

Received 27 September 2024, accepted 6 December 2024, date of publication 16 December 2024,
date of current version 26 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3517705

RESEARCH ARTICLE

Novel Prognostic Methods for System Degradation Using LSTM

WALEED BIN YOUSUF¹, SYED MUHAMMAD UMAR TALHA², ABDUL AHAD ABRO¹,
SADIQUE AHMAD³, (Member, IEEE), SYED MUHAMMAD DANIYAL¹, NAVEED AHMAD⁴,
AND ABDELHAMIED ASHRAF ATEYA^{3,5}, (Senior Member, IEEE)

¹Department of Computer Science, Faculty of Engineering Science and Technology, Iqra University, Karachi 75500, Pakistan

²Department of Electrical Engineering, National University of Sciences and Technology, Karachi 75500, Pakistan

³ELIAS Data Science Laboratory, College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia

⁴College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia

⁵Department of Electronics and Communications Engineering, Zagazig University, Zagazig 44519, Egypt

Corresponding author: Abdul Ahad Abro (abdul.ahad@iqra.edu.pk)

The authors would like to thank Prince Sultan University and ELIAS Data Science Lab for paying the APC of this publication.

ABSTRACT Accurate prognosis of degradation trend is considered very important for the maintenance of critical industrial equipment and assets. This leads to an increase in service availability and life expectancy. Accurate and timely degradation prognosis enables maintenance managers to efficiently plan maintenance regimes and reduce failure occurrences. Recently, Artificial Intelligence (AI) based prognosis and prediction techniques have been in the limelight and attracted the interest of the research community. One such popular AI-based technique is the Long Short-Term Memory (LSTM), which is very efficient in making predictions using time series and sequential data. This paper proposes a novel prognostic technique based on LSTM to predict the degradation trend. The proposed LSTM technique uses a dynamic training window approach with a fixed look-back window for forecasting future steps. The size of the training window is iteratively increased for each prediction as more data is available. This enables the model to utilize the complete sequence trends while making future degradation state predictions. To mitigate over-fitting during model training, the dropout technique and L2 regularization are also incorporated into the proposed generic LSTM model. The performance of the proposed LSTM-based technique is evaluated using experimental results on a real-world application and data. As a case study, the degradation trend of Aerial Bundled Cables (ABCs) using actual thermal degradation data acquired from in-service cables (ABC) is predicted. Moreover, the proposed LSTM-based technique is further compared with a particle filter-based statistical prognosis technique. Promising results validate the efficacy of the proposed LSTM-based approach for degradation prognosis.

INDEX TERMS Long short-term memory (LSTM), increasing sliding window, prognostic technique, degradation trend estimation, aerial bundled cables (ABCs).

I. INTRODUCTION

Condition monitoring plays a vital role in the predictive and preventive maintenance of critical industrial machinery and equipment. Condition monitoring measures specific parameter(s) and identifies any irregularities and variations, which may indicate a possible failure instance. The monitored data or measurements contain the key details pertaining

to the current and historical health state of the equipment. Sophisticated prognostic techniques can then be used on the historical data to predict the future degradation trend [1]. Accurate degradation trend prediction allows efficient planning of maintenance activities & regimes, and estimation of the Remaining Useful Life (RUL) [2]. Effective degradation prognosis ensures upkeep, service availability, and longevity of the asset. It also reduces the probability of failure and system downtime over the lifespan of the asset [3], [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Chuan Li.

Various statistical and artificial intelligence (AI) based prognostic methods have been reported in the contemporary literature to predict the RUL of a system or critical components [4]. Prognosis techniques can be broadly categorized into model-based, data-driven, and hybrid methods, depending on the available data, algorithms employed, and knowledge of the system [4], [5]. Model-based approaches predict the future state using an available physical model that outlines the physics of the failure mechanism. The acquired training data when combined with the physical model predicts the future trend. Although model-based techniques are generally accurate and effective in making predictions, these high-fidelity models are costly, time-consuming, and resource-intensive. In addition, as these models depend on the physics of the system, they have limited reusability [4]. The data-driven approaches make use of the current and previous states, without considering physical parameters. A data-driven approach can be further divided into artificial intelligence (AI) and statistical approaches [5]. As the name suggests, AI approaches make use of AI-based techniques, such as neural networks and fuzzy logic, whereas statistical methods use probabilistic models including gamma process, Gaussian process (GP) regression, Support Vector Machine (SVM), and Bayesian filters, such as Kalman and Particle filters, to make future state predictions [3], [4], [5]. In [4], data-driven methods are more user-friendly and faster to implement and deploy as compared to model-based methods. A practical illustration of one of the statistical data-driven methods, the Kalman Filter, is provided in [6] to predict the remaining useful life of the asset. In [7], a neural network and fuzzy logic-based approach to predict accelerated thermal aging of electric power cables. In another study [8], a hybrid approach is used for predicting the remaining service life of rotating machines. They combine the convolution operation of a neural network (NN) with the strength of a recurrent neural network (RNN), specifically the Long Short-Term Memory (LSTM), for efficient RUL prediction.

II. LITERATURE REVIEW

Recently, AI-based techniques have attracted considerable interest from the research community, in particular, Deep Learning (DL) based methods, such as LSTM, have been in the limelight for prognostics study of various systems [4]. LSTM has been reported to model damage propagation of aircraft gas turbine engines and lithium-ion batteries [9], [10]. In [11] a deep LSTM model for the dual task of assessing the degradation and then predicting the RUL of aero engines. RUL of a turbofan engine is predicted by Paulo da Costa et al. using the time series sensor data [12]. They combine global attention mechanisms with LSTM architecture. This enables it to learn RUL degradation and demonstrate effectiveness over classical algorithms [12]. Another study predicted the RUL of rotatory machines with the aid of LSTM and attention mechanisms. They use a one-dimensional Convolutional Neural Network (CNN) to extract local features from the

signal sequence and then feed them to a multi-layer LSTM architecture with an attention mechanism [13].

LSTM and other AI-based approaches generally encounter overfitting problems [14]. L2 regularization and dropout are frequently used with LSTM to avoid overfitting and improve the results of neural networks [15]. Various parameters, such as network complexity, number of hidden neurons, etc., control the performance of these techniques [15]. Dropout turns out to be more robust and improves performance in larger networks as compared to L2. Whereas L2 tends to yield higher predictive accuracy in small networks [15]. Dropout tries to combine the predictions of many “thinned” networks to boost performance [16]. The dropout technique drops some neurons during the training and allows the network to use all the neurons during testing to avoid overfitting and improve performance. L2 has become a norm to be applied with dropout or separately in LSTM neural networks for different applications such as speech recognition [17] and RUL prediction [18]. In this reported work, Long Short-Term Memory (LSTM) and its regularized versions are optimized for degradation prognosis in a real-world application using actual data.

In this research work, a novel prognostic technique based on LSTM deep learning algorithm is proposed. The novel technique modifies the classical LSTM model by incorporating an increasing training window approach in it. The degradation prognosis of Aerial Bundled Cables (ABCs) is selected as the case study to undertake the research and predict the degradation trend. Three different LSTM variants, namely generic LSTM (with increasing window), L2 regularized LSTM, and LSTM with dropout incorporation, are used for degradation trend estimation. A comparative analysis of the three LSTM variants is also reported in this study. Further comparison with particle filter-based statistical prognostic technique is also reported.

The next section describes the research methodology and details of the proposed novel LSTM approach. The following section presents the case study description, including details of data acquisition, model and data settings, data visualization, preprocessing, and implementation of the proposed LSTM-based scheme on the case study. This is followed by results and discussion; and finally, the conclusion and future work section is presented.

III. METHODOLOGY

Recently, Artificial Neural Network (ANN) models have been widely used for AI and machine learning applications as compared to conventional regression and statistical models [19]. Recurrent Neural Network (RNN) is a class of artificial neural networks with deep architecture that contains feedback connections to the preceding layers, thus it can process variable length data [20]. RNN is a common network used to model sequential data, such as the time-series data used in this study. However, RNN has long-term dependency limitations and vanishing gradient problems, and it often

fails to remember distant information. The Long Short-Term Memory (LSTM) networks, a variation of RNN, are however capable of capturing long-term dependencies by incorporating a memory cell that preserves the state over time [4]. Consequently, LSTMs have gained a lot of attention in a variety of time-series problems such as weather forecasting and speech recognition [21], [22]. Researchers in different studies [9], [10], [11] have applied LSTM as an efficient model for predicting the remaining useful life (RUL) of critical system components as well.

LSTM, like typical RNN architecture, can be considered as a chain of repeated blocks of neural networks. Each module/block of LSTM has four gates/layers interacting in a very special way. Figure. 1(a) shows the chain structure with three LSTM modules/blocks. The middle block shows an internal architecture for the four gates/layers. These layers help in understanding the internal mechanism of the LSTM network and demonstrate its superiority over traditional RNNs.

Figure. 1. Pictorial representation of repeating LSTM modules, each with four layers/gates. (a) A set of three LSTM blocks, (b) Forget gate representation, (c) Input gate and candidate cell update state representation, (d) Current cell state estimation using forget gate, input gate, and candidate update value, (e) Representation of the current state output.

The main theme of the LSTM architecture is the flow of information through the two horizontal lines connecting any two LSTM blocks, as shown in Figure. 1(a) [9]. Each line carries the output of the previous module to the input of the next module. The circle under and outside the module represents the input of the current state, while the circle above and outside the block represents the output of the block. The circles inside the modules show point-wise operations such as the addition of vectors and rectangular boxes representing the gates/layers. Lines that are merging show a concatenation operation, whereas a line branching out simply contains a copy of that signal. The bottom horizontal line on the left of each block represents a combination/ concatenation of the input of the current state (X_t) and the output from the previous state output (h_{t-1}). This is used to calculate the control signals for the three gates (namely forget, input, and output), and for the candidate update value of cell state, as shown in Figure. 1(b)-(e). The forget gate is represented in Figure. 1(b), while the input gate and its relation to the candidate cell update are highlighted in Figure. 1(c). The top horizontal line carries a copy of the previous state output (h_{t-1}). This is transferred to the current state output while being controlled through the forget gate, input gate, and the cell update, as shown in Figure. 1(d). The LSTM block output (h_t) is calculated by a nonlinear transformation of the current state (c_t) and output gate (o_t), as shown in Figure. 1(e). In this way, the LSTM can update its cell state by controlling what information to add and what not to add, by using the forget, input, and output gates. The stepwise details of LSTM working and signal propagation along with mathematical representation are given below.

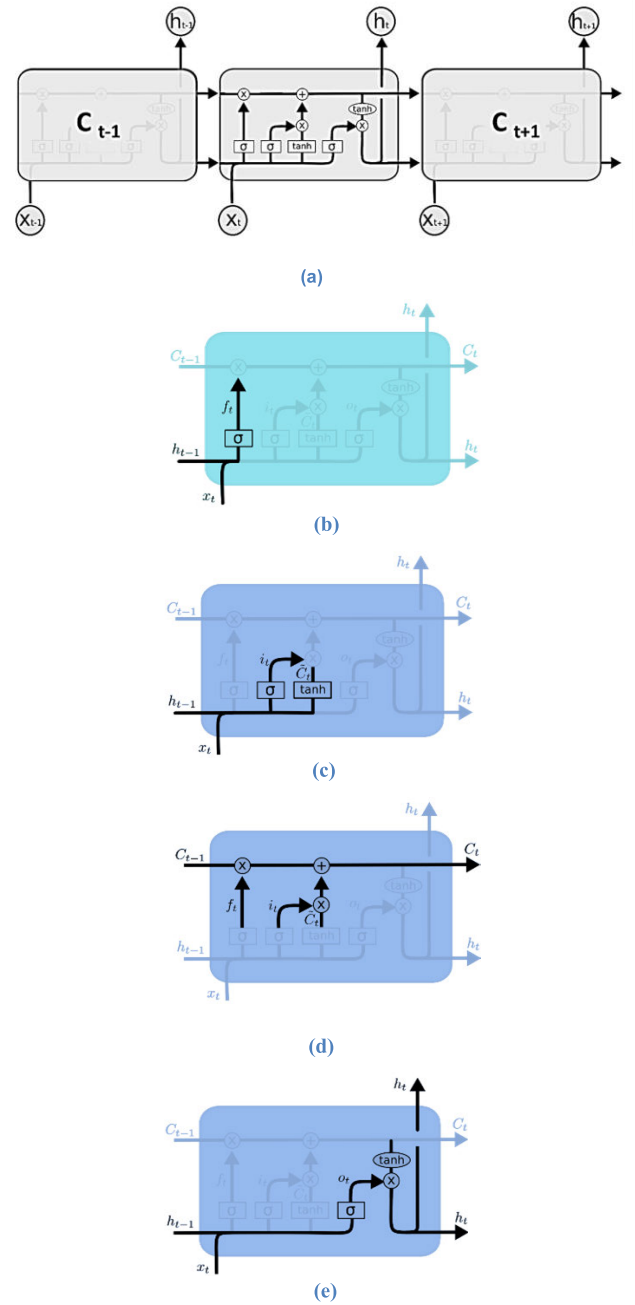


FIGURE 1. (a). Structure with three LSTM modules. (b). Forget gate representation. (c). Input gate and candidate cell update state representation. (d). Current cell state estimation using forget gate. (e). Representation of the current state output.

In the first step, the LSTM decides what information to discard from the cell state. This is decided by the forget gate (f_t), which is typically a sigmoid function used on the weighted previous state output (h_{t-1}) and the current input (X_t). The forget gate provides a number between 0 and 1, with 1 representing the information is kept, while 0 means the information is dropped. The forget gate is illustrated in Figure. 1(b) and is mathematically expressed as:

$$\bar{f}_t = W_f \cdot X_t + R_f \cdot h_{t-1} + b_f \quad (1)$$

$$f_t = \sigma(\bar{f}_t) \quad (2)$$

Equations (1) and (2) show how the forget gate output is generated where, $\bar{f}_t$ is the forget gate input at time t , while f_t is the corresponding output of the gate. W_f represents the forget gate weight for input signal X_t , R_f is the recurrent weight for h_{t-1} and b_f represents the bias. σ is the nonlinear sigmoid function, and “.” denotes matrix multiplication.

The next step decide what new information is to be updated in the LSTM cell state. This is undertaken in two steps as shown in (3)(4)(5), and (6). Firstly, the input gate (i_t) decides which values should be updated. Next, a nonlinear layer φ (tanh layer) creates a vector of new candidate values, that could be added to the state. This is further illustrated in Figure. 1(c) and mathematically expressed as:

$$\bar{i}_t = W_i \cdot X_t + R_i \cdot h_{t-1} + b_i \quad (3)$$

$$i_t = \sigma(\bar{i}_t) \quad (4)$$

$$\bar{C}_t = W_C \cdot X_t + R_C \cdot h_{t-1} + b_C \quad (5)$$

$$\tilde{C}_t = \varphi(\bar{C}_t) \quad (6)$$

where W_i , R_i , b_i and W_C , R_C , b_C are input weights, recurrent weights, and bias of input gate and cell state layer, respectively. $\bar{i}_t$ is the input of the input gate, while i_t is the gate output. Similarly, $\bar{C}_t$ is the input and $\tilde{C}_t$ is the corresponding output for the candidate update values.

The two outputs i_t and $\tilde{C}_t$ from the previous step are then combined to update the old cell state with the new cell state, as shown in Figure. 1(d) and mathematically expressed in (7):

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (7)$$

Finally, the output gate (o_t) controls the amount of information on the cell states being kept in LSTM output. The output gate's input is generated as shown in (8). The output gate typically uses a sigmoid function as shown in (9). Then, tanh is applied to the cell state to push the values between -1 and 1 . Lastly, the two outputs from the tanh layer and the output gate are combined to extract the LSTM unit output (h_t) at time instance t as illustrated in (10). This is shown in Figure. 1(e) and mathematically represented as:

$$\bar{o}_t = W_o \cdot X_t + R_o \cdot h_{t-1} + b_o \quad (8)$$

$$o_t = \sigma(\bar{o}_t) \quad (9)$$

$$h_t = o_t * \varnothing(C_t) \quad (10)$$

where, W_o , b_o represents the weights and bias of the output gate, $\varnothing$ is the activation function (tanh), the $*$ represents scalar multiplication, and (o_t , $\bar{o}_t$) have a similar meaning to (f_t , $\bar{f}_t$).

The weights involved in the above LSTM model are optimized during the training process. The classical LSTM framework is trained up to a certain point in the dataset, commonly referred to as the train-test split. Typically, the train-test split reserves the larger share for training purposes, such as keeping 70-90% for training and 30-10% for testing [8], [9], [10]. However, a training split covering

the majority of the dataset, say 90%, may result in overfitting [14], [15]. Whereas, keeping a relatively lower training proportion, say 60-70%, may reduce the prediction accuracy as the model has to predict a considerable portion in time. In the classical LSTM model, prediction of the future state is made based on the training and the look-back window. A “lookback period” determines the number of previous instances/ states used to predict the next timestep(s). For instance, a lookback period of **3** means the previous states at time $t-2$, $t-1$, and t are used to predict the next state at time $t+1$. Typically, the size of the lookback period/ window is also optimized/ determined in the training process. The subsequent future time step(s) prediction is made based on trained weights up till the train-test split and the look-back window input sequence. Therefore, the classical LSTM model heavily depends upon determining the lookback period and the train-test split. The lookback period combined with a number of future step predictions can enhance the prediction efficiency of the LSTM model. In this reported work, the proposed LSTM scheme modifies the classical train-test split framework and prediction mechanism.

In this study, an increasing sliding window-based training approach is used to predict the future trend. Unlike the classical LSTM model with a fixed train-test split, the proposed model uses a dynamic approach to the size of the training window. For each prediction, the size of the training window is iteratively increased, while keeping a fixed look-back window size. The number of future steps predicted, ahead of time, is also kept constant for each iteration. This will restrict encountering a large portion of the test set, even when the training set is smaller in size. The size of the training set increases as more data is made available. Moreover, this approach allows predictions to be made using all of the available data trends. For a better visualization and understanding of the dynamic training window approach, Figure. 2 presents an illustration to elaborate the iterative process. Consider a particular case where the training window size is set to 20 in the first iteration, then becomes 21 in the second iteration, and so on. However, the number of future predictions made stays the same, 10 in the case shown in Figure. 2; this will be referred to as a 10-step prediction.

The proposed LSTM-based prognostic technique is used for degradation trend estimation using real-world data. The next sub-section outlines the case study highlighting aerial bundled cables (ABCs), their composition, and the need for prognosis. In addition, the data acquisition and thermal degradation parameters are also discussed.

IV. CASE STUDY

A. AERIAL BUNDLED CABLES (ABC)

Aerial bundled cables (ABC) are an innovative solution for overhead power distribution cables, used in various parts of the world [23], [24]. Unlike classical uninsulated overhead power distribution cables separated by air gaps, aerial bundled cables are composed of multiple insulated

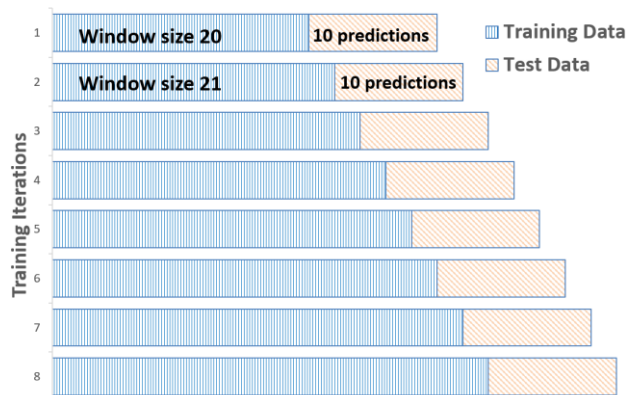


FIGURE 2. Change the proposed model's training and test set over each iteration.

phase conductors. The aluminum phase conductors are closely bundled together with a cross-linked polyethylene (XLPE) insulation layer and assembled with a neutral conductor. The XLPE insulation makes ABCs safer and more reliable as they offer lesser pilferage losses among other benefits [3]. ABC insulation makes it convenient to install in densely populated areas with narrow gaps between buildings, making it a suitable candidate for power distribution in congested urban areas. ABCs typically demonstrate a better ability to withstand heat, cold, and intense sunlight due to their insulation, and thus faults are less likely to occur on ABCs than on classical uninsulated copper cables [25].

Nevertheless, ABCs may experience XLPE insulation degradation due to various reasons including electrical, mechanical, and thermal loadings exceeding the listed ratings, and harsh environmental conditions, such as high temperatures & humidity [3], [26]. ABCs have been reported to have sudden failures, especially in marine and coastal areas with high moisture content. Such failures are typically due to moisture ingress through punctures/ ruptures in the XPLE insulation. This may lead to dangerous situations including voltage fluctuations and short circuits which leads to damage to precious equipment and even power supply outages. Moreover, the higher temperature of coastal areas causes the wire to overheat and degrade even further. All these factors in turn cause aging and reduce the remaining useful life of the cables.

Due to the insulated nature of the XLPE-based ABCs, physical or manual inspection of the cables is not possible. However, conditional monitoring of the installed cables followed by efficient prognosis is a candidate scheme to mitigate such sudden failures. The prognosis of cables includes the prediction of future health state based on historical monitoring and current cable conditions. It is thus crucial to develop degradation assessment schemes and predictive models that can monitor the changing state of ABCs. Accurate prognosis aids maintenance managers in preventive maintenance planning and in turn guarantees the upkeep and long life of cables [20], [26]. In this reported work, the proposed LSTM-based prognosis approach is applied to the actual

field-acquired ABC data, the details of which are provided in the next subsection.

B. DATA ACQUISITION AND THERMAL DEGRADATION PARAMETER

In this research, coastal areas of Pakistan, in particular Karachi with latitude and longitude coordinates of 24.86° N, and 67.0° E, have been selected for acquiring non-destructive data. Karachi is a metropolitan city located on the coast of the Arabian Sea and has a moisture-rich environment. The ABC cables installed in Karachi are XLPE insulated with aluminum core; a pictorial representation of a sample cable is displayed in Figure. 3.



FIGURE 3. A sample of aerial bundled cables installed at Karachi.

One of the main reasons for the insulation degradation in ABCs is the moisture ingress, which leads to cable failure. ABCs are typically installed between two distant poles using Insulation Piercing Connectors (IPCs). However, these connectors may puncture the insulation to make electrical contact. These punctures and subsequent cracks in cable insulation act as a point of entry (POE) for moisture and humidity. Typically, there is an opening near the connectors at both ends of the cable, the opening and POE are shown in Figure. 4. The penetrated moisture content and the passing current heating effect degrade the ABC insulation over time.

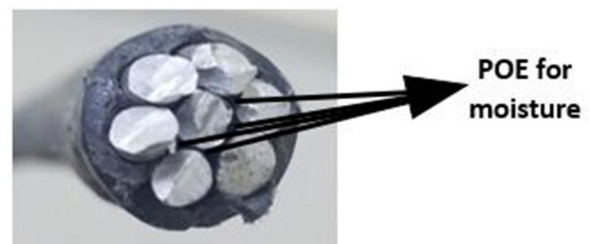


FIGURE 4. Point of entry for moisture in abc.

In this research work, thermal conditional monitoring is reported. A FLIR[®] E40 thermal imaging camera (TIC) has been used for collecting thermography information (temperature-energy distribution) of the in-service ABC. The specifications of this camera are listed in Table 1.

The TIC-based thermographic results are used to monitor the degradation of the installed ABCs. The data is acquired at a constant time of the day to make effective comparisons and minimize environmental variations, such as light

TABLE 1. Specifications of FLIR-E40 thermal imaging camera.

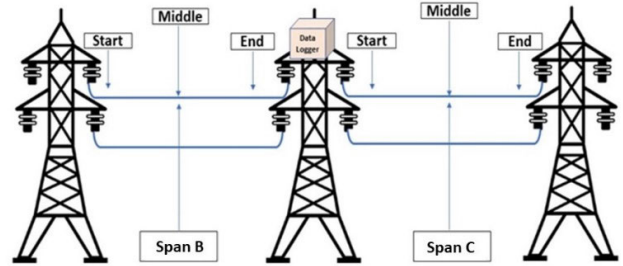
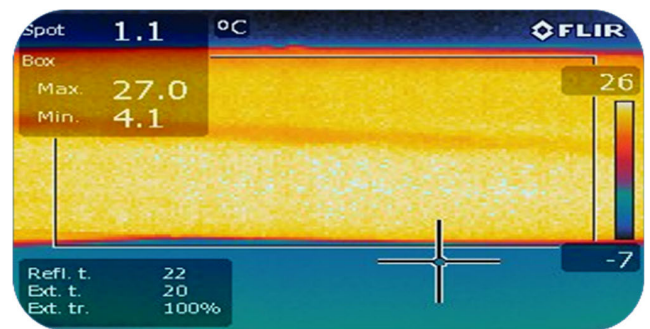
Characteristics	Detail
Temperature range	-4°F to 120°F (- 20°C to 650°C)
Frame rate	60Hz
Field of view/min focus distance	25° x 19° /1.31 ft (0.4m)
Resolution	19,200 pixels (160x120) Infrared resolution
Accuracy	+/- 2% (reliable temperature measurement)
Thermal sensitivity	<0.07°C at 30°C

and temperature intensity. Moreover, to mitigate inaccuracies about atmospheric conditions, the TIC is kept at a constant distance from the ABC; 5-6 inches at a horizontal angle. Actual thermographic data acquisition instances are shown in Figure 5.

**FIGURE 5.** Data acquisition using thermal imaging camera.

Two spans of ABC were used in this reported work for data acquisition, named span B and span C. The span of each cable is further divided into three smaller segments (namely, start, middle, and end segments) due to the long cable length and limited field of view of the TIC. Six thermal images were taken at each test instance at predefined and fixed sample points along the length of ABC as shown in Figure 6. Thermal images are then captured at each of these smaller segments on a monthly basis and a historical health monitoring database of ABC is generated.

Six images were captured at each measuring instant (3 images for each span). An example of the captured thermal image of ABC is shown in Figure 7. The acquired thermal images were then processed and the corresponding absolute temperatures of the cable were computed. The data used in the reported research is acquired at six different locations at each measuring instant.

**FIGURE 6.** Nomenclature used for thermographic data acquisition. Two ABCs are used with each's span further divided into three smaller segments (start, middle, and end).**FIGURE 7.** A sample of the captured thermal image of ABC.

The current(load) in the cable for each data acquisition instance varies based on consumer usage. This loading effect should be considered while generating the historical database. Therefore, a thermal degradation parameter (TDP) is used to incorporate the thermal parameters corresponding to the electrical loading. TDP is equivalent to the temperature of the cable per unit ampere current. A sample TDP data instance used in the study is tabulated in Table 2:

TABLE 2. A sample thermal degradation parameter (TDP) data instance.

Thermal Degradation Parameter (°C/A)			
	Start Segment	Middle Segment	End Segment
Span B	0.5726	0.5733	0.5800
Span C	0.5797	0.5713	0.5688

C. IMPLEMENTATION /DATA SETTING

Applying LSTM to ABC thermal degradation data involves understanding the nature of the data, preprocessing, and visualizing the data, and then final training the neural network. Once a model is learned, it is validated using a cross-validation technique and then finally tested for accuracy in predicting the degradation trend. The following sub-sections outline the implementation details of the model.

D. DATA VISUALIZATION AND PRE-PROCESSING

Thermal degradation parameter (TDP) data collected for aerial bundled cables over 630 days or 90 weeks is used in this research. The recorded sensor readings are then tabulated; a visualization representation of the processed TDP values is shown in Figure 8. Typically, from week 30 to 70 the recorded values showed a quadratic trend, and then after 70 weeks, the values increased exponentially.

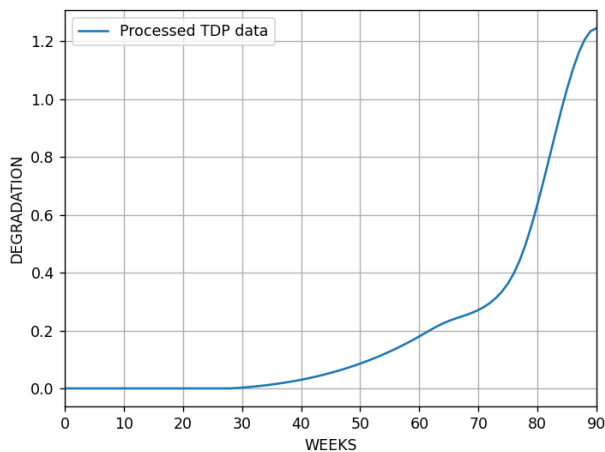


FIGURE 8. Visualization of thermal degradation parameter values.

The rate of change feature is extracted from a data set of 90-week values and used as input to train the model. Thus 90 weeks of data is available for model training. This reported work presents a novel approach of “Increasing Window”; in which the model is trained in each iteration with a prediction step of 4 weeks. Increasing the window approach is useful in processing the more recent data as the window size grows, which is also beneficial for non-stationary and evolving sequences. Data used in the first iteration is 40 weeks then for the next iteration window size will increase by one week. For prediction in each iteration, the last 20 weeks values from the training dataset of the total available data (i.e. 90 weeks) are provided as a test dataset to predict data value for the next 1 week. This predicted week is then added to the test dataset for new predictions. Since the predicted value is the rate of change therefore first prediction of each iteration will be added to the actual value from the dataset. The rest predictions will be added to the predicted ones. Our validation strategy ensures that the model is evaluated on data it has not seen during training, helping to identify any issues with generalization. Early stopping prevents overfitting by halting training when validation performance stops improving. We also ensure that the final model is evaluated on an independent test set in each iteration to gauge its performance in real-world scenarios.

E. NETWORK CONFIGURATION AND TRAINING

In this reported work, three different variants of the proposed LSTM models have been configured followed by a

comparison of these models. The first model is a generic model with an LSTM layer and a fully connected dense layer, the second model uses regularized LSTM with L2 regularization applied to it and the third model incorporates dropout with LSTM.

The first model uses a generic LSTM for predicting the degradation of ABCs. It is a sequential model with an LSTM layer followed by a dense layer. Hyperbolic tangent (tanh) is used as an activation function for the LSTM layers. An Adaptive Moment Estimator Adam optimizer is used in this research to train the LSTM as it is known to provide good results in sparse gradients and noisy data [27]. In addition, Adam’s optimizer is considered more efficient in terms of time and accuracy. In contrast, traditional batch or stochastic gradient descent-based training algorithms for deep neural networks are typically time-consuming in weight optimization [27]. The epochs are set to a maximum of 100 and mean square error is used as an evaluation metric for cross-validation.

The second model (L2 Model) aims to avoid overfitting and thus adds regularization to the generic model. Overfitting is a serious concern while training a model using neural networks, where the model learns the training data perfectly and fails to generalize on test data. To avoid over-fitting, the L2 regularizer is incorporated in Model 2 implementation. This paper uses a ‘kernel regularizer which is applied to the layers of the LSTM model rather than to the bias or the activity of the network. The L2 regularizer has a learning rate set to 1×10^{-10} and is added to the LSTM layer of the network. L2 regularization is also known as ridge regression in literature, it aims to prevent over-fitting by penalizing complex models [28]. Euclidean distance is used as a penalty term and the model learns from features that have small weights rather than ignoring them [28].

The third model incorporates a dropout layer in it, keeping the configurations of the first model. The core idea of dropout is to randomly drop units and relevant connections from the neural networks during training. Figure 9a shows a standard neural network where each neuron in a layer relates to all the other neurons of the next layer. However, in the case of dropout, some neurons are dropped as can be seen in Figure 9b, neurons with a cross (X) in them have no connections in the network. This prevents units from co-adapting too much and consequently prevents the model from overfitting [28]. In the reported model, dropout with a probability of 0.5 is used after the LSTM layer during the training phase.

F. MODEL PERFORMANCE EVALUATION METRIC

The performance of the three models is evaluated using the root mean square error (RMSE). It is a good estimator for calculating the standard deviation of the distribution of errors. It helps in understanding the accuracy of the model while predicting degradation. RMSE is mathematically

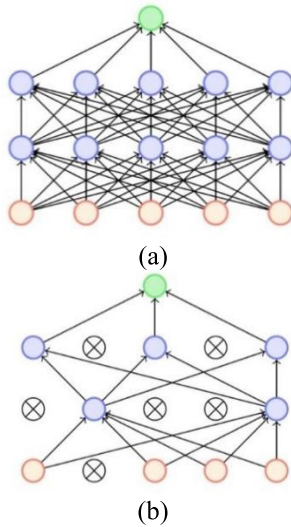


FIGURE 9. Neural networks and dropout (a) Standard neural network without dropout (b) Neural network with dropout incorporated in it.

represented as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{n}} \quad (11)$$

Equation (11) shows how to calculate RMSE where $\hat{Y}$ denotes the predicted TDP value and Y_i is the actual or true value from the dataset; while n is the size of the test set or a number of predicted values.

V. RESULTS AND DISCUSSION

In this paper, the degradation of two spans of aerial bundled cable; span B and C as explained above in the case-study section, which are each further divided into three segments; start, middle, and end is predicted. In the first experiment, the LSTM is applied using various look-back sizes, and the performance is compared based on RMSEs. The experiment uses 5,8,10 and 20-step look-back values, while it predicts the next 10 weeks prediction of which only the first 4 weeks' values are reported in this study. For this experiment, TDP data of the Span B start segment is used. The graphical representation of the experiment is shown in Figure. 10; which shows that the size of 20 steps has the lowest RMSEs through all weeks of prediction. Therefore, a 20-step look-back will be used in further experiments.

Next, the classical LSTM approach (no fixed lookback, iterations, and increasing window) is compared with the generic/proposed LSTM approach with an increasing sliding window. Both the classical and proposed generic LSTM models predict the future 4 weeks of degradation from the input data fed to them. The classical model takes as input the first 86 weeks of data to train and predicts the last 4 weeks. On the other hand, the generic model is provided with 40 weeks of data for training, and it predicts the next 4 weeks using 20 look-back. Figure. 11 shows the prediction

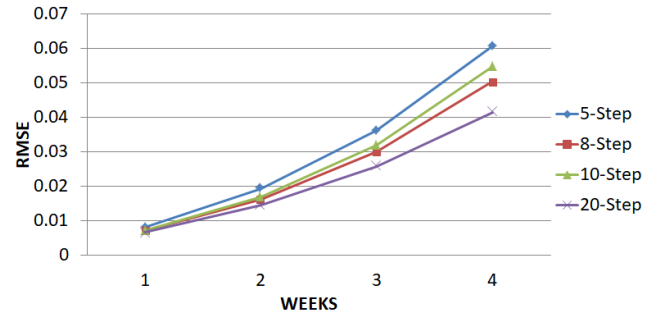


FIGURE 10. Comparison between different look-back sizes on four-step prediction.

curves for both classical and proposed generic LSTM compared with the original TDP trend. The proposed approach outperforms the classical method as its prediction curve closely matches the original TDP curve. Similar outcomes are verified from the RMSE that was recorded for both models. The proposed approach presented an RMSE of 0.0067, whereas, classical presented 0.1734 on TDP data.

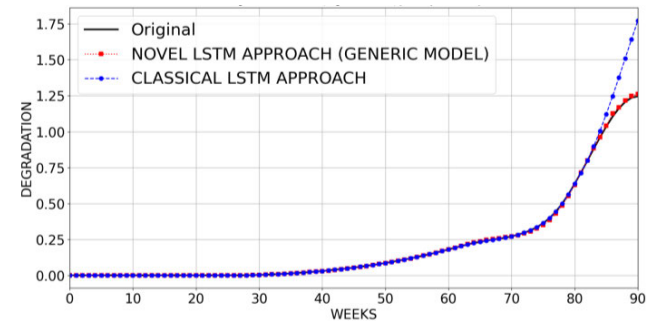


FIGURE 11. Comparison of classical and generic LSTM model's prediction.

Figure 10 demonstrates the superiority of the proposed generic LSTM approach. The proposed technique and its variants, Generic, Dropout, and L2, were further tested with the acquired ABC TDP data. The proposed LSTM model then predicts a 4-step future degradation trend, i.e. four weeks ahead in time. 1st Step predictions of the three LSTM approaches are shown in Figure. 11, where 1st step means predicting one step/week ahead. TDP data pertaining to Span B, the start segment is used in this experiment. As a reference, the original TDP values are also shown in Figure. 12. The three LSTM approaches capture the degradation trend very well, as shown in Figure 12.

Similarly, 2nd Step, 3rd Step, and 4th Step predictions, corresponding to 2, 3, and 4 weeks ahead, are shown in Figure. 13-15, respectively. All these experiments use span B, start segment data. The three LSTM approaches are compared with the original TDP values in each of the figures, Figure. 13-15.

As evident from Figure. 13-15, all three LSTM-based approach predictions are very close to the original value, and it is hard to conclude which scheme performs the best.

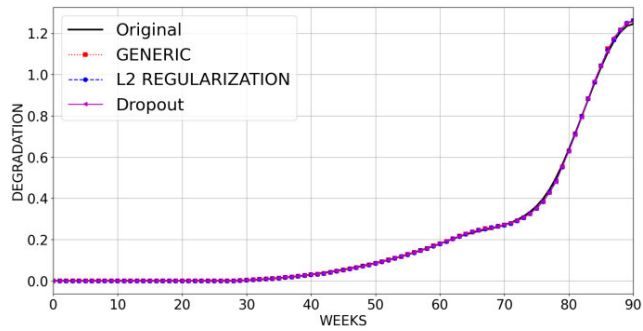


FIGURE 12. Degradation of span B starts segment-1st step prediction.

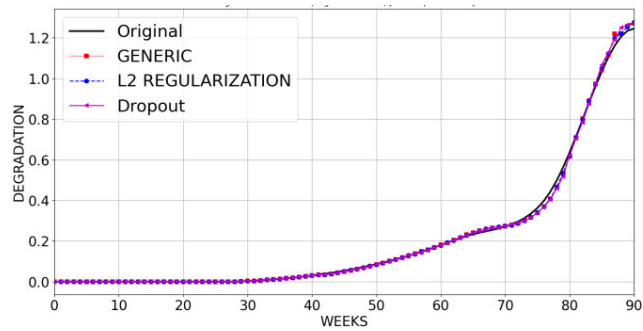


FIGURE 13. Degradation of span B starts segment-2nd step prediction.

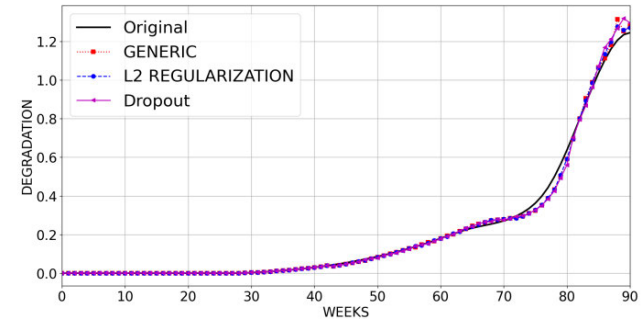


FIGURE 14. Degradation of span B starts segment-3rd step prediction.

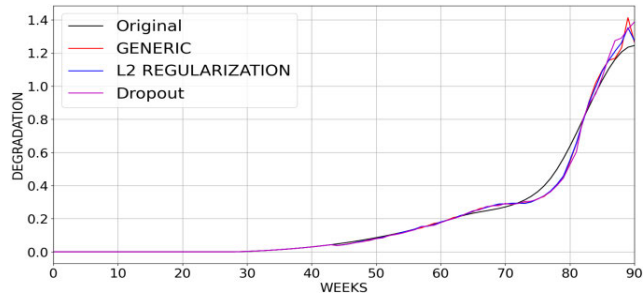


FIGURE 15. Degradation of span B starts segment-4th step prediction.

Therefore, for comparison between the LSTM models, RMSE values for the 1st – 4th step of prediction for each segment (start, middle, and end segments) of both spans, B and C, are calculated. The 1-4 step error-based comparison

for Generic LSTM for each span of the cable’s segments is tabulated in Table 3. The mean value, across all span segments, for each of the 1st, 2nd, 3rd, and 4th step errors are also tabulated in Table 3.

TABLE 3. RMSE-based comparison for generic LSTM model.

SPAN B OF CABLE	1 st Step Error	2 nd Step Error	3 rd Step Error	4 th Step Error
SEG: START	0.006746	0.015375	0.027560	0.044581
SEG: MID	0.014992	0.036163	0.063308	0.095608
SEG: END	0.015309	0.036952	0.062892	0.092925
SPAN C OF CABLE				
SEG: START	0.014368	0.034568	0.063088	0.102634
SEG: MID	0.012532	0.030344	0.053660	0.082300
SEG: END	0.010505	0.026050	0.053012	0.096825
MEAN	0.012409	0.029909	0.053920	0.085812

Similarly, the RMSE-based comparison for L2 Regularization and dropout-based LSTM models are tabulated in Tables 4 and 5, respectively. Each of these tables lists all span segments and presents 1-4 step errors; along with mean values.

TABLE 4. RMSE of dropout model.

SPAN B OF CABLE	1 st Step Error	2 nd Step Error	3 rd Step Error	4 th Step Error
SEG: START	0.007563	0.018092	0.033951	0.056081
SEG: MID	0.015181	0.036072	0.062274	0.093081
SEG: END	0.014381	0.033866	0.057915	0.085318
SPAN C OF CABLE				
SEG: START	0.015288	0.037351	0.069136	0.112515
SEG: MID	0.013129	0.030953	0.053104	0.078575
SEG: END	0.012042	0.032237	0.065743	0.114366
MEAN	0.012931	0.031428	0.057020	0.089989

TABLE 5. RMSE Of L2 regularization model.

SPAN B of CABLE	1 st Step Error	2 nd Step Error	3 rd Step Error	4 th Step Error
SEG: START	0.006482	0.014305	0.024698	0.039785
SEG: MID	0.014546	0.035043	0.061392	0.092346
SEG:END	0.015058	0.035568	0.060476	0.089145
SPAN C of CABLE				
SEG: START	0.014374	0.034939	0.064972	0.107266
SEG: MID	0.012641	0.030512	0.053848	0.081855
SEG:END	0.010343	0.026250	0.054280	0.099472
MEAN	0.012241	0.029436	0.053278	0.084978

The mean values, as listed in Table 3-V, enable a comparison of the three LSTM models. For better visualization of this comparison, these mean values are also plotted and shown in Figure. 16. As inferred from Figure. 16,

the L2-based LSTM model outperforms the other reported models and has the lowest mean values of errors for all predictions.

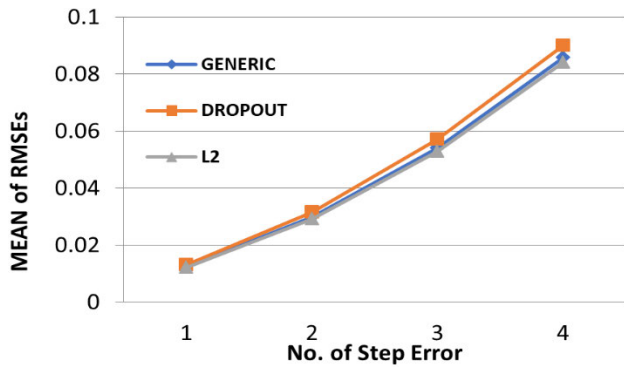


FIGURE 16. Comparison between Generic, Dropout, and L2 models.

To further demonstrate the efficacy of the proposed approach, the predicted thermal degradation of aerial bundled cables is further compared to a statistical method using the same approach. A particle filter (PF) statistical method [26] to predict the insulation degradation of ABCs using the same dataset. PF is used to estimate the future health of cable and the RMSE for each cable is recorded after prediction. 1st step prediction for both LSTM (L2 Regularization model) and Particle filter is shown in Figure. 17. Span B, start segment TDP data is used for this experiment.

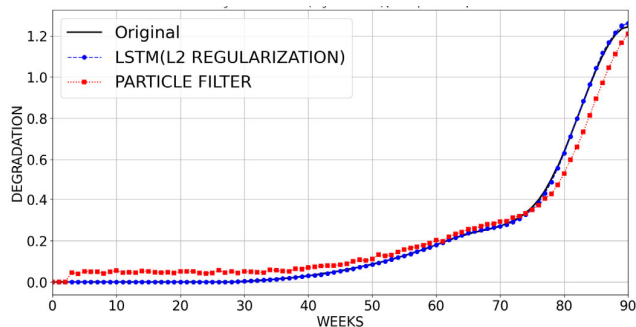


FIGURE 17. Degradation of span B starts segment-1st step prediction comparison of the proposed LSTM-based prognosis scheme with a particle filter-based statistical method.

The LSTM-based model and PF-based statistical model for ABCs degradation trend prediction are further compared for 1-8 step predictions. The RMSE-based comparison of both approaches is tabulated in Table 6. For a better visual representation, the RMSE-based comparison is also illustrated in Figure. 18.

It is evident from Figure 18, that the proposed method with L2 regularization outperforms the PF-based statistical method. The LSTM method predicts an error of 0.057 as compared to 0.114 of the particle filter method. LSTM performs 2 times better in terms of evaluation metric than the statistical (PF-based) method used for ABCs degradation prognosis.

TABLE 6. Comparison of the proposed LSTM-based prognosis scheme with a particle filter-based statistical method using the same dataset.

NUMBER OF STEPS	LSTM	PARTICLE FILTER
1st Step Error	0.00648263	0.041196919
2nd Step Error	0.01430561	0.054876762
3rd Step Error	0.02469821	0.079982197
4th Step Error	0.03978555	0.104727670
5th Step Error	0.05992040	0.128036344
6th Step Error	0.07693349	0.149832892
7th Step Error	0.10334202	0.168782621
8th Step Error	0.13435108	0.184706160

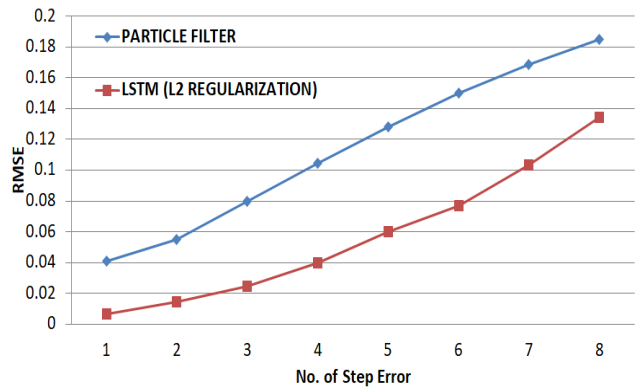


FIGURE 18. Comparison of the proposed LSTM-based prognosis scheme with a Particle filter-based statistical method using the same dataset.

LSTM models reported in this research work produced precise predictions of the TDP dataset. The model testing result proved the effectiveness of training as future trends were accurately predicted. However, the results can be improved in future research by implementing a deeper LSTM architecture with multiple LSTM layers stacked on top of each other and then connected with a dense layer [29]. More layers could become harder to train but carries the potential to offer improved results. Additionally, future work can include optimization of the regularization techniques used in the reported research work and applying the proposed model to similar degradation prediction problems.

VI. CONCLUSION AND FUTURE WORK

The degradation prognosis of critical industrial component(s) is of particular interest to the stakeholders for the upkeep of infrastructure and timely planning of maintenance activities. This research reports a novel LSTM-based prognosis algorithm for the degradation trend assessment of critical assets. The proposed scheme presents a dynamic training window approach, where the size of the training window is iteratively increased while keeping constant values of the lookback size and number of steps predicted ahead of time. The prediction horizon of the proposed approach is validated for up to 4 weeks as tabulated in Table 6. The reported study focuses on predicting the degradation of Aerial bundled cables (ABC) by applying the proposed LSTM to actual thermal degradation data recorded through a thermal

imaging camera. The proposed increasing sliding window-based LSTM scheme outperforms the classical approach. Three different variations of the proposed LSTM models were investigated for degradation assessment. The reported models include a generic LSTM with an Adam optimizer, an L2-regularization-based LSTM model, and an LSTM with a dropout layer. The three proposed LSTM models present promising results with high prediction accuracy. However, the L2 regularization-based proposed LSTM scheme outperforms the other presented methods. The L2 regularization-based proposed LSTM model was also compared to a particle filter-based statistical method. The proposed LSTM presented better results as compared to the statistical method. In the future, the model can be used for predicting the remaining useful life (RUL) or service life till the failure of the cables. LSTM can be used for dual purposes, first to predict the degradation of cables and then to estimate the remaining service life of cables to allow the maintenance managers to undertake timely repairs/ replacement of the damaged wires. In addition, different regularization techniques can be explored, such as the Monte Carlo dropout; which is a Bayesian approximation of the regular dropout. Furthermore, the proposed work can be used for cross-validation across different cable segments to verify generalization and with other time series data for degradation prediction in other domains, such as aircraft wings, structural health monitoring (SHM) of bridges, bearings, etc.

REFERENCES

- [1] B. Iung, M. Monnin, A. Voisin, P. Cochetoux, and E. Levrat, "Degradation state model-based prognosis for proactively maintaining product performance," *CIRP Ann.*, vol. 57, no. 1, pp. 49–52, 2008.
- [2] B. Iung, M. Véron, M. C. Suhner, and A. Muller, "Integration of maintenance strategies into prognosis process to decision-making aid on system operation," *CIRP Ann.*, vol. 54, no. 1, pp. 5–8, 2005.
- [3] W. B. Yousuf, T. M. R. Khan, S. T. Tariq, M. Ul-Hassan, and A. Shah, "Remaining useful life prediction of aerial bundled cables in coastal areas using thermal and corrosion degradation models," *IEEE Trans. Power Del.*, vol. 37, no. 4, pp. 2543–2550, Aug. 2022.
- [4] K. Alrajhi, "Automatic Arabic part-of-speech tagging: Deep learning neural LSTM versus Word2Vec," *Int. J. Comput. Digit. Syst.*, vol. 8, no. 3, pp. 307–315, Jul. 2019.
- [5] D. An, N. H. Kim, and J.-H. Choi, "Practical options for selecting data-driven or physics-based prognostics algorithms with reviews," *Rel. Eng. Syst. Saf.*, vol. 133, pp. 223–236, Jan. 2015.
- [6] B. H. Shah, "Kalman filter-based prediction of remaining useful life of aerial bundled cables," in *Proc. 3rd Int. Conf. Comput.*, 2020, pp. 1–6.
- [7] L. Boukezzi, L. Bessissa, A. Boubakeur, and D. Mahi, "Neural networks and fuzzy logic approaches to predict mechanical properties of XLPE insulation cables under thermal aging," *Neural Comput. Appl.*, vol. 28, no. 11, pp. 3557–3570, Nov. 2017.
- [8] M. Ma and Z. Mao, "Deep-Convolution-Based LSTM network for remaining useful life prediction," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1658–1667, Mar. 2021.
- [9] Y. Wu, M. Yuan, S. Dong, L. Lin, and Y. Liu, "Remaining useful life estimation of engineered systems using vanilla LSTM neural networks," *Neurocomputing*, vol. 275, pp. 167–179, Jan. 2018.
- [10] A. Muneer, S. M. Taib, S. M. Fati, and H. Alhussian, "Deep-learning based prognosis approach for remaining useful life prediction of turbofan engine," *Symmetry*, vol. 13, no. 10, p. 1861, Oct. 2021.
- [11] H. Miao, B. Li, C. Sun, and J. Liu, "Joint learning of degradation assessment and RUL prediction for aeroengines via dual-task deep LSTM networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5023–5032, Sep. 2019.
- [12] P. R. D. O. D. Costa, A. Akcay, Y. Zhang, and U. Kaymak, "Attention and long short-term memory network for remaining useful lifetime predictions of turbofan engine degradation," *Int. J. Prognostics Health Manage.*, vol. 10, no. 4, pp. 1–12, Jun. 2023.
- [13] H. Zhang, Q. Zhang, S. Shao, T. Niu, and X. Yang, "Attention-based LSTM network for rotatory machine remaining useful life prediction," *IEEE Access*, vol. 8, pp. 132188–132199, 2020.
- [14] E. Xu, Y. Li, Z. Han, J. Du, M. Yang, and X. Gao, "A method for predicting the remaining life of equipment based on WTTE-CNN-LSTM," *J. Adv. Mech. Design, Syst., Manuf.*, vol. 16, no. 1, 2022, Art. no. JAMDSM0001.
- [15] E. Phaisangittisagul, "An analysis of the regularization between L2 and dropout in a single hidden layer neural network," in *Proc. 7th Int. Conf. Intell. Syst., Model. Simul.*, 2016, pp. 174–179.
- [16] K. Li, X. Zhao, J. Bian, and M. Tan, "Sequential learning for multimodal 3D human activity recognition with long-short term memory," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Aug. 2017, pp. 1556–1561.
- [17] T. He and J. Droppo, "Exploiting LSTM structure in deep neural networks for speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 5445–5449.
- [18] Z.-H. Liu, X.-D. Meng, H.-L. Wei, L. Chen, B.-L. Lu, Z.-H. Wang, and L. Chen, "A regularized LSTM method for predicting remaining useful life of rolling bearings," *Int. J. Autom. Comput.*, vol. 18, no. 4, pp. 581–593, Aug. 2021.
- [19] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," *Heliyon*, vol. 4, no. 11, Nov. 2018, Art. no. e00938.
- [20] M. U. Hassan, T. M. Khan, S. Abbas, M. A. Shahid, S. T. Tariq, and F. Amir, "Degradation assessment of in-service aerial bundled cables in coastal areas leading to prognosis using infrared thermography," *IET Gener., Transmiss. Distrib.*, vol. 15, no. 8, pp. 1348–1356, Apr. 2021.
- [21] A. S. Almasoud, T. Abdalla Elfadil Eisa, F. N. Al-Wesabi, A. Elsaifi, M. Al Duhayyim, I. Yaseen, M. Ahmed Hamza, and A. Motwakel, "Parkinson's detection using RNN-Graph-LSTM with optimization based on speech signals," *Comput., Mater. Continua*, vol. 72, no. 1, pp. 871–886, 2022.
- [22] A. Shewalkar, D. Nyavanandi, and S. A. Ludwig, "Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU," *J. Artif. Intell. Soft Comput. Res.*, vol. 9, no. 4, pp. 235–245, Oct. 2019.
- [23] M. R. Soares, "Spacer cable and ABC distribution lines long-term analysis," in *Proc. Transmiss. Distrib. Conf. Expo.*, 1996, pp. 219–224.
- [24] V. T. Morgan, "The external thermal resistance of aerial bundled cables," *Generation, Transmiss. Distrib.*, vol. 140, no. 2, pp. 65–72, 1993.
- [25] (2021). *Advantages of Aerial Bundled Cables for Power Distribution*. Accessed: Aug. 16, 2021. [Online]. Available: <https://axis-india.com/2018/05/advantages-of-aerial-bundled-cables-for-power-distribution/>
- [26] W. B. Yousuf, T. Mairaj Rasool Khan, S. Abbas, and M. F. Hashmi, "Prognostic algorithm for degradation prediction of aerial bundled cables in coastal areas," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Qingdao)*, Oct. 2019, pp. 1–6.
- [27] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*.
- [28] X. Ying, "An overview of overfitting and its solutions," *J. Phys., Conf. Ser.*, vol. 1168, Feb. 2019, Art. no. 022022.
- [29] U. Yayan, A. T. Arslan, and H. Yucel, "A novel method for SoH prediction of batteries based on stacked LSTM with quick charge data," *Appl. Artif. Intell.*, vol. 35, no. 6, pp. 421–439, May 2021.



WALEED BIN YOUSUF received the B.E. degree in telecommunication engineering from Iqra University, Pakistan, in 2013, and the M.S. and Ph.D. degrees in electrical engineering from the National University of Sciences and Technology (NUST), Pakistan, in 2016 and 2022, respectively. He is currently an Assistant Professor with Iqra University. His research interests include structural health monitoring, bio-medical imaging, and reliability availability and maintainability (RAM) schemes for industrial infrastructure.



SYED MUHAMMAD UMAR TALHA received the Bachelor of Engineering and master's degrees from Iqra University, Pakistan, in 2008 and 2011, respectively, and the Ph.D. degree in electrical engineering from the National University of Science and Technology, Pakistan, in 2021. His research interests include computer vision, medical imaging, machine learning, optimization, inverse problems, and signal processing.



SYED MUHAMMAD DANIYAL received the M.S. degree in computer science from Iqra University, Pakistan, with a focus on artificial intelligence, medical imaging, computer vision, and machine learning. He is currently a Junior Lecturer with the Faculty of Engineering Science and Technology, Computer Science Department, Iqra University.



ABDUL AHAD ABRO received the Ph.D. degree in computer engineering from Ege University, Türkiye, with a specialization in artificial intelligence, deep learning, computer vision, and medical imaging. He is currently the Head of the Cyber Security Department and the Postgraduate Program, Faculty of Engineering Science and Technology, Iqra University, Pakistan. His research interests include conceptualization, design, and implementation of artificial intelligence solutions. He has made significant contributions to international journals and conferences in various roles, including an author, a reviewer, an editor, and a program committee member.

NAVEED AHMAD is currently an Associate Professor with the College of Computer and Information Science, Prince Sultan University. His research interests include critical areas of cybersecurity, privacy, blockchain technology, and intelligent transport systems. He is dedicated to advancing knowledge in these fields, contributing to the development of secure and efficient technologies that address contemporary challenges in digital privacy and transportation. With a strong background in academia and research, he continues to play a pivotal role in shaping the future of secure computing systems.



SADIQUE AHMAD (Member, IEEE) received the master's degree from the Department of Computer Sciences, IMSciences University, Peshawar, Pakistan, in 2015, and the Ph.D. degree from the Department of Computer Sciences and Technology, Beijing Institute of Technology, China, in 2019. He is currently a Researcher with Prince Sultan University, Riyadh, Saudi Arabia. He has achieved over 80 research papers in peer-reviewed journals and conferences. He is also focusing on deep cognitive modeling in social cybersecurity, the IoT trust management, and object detection.



ABDELHAMIED ASHRAF ATEYA (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Zagazig University, Egypt, in 2010 and 2014, respectively, and the Ph.D. degree from Saint Petersburg State University of Telecommunications, Russia, in 2019. He is currently an Assistant Professor with the Department of Electronics and Communications Engineering, Faculty of Engineering, Zagazig University. He has co-authored more than 50 publications in high-ranked journals. His current research interests include machine learning applications in communication networks, 5G/6G communications, the Internet of Things, tactile Internet and its standardization, and vehicular communications. He is a member of many scientific communities and an ACM professional member. He has been an active member of several international journals and conferences, with a contribution as an author, a reviewer, an editor, and a member of program committees.

...