. It can be wrapping the entire sequence of CNN layers, and as an input layer for the LSTM model, the structure of the CNN-LSTM model is shown in Fig. 5. In this study, three scenarios were introduced. In scenario 1, the collected data are extracted from the arcing interval and closing interval and then modelled based on the popular machine learning and deep learning models. In scenario 2, for the interval of feature extraction, one is to extract the arc interval and closing interval, and the other is to extract all the data sets, but the parameters and algorithm of the model remain unchanged. In scenario 3, select the same model algorithm, but change the parameters of machine learning, the number of layers, and the activation function of deep learning. In scenario 4, select the same percentage as the training set and the rest as the test set, with the percentage of 20%, 40% and 80%.

Table 1 The original feature relevant to RUL





4.3 Performance evaluation

The parameter *m* has a great impact on the experimental results. Based on the RUL prediction, mean absolute error (MAE) and root mean squared error (RMSE) are selected to evaluate the performance of algorithms. The RMSE can be expressed mathematically as:

RMSE
$$(X, h) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (h(x^i) - y^i)^2}.$$
 (2)

The $h(x^i)$ is the observed value and the y^i is the predicted value, the RMSE is 0 if the prediction value equals the actual value, and the MAE can be expressed mathematically as:

MAE
$$(X, h) = \frac{1}{m} \sum_{i=1}^{m} |h(x^i) - y^i|.$$
 (3)

5 Experimental results

5.1 Scenario1: prevailing machine learning algorithms VS deep learning algorithms

In this scenario, the traditional machine learning model will be compared with the deep learning model. The selected models are the data set after feature extraction based on arc burning interval and closing interval. The feature extraction includes time domain, frequency domain, and wavelet packet. An SVM model is mainly optimized by a hyper-parametric algorithm, and the results are obtained. CNN selects regression functions to predict equipment life through the input layer, hidden layer, and output layer. all the tests were conducted on the Intel(R) Core (TM)i5-10210U CPU @1.60 Hz 2.11 GHz. The test set and training set are divided by selecting the equipment data of the same batch of experiments, extracting 30% batch as the test set and the rest as the training set. The results show that the RMSE of the SVM model and CNN model

are 88.7 and 64,1 respectively, and the values of MAE are 65.4 and 56.4 respectively, the values of RMSE and MAE of LSTM are 58.1 and 53.0 respectively. It can be found that LSTM, as a classical time series model, improves the accuracy of RMSE by 33.5% and MAE by 19.0% compared with the traditional machine learning model SVM. Compared with CNN RMSE decreased by 11% and MAE decreased by 6%. However, this experiment introduces a new model CNN-LSTM to explore the promotion effect of the merged model on the deep learning model. The results show that the RMSE of the CNN-LSTM model is 54.7 and the MAE value is 51.8, the RMSE value decreased by 14.7% in the CNN model, MAE decreased by 8.1%. Compared with LSTM, RMSE and MAE decreased by 6.2% and 2.3% respectively. Therefore, the merge neural network CNN-LSTM model performs better in several models. The modelling results are shown in Fig. 6.

5.2 Scenario2: different intervals are used for feature extraction

In this scenario, it compares the prediction results of arc firing interval and closing interval extracted according to the mechanical principle and the results of feature extraction of all data sets. The algorithm model still selects SVM, LSTM, CNN, and CNN-LSTM. The results show that the values of RMSE and MAE of SVM are 107.7 and 81.2 respectively, the values of RMSE and MAE of LSTM are 71.5 and 59.3 respectively, and the error values of CNN are 81.7 and 69.4 respectively, the value of RMSE of CNN-LSTM is 65.5, and the value of MAE is 58.1. From the overall results, the model results of arcing interval and closing interval based on domain knowledge extraction are higher than those of whole interval extraction. The modelling results are shown in Fig. 7.





5.3 Scenario 3: different layers of deep learning models and machine learning parameters

In this scenario. The machine learning model uses grid search to optimize the parameters *c* and **gamma** of SVM and obtains the best parameters *c* and **gamma**. For the deep learning model, increase the number of hidden layers and increase the activation function, keep the optimizer unchanged, and add a full connection layer in the output layer. The results show that SVM has a certain improvement, the RMSE and MAE are 83.2 and 63.1 respectively, but the improvement of the results of the deep learning model is not obvious. The modelling results are shown in Fig. 8.

5.4 Scenario 4: different data sets are selected as the training set

In this scenario. Since the usable life lengths of different equipment experiments are different, the same percentage is selected for the experiment, mainly 20%, 40%, and 80% of the data sets are selected as the training sets for training. Due to the need to study the impact of the data sets on the model accuracy, the feature extraction interval of the data and the model parameters are consistent with scenario 1. The results show that with the increase of the data set, the model accuracy will also increase. On the contrary, when 40% and 20% of the data set are selected as the training set, the model results will decline to vary degrees. The modelling results are shown in Table 2.



 Table 2
 the impact of different training sets on the performance of different models

Different percentages of training sets	Models	RMSE	MAE
20% training set based on feature extraction	SVM	153.8	130.1
	CNN	147.1	127.1
	LSTM	145.4	125.1
	CNN-LSTM	144.3	124.7
	SVM	143.8	127.1
40% training set based on feature extraction	CNN	127.7	121.1
	LSTM	125.4	119.1
	CNN-LSTM	120.4	117.4
	SVM	89.1	67.4
80% training set based on feature extraction	CNN	63.4	57.1
	LSTM	57.4	53.1
	CNN-LSTM	54.1	50.9

6 Discussion

PdM has been widely studied to improve the ability of companies to optimize problems. Residual life prediction is one of the hot research topics. For example, electrical companies predict the residual life of electrical equipment, collect current and voltage data by installing sensors and study the loss of equipment through the use process, by introducing the current novel machine learning model and deep learning to help improve the accuracy of prediction. Therefore, this research mainly proves that the merge deep learning model can further improve the accuracy of the RUL and bring tangible benefits to companies.

6.1 Results discussion

Through the analysis of the three parts, the following two conclusions can be drawn. The accuracy of the feature extraction model through the study of the mechanism features of Feature Engineering, arcing interval, and closing interval, has been improved to varying degrees than that of the feature extraction of the full data set, which shows that the prediction of RUL by professional knowledge can help to improve the accuracy to a certain extent. On the other hand, the accuracy of the model can also be improved by increasing the number of layers and parameters of the model. Machine learning is mainly optimized by increasing the number of layers and changing the parameters and the accuracy of the model is affected by the amount of data. In this experiment, a new method CNN-LSTM network is introduced. The results show that RMSE can improve the accuracy by 5.7% compared with the highest model accuracy, and MAE can improve the accuracy by 2.1%. Therefore, it can bring new insights to the prediction of RUL. Hence, according to these experiment results, there are three key findings in our study. Firstly, as a research hot-spot in the current industrial field, deep learning is more accurate than most machine learning models in terms of model accuracy, but the structural complexity of the deep learning model is higher than machine learning, at the same time, deep learning needs to rely on higher computer configuration, which may have higher investment costs for some projects, By introducing a merged model CNN-LSTM, this experiment can better reflect the model of deep learning, which will be the focus of RUL prediction in the future industry. Secondly, the accuracy of the model obtained by extracting the feature of arcing interval and closing interval is higher than that of all intervals. A professional's knowledge is therefore crucial to developing a model using machine learning and deep learning. Studying the working characteristics of the equipment can help to improve the accuracy of the model. Finally, because the experimental data comes from the data collected by sensors in the real industrial environment, there is a large gap in the total life of equipment in different batches.

6.2 Future works

In this experiment, the accuracy of CNN-LSTM's model after feature extraction is mainly verified. For the previous data set, due to the large difference in the service life of different batches of equipment and the large gap in the usable range of service life, the current sample has certain limitations, and it may be necessary to increase the sample size to continue the research. For the accuracy of the model, because the model is completed under certain configuration restrictions, The computer configuration may have a certain impact on the structure of the model. If the number of layers of the model is increased or the training time is increased again, it may be improved to some extent. Therefore, in the later stage, it may be necessary to further collect the data set and improve the structure of the deep learning model if the configuration allows, to improve the accuracy of the model.

7 Conclusions

PdM is essential to various industries such as aircraft, automobiles, and railways. RUL prediction can provide tangible benefits to the industry in terms of maintenance planning. In this paper, the main focus is on the modelling and prediction of electrical equipment's RUL. This study covers the statistical techniques (time domain, frequency domain, wavelet packet) in PdM, the prevailing machine learning application in PdM, and the deep learning application in PdM. The main contribution of this study is to propose an electrical equipment remaining useful life prediction approach based on the CNN-LSTM model. In the data pre-processing stage, noise elimination, outlier removal, and average value replacement of the original data are implemented. Then the key features are extracted, before it is smoothed by a sliding window to increase the amount of information in the data. Subsequently, the benchmarking machine learning models are selected to establish the model by using different parameters, and the evaluation metrics of RMSE and MAE are selected. According to the model results, it can be seen that the industrial environment has a great impact on the life prediction of products, resulting in some negative performance results when the model predicts different batches. Therefore, the dimension of monitoring quantity plays a vital role in equipment life prediction based on data. In light of the different material characteristics and levels of equipment use, there are large differences in equipment prediction, which will be investigated further.

Funding

Not applicable.

Availability of data and materials Not applicable.

Code availability Not applicable.

Declarations

Competing interests

The authors declare no competing interests.

Author contribution

HF: Conceptualization, Methodology, Writing– original draft and Validation. YL: Conceptualization, Methodology, Review, Resources & Editing. All authors read and approved the final manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 15 May 2022 Revised: 14 July 2022 Accepted: 17 July 2022 Published online: 27 July 2022

References

- R. Ahmad, S. Kamaruddin, An overview of time-based and condition-based maintenance in industrial application. Comput. Ind. Eng. 63(1), 135–149 (2012)
- P. Andersson, L.-G. Mattsson, Service innovations enabled by the "internet of things". IMP J. (2015)
- Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Machinery health prognostics: a systematic review from data acquisition to RUL prediction. Mech. Syst. Signal Process. **104**, 799–834 (2018)
- 4. M. Haarman, M. Mulders, C. Vassiliadis, Predictive maintenance 4.0: predict the unpredictable. PwC and Mainnovation 4 (2017)
- C. Chen, Y. Liu, X. Sun, S. Wang, C.D. Cairano-Gilfedder, S. Titmus, A.A. Syntetos, Reliability analysis using deep learning, in ASME IDETC-CIE (ASME, Quebec, Canada)
- T. Huuhtanen, A. Jung, Predictive maintenance of photovoltaic panels via deep learning, in *IEEE Data Science Workshop (DSW)* (IEEE, 2018)
- T.A. Shifat, H. Jang-Wook, Remaining useful life estimation of BLDC motor considering voltage degradation and attention-based neural network. IEEE Access 8, 168414–168428 (2020)
- Y. Cao, M. Jia, P. Ding, Y. Ding, Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network. Measurement 178, 109287 (2021)
- Y. Liu, C. Chen, T. Wang, L. Cheng, An attention enhanced dilated CNN approach for cross-axis industrial robotics fault diagnosis. Intell. Auton. Syst. 2(1), 1–11 (2022)
- A.-A. Tulbure, A.-A. Tulbure, E.-H. Dulf, A review on modern defect detection models using DCNNs-deep convolutional neural networks. J. Adv. Res. 35, 33–48 (2022)
- C. Cheng, G. Ma, Y. Zhang, M. Sun, F. Teng, H. Ding, Y. Yuan, A deep learning-based remaining useful life prediction approach for bearings. IEEE/ASME Trans. Mechatron. 25(3), 1243–1254 (2020)
- M.Z. Alom, T.M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M.S. Nasrin, M. Hasan, B.C. Van Essen, A.A. Awwal, V.K. Asari, A state-of-the-art survey on deep learning theory and architectures. Electronics 8(3), 292 (2019)
- X. Xu, Q. Hua, Industrial big data analysis in smart factory: current status and research strategies. IEEE Access 5, 17543–17551 (2017)
- 14. L. Chun-Lin, A tutorial of the wavelet transform. NTUEE, Taiwan 21, 22 (2010)
- 15. S. Selcuk, Predictive maintenance, its implementation and latest trends. Proc. Inst. Mech. Eng., B J. Eng. Manuf. 231(9), 1670–1679 (2017)
- L. Ren, L. Zhao, S. Hong, S. Zhao, H. Wang, L. Zhang, Remaining useful life prediction for lithium-ion battery: a deep learning approach. IEEE Access 6, 50587–50598 (2018)
- C.-G. Huang, X. Yin, H.-Z. Huang, Y.-F. Li, An enhanced deep learning-based fusion prognostic method for RUL prediction. IEEE Trans. Reliab. 69(3), 1097–1109 (2019)
- N. Kolokas, T. Vafeiadis, D. Ioannidis, D. Tzovaras, Forecasting faults of industrial equipment using machine learning classifiers, in *Innovations in Intelligent Systems and Applications (INISTA)*, (IEEE, 2018)
- I. Alsyouf, M. Shamsuzzaman, G. Abdelrahman, M. Al-Taha, Improving reliability of repairable systems using preventive maintenance and time-between-failures monitoring. Eur. J. Ind. Eng. 10(5), 596–617 (2016)
- A. Garg, S. Deshmukh, Maintenance management: literature review and directions. J. Qual. Maint. Eng. 12(3), 205–238 (2006)

- A. Malhi, R. Yan, R.X. Gao, Prognosis of defect propagation based on recurrent neural networks. IEEE Trans. Instrum. Meas. 60(3), 703–711 (2011)
- 22. D. Wang, K. Liu, X. Zhang, A generic indirect deep learning approach for multisensor degradation modeling, in *IEEE Transactions on Automation Science and Engineering* (2021)
- C. Liu, D. Tang, H. Zhu, Q. Nie, A novel predictive maintenance method based on deep adversarial learning in the intelligent manufacturing system. IEEE Access 9, 49557–49575 (2021)
- L. Zhang, J. Lin, B. Liu, Z. Zhang, X. Yan, M. Wei, A review on deep learning applications in prognostics and health management. IEEE Access 7, 162415–162438 (2019)
- J. Liu, C. Pan, F. Lei, D. Hu, H. Zuo, Fault prediction of bearings based on LSTM and statistical process analysis. Reliab. Eng. Syst. Saf. 214, 107646 (2021)
- C. Chen, Y. Liu, X. Sun, C. Di Cairano-Gilfedder, S. Titmus, An integrated deep learning-based approach for automobile maintenance prediction with GIS data. Reliab. Eng. Syst. Saf. 216, 107919 (2021)
- 27. H.M. Hashemian, State-of-the-art predictive maintenance techniques. IEEE Trans. Instrum. Meas. **60**(1), 226–236 (2010)
- B. Zhao, X. Zhang, Z. Zhan, Q. Wu, A robust construction of normalized CNN for online intelligent condition monitoring of rolling bearings considering variable working conditions and sources. Measurement 174, 108973 (2021)
- X. Li, W. Zhang, Q. Ding, Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. Reliab. Eng. Syst. Saf. 182, 208–218 (2019)
- B. Wang, Y. Lei, N. Li, T. Yan, Deep separable convolutional network for remaining useful life prediction of machinery. Mech. Syst. Signal Process. 134, 106330 (2019)
- L. Cummins, B. Killen, K. Thomas, P. Barrett, S. Rahimi, M. Seale, Deep learning approaches to remaining useful life prediction: a survey, in *IEEE* Symposium Series on Computational Intelligence (SSCI), (IEEE, 2021)
- Y. Zhou, M. Huang, Y. Chen, Y. Tao, A novel health indicator for on-line lithium-ion batteries remaining useful life prediction. J. Power Sources 321, 1–10 (2016)
- Y. Zhang, R. Xiong, H. He, M.G. Pecht, Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. IEEE Trans. Veh. Technol. 67(7), 5695–5705 (2018)
- H. Yan, J. Wan, C. Zhang, S. Tang, Q. Hua, Z. Wang, Industrial big data analytics for prediction of remaining useful life based on deep learning. IEEE Access 6, 17190–17197 (2018)
- S. Wang, J. Wan, D. Zhang, D. Li, C. Zhang, Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. Comput. Netw. 101, 158–168 (2016)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- ► Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com