Article

Deep-Learning Based Prognosis Approach for Remaining Useful Life Prediction of Turbofan Engine

Amgad Muneer ^{1,2,*} , Shakirah Mohd Taib ^{1,2} , Suliman Mohamed Fati ³ and Hitham Alhussian ^{1,2}

¹ Computer and Information Sciences Department, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia; shakita@utp.edu.my (S.M.T.); seddig.alhussian@utp.edu.my (H.A.)

² Centre for Research in Data Science (CERDAS), Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia

³ College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia; smfati@yahoo.com

* Correspondence: amgad_20001929@utp.edu.my

Abstract: The entire life cycle of a turbofan engine is a type of asymmetrical process in which each engine part has different characteristics. Extracting and modeling the engine symmetry characteristics is significant in improving remaining useful life (RUL) predictions for aircraft components, and it is critical for an effective and reliable maintenance strategy. Such predictions can improve the maximum operating availability and reduce maintenance costs. Due to the high nonlinearity and complexity of mechanical systems, conventional methods are unable to satisfy the needs of medium- and long-term prediction problems and frequently overlook the effect of temporal information on prediction performance. To address this issue, this study presents a new attention-based deep convolutional neural network (DCNN) architecture to predict the RUL of turbofan engines. The prognosability metric was used for feature ranking and selection, whereas a time window method was employed for sample preparation to take advantage of multivariate temporal information for better feature extraction by means of an attention-based DCNN model. The validation of the proposed model was conducted using a well-known benchmark dataset and evaluation measures such as root mean square error (RMSE) and asymmetric scoring function (score) were used to validate the proposed approach. The experimental results show the superiority of the proposed approach to predict the RUL of a turbofan engine. The attention-based DCNN model achieved the best scores on the FD001 independent testing dataset, with an RMSE of 11.81 and a score of 223.

Keywords: remaining useful life; fault diagnosis; CMAPSS; turbofan engine; deep learning; convolutional neural network

Citation: Muneer, A.; Taib, S.M.; Fati, S.M.; Alhussian, H. Deep-Learning Based Prognosis Approach for Remaining Useful Life Prediction of Turbofan Engine. *Symmetry* **2021**, *13*, 1861. <https://doi.org/10.3390/sym13101861>

Academic Editor: Vyacheslav Yukalov

Received: 17 September 2021

Accepted: 29 September 2021

Published: 3 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the aerospace industry, safety and reliability are key factors that require great attention due to the tough working conditions and long operating hours for aerospace systems [1]. One of the safety issues is turbofan engine failure, which is a critical component in airplanes. In addition, the turbofan engine is a very sophisticated and accurate piece of thermal equipment [2] in an airplane, involved with 60% of airplane issues. Therefore, any failures should be detected at the soonest possible moment. Such early detection can help to avoid any catastrophic damage or abrupt halting that may lead to economic and human losses [3]. Therefore, predictive maintenance and monitoring are a necessity to build a cost-effective maintenance strategy. The maintenance approach should be stable and flexible to increase the system's reliability and efficiency, and availability results in reduced downtime and operating costs [4,5].

Therefore, mechanical equipment prognostics and health management (PHM) have received considerable attention, and the prediction of RUL is at the center of PHM. Thus, with the internet industrial revolution, maintenance strategies have shifted from preventive

maintenance based on reliability assessments to predictive (condition-based) maintenance (CBM) [6]. The implementation process of condition-based maintenance to PHM includes data collected from advanced sensors, signal processing, feature extractions of the degradation characteristics of the rotary equipment, anomaly detection [7], fault identification and classification, fault prediction, and maintenance scheduling.

RUL prediction methods are mainly classified into two categories: model-based methods and data-driven methods. Model-based methods construct the model using mechanical principles, which takes some time. However, developing an accurate model is challenging due to the complex system structure and uncertain environment [8]. The empirical knowledge-based strategy necessitates the utilization of considerable prior information from industry specialists to develop a matching knowledge base. The knowledge-based strategy does not require a precise model; therefore, prediction accuracy cannot be guaranteed [9]. Data-driven approaches generate estimate models based on historical run-to-failure data, avoiding the limitations of physical failure models and expert knowledge [9,10]. Furthermore, data-driven methods overcome the gaps of both methods [11] and benefit from low computing costs and excellent accuracy. This study focuses primarily on data-driven methods for predicting the RUL.

One of the data-driven RUL prediction directions involves the use of deep learning (DL). DL is emerging as a new paradigm in the machine learning field and has provided an attractive opportunity to handle historical data and has achieved remarkable results in RUL-related fields [12]. The authors in [13] were the first to use the deep CNN method for RUL prediction. The results showed that CNN outperformed the multi-layer perceptron approach, support vector regression, and significance vector regression. In [14], a new CNN-based method for predicting bearing RUL was proposed, whereby the CNN model was combined with a smoothing process. The authors in [15] suggested a CNN-LSTM approach for RUL estimation, and the CNN method was found to be effective for local feature extraction, whereas the long short-term memory (LSTM) method was utilized to capture the degradation process. On the other hand, traditional deep learning methods analyze and mine sensor data using signal processing techniques, extract features that reflect system degradation and failure, and thus implement RUL predictions for rotary equipment. However, it remains challenging to develop an effective approach to mining complex time series data and to achieve highly accurate predictions [16], because data collected from the operation of a turbofan engine are non-linear [17], time-variant [8,18], large in scale, and high in dimension [7,19,20]. This leads to difficulty in discovering the failure trends in the input data [11] and results in the inability of models to extract highly abstract features [9,10,21], which hinders models from mapping the non-linear relationship between the extracted features and the corresponding RUL [11,12,21–23], leading to high computing costs and low prediction accuracy. To cope with this problem, feature extraction is critical in order to effectively capture the most useful information from high-dimensional data. Thus, we have proposed an attention-based deep convolutional neural network (DCNN) model by incorporating the prognosability method for the evaluation and selection of features, which results in a significant dimensionality reduction. DCNN has excellent learning ability, which is mainly achieved by using multiple non-linear feature extraction [14]. It can automatically learn hierarchical representations from numerical data [24]; in addition, the fully connected layer of the CNN simply learns the non-linear combination features after enhancing the ability of the DCNN model to learn new features, using an attention mechanism. The attention mechanism is an excellent way of enhancing the ability of a model to learn new features [25]. It acts as a secondary screening process of data information to emphasize the most critical pieces of information for analysis, ensuring an accurate prediction [26].

Therefore, we have proposed an attention-based DCNN model incorporating the time window (TW) approach for RUL prediction. The proposed model can extract and learn the turbofan engine system degradation-related features by applying a time window approach to the raw sensor data. Then, a new attention-based DCNN model can successfully extract

high-level abstract features through a deep learning network. The corresponding RUL value may be estimated using the learned representations. The suggested model, which employs a time window, attention mechanism, and a deep CNN structure, is intended to provide higher prognostic accuracy than shallow or typical machine learning methods presented in the literature.

The effectiveness of this approach was validated on C-MAPSS turbofan engine benchmark datasets provided by the National Aeronautics and Space Administration (NASA). The main contribution of this paper is twofold. First, in this paper we propose a data-driven approach to predicting the RUL of a turbo engine system using an attention-based DCNN model and a time window approach used for preparing the sample to results in better feature extraction.

Second, we have employed a measure of prognosability (feature ranking and selection) as indicators of the engine's condition at failure. The features with less variability were thereby eliminated to improve the prediction accuracy.

The rest of this paper is structured as follows. Section 2 highlights the background of the study and the related works. Then, Section 3 outlines the proposed research methodology, whereas Section 4 describes the experimental findings. Lastly, Section 5 concludes the paper and highlights the future work.

2. Related Works

Prognostic health management (PHM) refers to a process system that can forecast the future health status of mechanical systems in the engineering field. PHM is critical to ensuring the reliability of machinery systems, and it depends on the sensor's capabilities and analysis by monitoring the condition of mechanical components to measure their health portion [20,25]. Therefore, accurate RUL predictions are essential in PHM for many fields, including manufacturing and various industrial cyber-physical systems [27]. Furthermore, if the exact mechanical equipment RUL is known, manufacturing industries can plan future maintenance ahead of time and guarantee a seamless repair and maintenance process. Consequently, RUL prediction methods are mainly classified into physical model-based methods and data-driven methods.

2.1. RUL Prediction Based on Physical Models

Model-based approaches typically build a degradation model for rotary equipment such as a turbofan engine based on its physical structure, which is then used to predict RUL. For example, the authors in [28] present a fatigue-crack growth law, using fracture mechanics knowledge to demonstrate the fatigue-crack development model's application. Jiang [29] suggested a method for predicting RUL based on a model of convex optimization-life parameter deterioration. Gao et al. [30] proposed a physical model for RUL prediction based on Bayesian theory. The authors in [31] developed a Hertzian contact dynamic theory model of a ball bearing and raceway and showed that appropriate damping can extend the life of the bearing. Model-based solutions necessitate the establishment of an accurate degradation model; however, the complex structure of components and operating mechanisms and the inherent uncertainties in engineering practices make this problematic.

Additionally, as the structure of the mechanical system becomes more complex, those physical models that are primarily concerned with exploiting the system fault mechanism may not be the most feasible for practical prognostications of complex mechanical systems, such as turbofan engines or solar applications [32], because the uncertainty in the machining process and measurement noise are not incorporated into the physical models.

2.2. RUL Prediction Based on Data-Driven Models

Over the last few years, data-driven prognostics has shown a massive interest in establishing a link between the data collected from monitoring rotary equipment and the relevant RUL. As a result, numerous machine learning algorithms, most notably those based on neural networks, have been utilized to perform mapping between the gathered feature data and the related RUL. The benefit of using DL-based methods in prognosis models is that DL

can accurately simulate extremely non-linear, complex, multi-dimensional structures without previous information on the system's physical behavior. In addition, numerous forms of engineering system data, including raw sensor measurements, can be utilized directly as model inputs to estimate the RUL based on history trajectory data, which is significant for enhancing the reliability and safety of turbofan engine systems. Data-driven approaches generate estimate models based on historical run-to-failure data, avoiding the limitations of physical failure models and expert knowledge [9,10]. For instance, the authors in [33] suggested an LSTM-based scheme for the RUL estimation of aeroengines in the case of high noise levels, hybrid faults, and complex operations, as an enhancement of the standard recurrent neural network (RNN) approach. The authors in [34] used LSTM for a tool wear health monitoring task. The authors in [23] suggested an optimized DL-based method for multi-bearing RUL collaborative predictions by integrating both time and frequency domain functions. Numerical tests validated the developed method's feasibility and its superiority on a real dataset. Therefore, a restricted Boltzmann machine method was proposed by the authors in [35] for the learning of feature representation to calculate the RUL of machines, using the new concept of regularization and an unsupervised algorithm for a self-organizing map. The authors in [36] suggested a multi-objective deep belief network (MODBN) ensemble approach. An evolutionary method was combined with a standard DBN train method to develop several DBNs simultaneously, while keeping accuracy and diversity in mind.

In [37], the LSTM was suggested as a version of the RNN for turbofan engine RUL estimations. In the field of sequence prediction, the use of LSTM is popular, but it is time-consuming. The use of CNN for RUL estimations for the same aeroengine was proposed in [14]. The authors used a TW method to prepare the raw data samples, which helps to collect more degradation data. As a result, the model dimension inputs grow, making the development of a neural network model a difficult task. This raises the question of how to build network layers and network nodes to prevent overfitting, and to minimize time and computational costs, while avoiding local minima.

Another recent study was conducted by Peng et al. [2]. The integration of a 1-D convolutional neural network with a complete convolutional layer (1-FCLCNN) and LSTM was suggested as a tool for predicting RUL. To extract the spatial and temporal features of the FD001 and FD003 datasets developed with a turbofan engine, this method uses LSTM and 1-FCLCNN. CNN applications in RUL-related fields have also attracted much attention from various researchers [12]. The authors in [13] were the first to use the deep CNN approach for RUL predictions. The results showed that CNN outperformed the support vector regression, multi-layer perceptron, and significance vector regression approaches. The CNN approach that was proposed in [13] has been tested and evaluated on the C-MAPSS dataset, for which the RMSE obtained was reported to be 18.45.

Another study [38] presented a novel method for deep feature learning for RUL prediction using time-frequency representation (TFR) and multi-scale CNN networks (MSCNN). TFR can effectively disclose the non-stationary nature of the bearing degradation signal. Using wavelet transform, the authors obtained TFRs that contain a wealth of important information after accumulating time series degradation signals. Due to their high dimensionality, bilinear interpolation was used to minimize the size of these TFRs, which were then employed as inputs for the deep learning model. However, the proposed approach [38] still has some drawbacks. First, the training time of the algorithm is slow, and the computation speed needs to be increased. Furthermore, a graphic processing unit (GPU) is needed to help to process the original TFR. A different study was proposed by Zhang et al. [39], known as CNN-XGB (eXtreme gradient boosting), with an extended time window to tackle an issue affecting aero-engine systems, namely, that these systems typically operate under a variety of operating conditions, which may affect the deterioration trajectory of the system differently and hence impair the accuracy of the RUL prediction. The proposed approach was validated using NASA C-MAPSS turbofan aeroengine datasets. The RMSE obtained was 20.3, and the training time was reported to be 621.7 s. A summary of the literature is provided in Table 1.

Table 1. Summary of the findings of the state-of-art methods.

Authors	Year	Method Used	Benchmark Dataset	Results Achieved	Limitations/Gaps
Peng et al., [2]	2021	FCLCNN-LSTM	C-MAPSS	This model verified only with FD001 (11.17) and FD003 (9.99) subset of data	The key drawback of this model is the need to incrementally update the prognosis results.
Wen, Dong and Gao [12]	2019	ResCNN	C-MAPSS	RMSE for FD001 (12.16)— RMSE for FD002 (20.85)— RMSE for FD003 (12.01)— RMSE for FD004 (24.79)	The limitations of the proposed method are that the imbalance of signal data is ignored, and the tuning parameter process of the ensemble ResCNN is very time-consuming.
Babu et al., [13]	2016	First attempt at a deep CNN	C-MAPSS	RMSE for FD001 (18.44)— RMSE for FD002 (30.29)— RMSE for FD003 (19.81)— RMSE for FD004 (29.15)	The limited accuracy of the RUL estimation, means this method is not practical for real-world applications.
Li, Ding & Sun, [14]	2018	DCNN	C-MAPSS	RMSE for FD001 (18.45)— RMSE for FD002 (22.36)— RMSE for FD003 (12.64)— RMSE for FD004 29.16	Additional architecture improvements are required, as the current training time exceeds that of the majority of shallow networks in the literature.
Zhang et al., [36]	2016	Multi-objective deep belief network ensemble	C-MAPSS	RMSE for FD001 (15.04)— RMSE for FD002 (25.05)— RMSE for FD003 (12.51)— RMSE for FD004 (28.66)	This model suffers from slow prediction process and limited accuracy of RUL estimation, which made it not cost-effective method in industrial contexts.
Zheng et al., [37]	2017	Deep LSTM	C-MAPSS	RMSE for FD001 (16.14)— RMSE for FD002 (24.49)— RMSE for FD003 (16.18)— RMSE for FD004 (28.17)	The main drawback can be summarised in twofold. First, the limited accuracy of the RUL prediction, which make this method is not practical for industrial contexts. Second, high computational load.
Zhu, Chen and Peng [38]	2018	multi-scale CNN	PRONOSTIA	Tested on bearing dataset	Further architecture improvements are required, as the current model need more optimization.
Zhang et al., [39]	2019	CNN-XGB	C-MAPSS	RMSE for FD001 (12.61)— RMSE for FD002 (19.61)— RMSE for FD003 (13.01)— RMSE for FD004 (19.41)	This main drawback of this method is the computational speed and cost, with a prediction time of around 621.7 s. It not cost-effective model in industrial contexts.
This study	2021	Attention-based DCNN	C-MAPSS	RMSE for FD001 (11.81)— RMSE for FD002 (18.34)— RMSE for FD003 (13.08)— RMSE for FD004 (19.88)	The proposed model training time is 142 s, which shows its superiority in reducing the training time and model complexity compared to several popular methods in the literature

When the operating conditions are more complex, the RUL prediction is more challenging, and this kind of problem deserves further study.

3. The Proposed Approach

This study provides an optimized DCNN-based method architecture. Figure 1 indicates the approach to RUL prediction proposed in this study. Essentially, it consists of four distinctive parts: data pre-processing, feature extraction, model training, and estimating remaining useful life.

Figure 1. Flowchart of the proposed RUL prediction approach.

The proposed DCNN-based model is trained and evaluated, applying standard performance evaluators for prediction models to achieve the optimum model for predicting the RUL of different engine units. In the first section, the C-MAPSS benchmark dataset is introduced. The second section focuses on the proposed deep candidate model, whereas the final two steps of the suggested approach are discussed in subsequent sections.

3.1. C-MAPSS Benchmark Dataset

In this study, we selected the C-MAPSS aero-engine degradation dataset provided by NASA to verify the effectiveness of the proposed DNN candidate models. The primary control system comprises three components: a fan controller, a regulator, and a limiter. The fan maintains normal flight conditions by directing air into the inner and outer culverts (Figure 2). The combustor is supplied with compressed high-temperature, high-pressure gases via a low-pressure compressor (LPC) and a high-pressure compressor (HPC). Low-pressure turbines (LPTs) can decelerate and pressurize air, increasing aviation kerosene's chemical energy conversion efficiency. High-pressure turbines (HPT) generate mechanical energy by striking turbine blades with high temperatures and high-pressure gas. The low-pressure rotor (N1), high-pressure rotor (N2), and nozzle all contribute to the engine's combustion efficiency.

Figure 2. C-MAPSS aero-engine diagram [40].

The C-MAPSS is widely used in prognostic studies, containing four sub-datasets under different operating conditions (OCs) and failure modes (FMs). Every sub-dataset includes a training set, a testing set, and testing RUL values, and consists of twenty-one sensors and three operation settings [41]. Each engine unit has varying degrees of wear. Over time, the engine units start to degrade until they reach system failure, which is described as an unhealthy time cycle. Therefore, the sensor records in the testing set are terminated before the occurrence of a system fault. Table 2 provides information on the turbofan degradation engine systems dataset.

Table 2. The information of the different subsets of data in the C-MAPSS dataset.

Dataset	C-MAPSS			
	FD001	FD002	FD003	FD004
Training Units (N)	100	260	100	249
Testing Units	100	259	100	248
Operating Conditions (OC)	1	6	1	6
Fault modes (FM)	1	1	2	2
Training samples (default)	17,731	48,819	21,820	57,522
Testing samples	100	259	100	248

In this experiment, we aimed to predict the RUL of a single-engine unit randomly selected from the testing set. In this study, the four subsets of data, FD001–FD004, are used for attention-based DCNN model verifications. FD001 is the simplest subset of data and FD004 is the most complex subset of data, in which the engines have six OCs and two FMs. When the operating conditions are more complex, the RUL prediction is more challenging, and herein the proposed method's effectiveness is verified and validated with four different subsets of data.

3.2. Feature Selection Using the Prognosability Algorithm

To improve the RUL prediction accuracy, we have employed a feature selection metric called the prognosability algorithm. These metrics assign a numerical value to the recognized condition indicators on a scale from zero to one. A higher-ranked feature more accurately monitors the degradation process and is thus more suitable for training the RUL prediction model. Prognosability is a measure used to determine a feature's variability during failure in contrast to the range between its initial and end values. A more prognosis-able feature exhibits less volatility during failure in respect to the range between its initial and ultimate values. The Y values vary from zero to one, with one indicating that X is perfectly prognosable and zero indicating that X is not prognosable. The computation of prognosability uses this formula:

$$\text{prognosability} = \exp\left(-\frac{\text{std}_j(x_j(N_j))}{\text{mean}_j|x_j(1) - x_j(N_j)|}\right), j = 1, \dots, M \quad (1)$$

where x_j denotes the measurements vector made on the j^{th} system, M is the number of monitored systems, and N_j denotes the number of measurements made on the j^{th} system. Therefore, we have ranked the 21 features using prognosability, as presented in Figure 3. The selected features are (s2, s3, s4, s7, s8, s9, s11, s12, s13, s14, s15, s17, s20, s21) and the irregular or unchanged sensor data have been eliminated (e.g., s1, s5, s6, s10, s16, s18, s19). Thus, the same features were confirmed and used by Zhang et al. [36] and Duan et al. [42] as the feature inputs of the DCNN-based model after normalizing the inputs and preparing the raw samples using the time window approach.

Figure 3. Importance ranking of all features: feature selection using the prognosability metric.

3.3. Data Normalization

Data from several raw sensors, operating parameters, and runs to failure are granted in real-world applications. Sensor data must be standardized in regard to each sensor before training and testing since the value scales of different sensors can vary.

Therefore, sensor data can be normalized by letting μ_i denote the mean of the i -th sensor data from the engine and σ_i denote the standard deviation. In the Z-score normalization, x'_i is the normalized sensor output. The raw sensor data is scaled within the range of [0,1] using Min-Max normalization as follows.

$$x'_i = \frac{x_i - \mu_i}{\sigma_i} \quad (2)$$

$$x'_i = \frac{x_i - \min x_i}{\max x_i - \min x_i} \quad (3)$$

where $x^{i,j}$ is the i -th measuring point of the j -th sensor. $x^{i,j}_{norm}$ is the $x^{i,j}$ normalized result. $x^{j}_{\max}$ and $x^{j}_{\min}$ denote the maximum and minimum values of the j th sensor. Because the time series contains more information than a single point, in this study we implemented a TW to make use of multivariable temporal information, as in [14,36,42]. For the training data set FD0001, the time window length was selected as 35. All the historical data in the TW were gathered at each step to form a high-dimensional vector of length 14×35 as the input data. Thus, 14 sensors' measurements out of 21 sensors were employed as the raw input features, as performed in [14,43]. The dynamic characteristics of turbofan engine operating data under different operating conditions are significantly different, which leads to different network structures for the extraction of features. The proposed attention-based DCNN model structure was designed to predict the RUL of turbofan engines under both single and multiple OCs. Therefore, this paper utilizes the four subsets of the dataset

shown in Table 2, and the FD001 subset of data was used for experimental analysis, as in most of the literature [14,36,42,43]. Figure 4 shows the normalized sensor measurements.

Figure 4. Samples of normalized sensors' measurements.

3.4. Deep Convolutional Neural Networks

CNNs are designed to handle learning problems involving high-dimensional input data with complex spatial structures, such as image classification [44,45], text classification [46] video processing [47,48], amino acid sequence prediction [49,50], and time series failure signals. The main three layers of CNN are as follows.

3.4.1. Convolutional Layer

The convolutional layer is the most significant component of convolutional networks. Feature maps are generated by sliding the convolution kernel over the data and convolving with the covered data. Furthermore, the mutual weights property reduces model parameters and the possibility of overfitting. The calculation process of the i -th feature map of the l -th convolutional layer x_i^l , is as follows:

$$x_i^l = \varphi(z_i^l) \quad (4)$$

$$z_i^l = k_i^l * x^{l-1} + b_i^l = \sum_{c=1}^C k_{i,c}^l * x_c^{l-1} + b_i^l \quad (5)$$

where z_i^l represents the convolution operation's output, $*$ denotes the convolution operator, k_i^l is the i -th convolution kernel, x^{l-1} is the input volume, and b_i^l and $\varphi(z_i^l)$ represent the bias term and non-linear activation function. Finally, C denotes the number of input channels.

3.4.2. Pooling Layer

The pooling layer aims to combine similar features into one and speed up the calculation using a non-linear down-sampling function [51]. The most popular pooling layer is the max-pooling layer. The pooling layer's inputs are the previous layers' function maps, and the outputs are the limit of a local patch of the inputs. The following is the function:

$$x_i^l = \max(x_i^{l-1}, p, s) \quad (6)$$

where x_i^l is the i -th feature map of the l -th pooling layer, x_i^{l-1} is the i -th feature map in the previous layer $l-1$, $\max(\cdot)$ means the max-pooling, and p and s represent the pooling size and the stride size.

3.4.3. Fully Connected Layer

The fully connected layer summarizes the features and outputs prediction results as the final layer of the convolutional neural network [51]. The output x^l of the l -th fully connected layer is as follows:

$$x^l = \varphi(\omega^l x^{l-1} + b^l) \quad (7)$$

where x^{l-1} is the output of the previous layer $l-1$, ω^l and b^l represent the weight matrix and the bias vector.

CNNs attempt to learn hierarchical filters which can transform large input data to accurate class labels using minimal trainable parameters. This is accomplished by enabling sparse interactions between input data and trainable parameters through parameter sharing to learn equivariant representations (also called feature maps) of the complex and spatially structured input information [52]. In a deep CNN, units in the deeper layers may indirectly interact with a large portion of the inputs due to the usage of pooling operations, which replaces the output of net at a certain location with a summary statistic and allows the network to learn complex features from this compressed representation [14]. The so-called “top” of the CNN is usually composed of many fully connected layers, including the output layer, which uses the complex features, learned by previous layers, to make predictions.

The attention-based DCNN has excellent learning ability, which is mainly achieved by employing multiple non-linear feature extraction. The learning ability of a DCNN is enhanced by utilizing multiple non-linear feature extraction. Furthermore, it can learn hierarchical representations from data on its own. As a result, the scale of the convolution kernel and the number of convolution layers significantly affect the prediction performance. The proposed network architecture for the RUL estimation undertaken in this study is shown in Figure 5. The input data is two-dimensional (2D). The feature number is one dimension, and the sensor’s time sequence is the other.

Figure 5. The proposed architecture of the attention-based DCNN for RUL prediction.

The feature maps are then combined using a convolutional layer with one filter with size of 3×1 . The attention mechanism is used to extract high abstract degradation and trend features. After attention layer, the extracted feature will be connected with a fully connected layer. In addition, the dropout method will be used to relieve overfitting. Additionally, ReLU is the activation feature of each layer. In this study, the Adam optimization algorithm will serve as the optimizer. Adam is a rate-optimized adaptive learning method

utilized to train deep neural networks [53]. Given the current state of the turbofan engine datasets, we increased the penalty for lag prediction; the loss is denoted as follows.

$$\text{loss} = \frac{1}{N} \sum_{i=1}^N \omega (y_i - \hat{y}_i)^2 \quad (8)$$

where y_i is the actual value and $\hat{y}_i$ is the predicted value. N is the number of validations set. When the true value y_i is greater than the expected value $\hat{y}_i$, the penalty coefficient $\omega = 1$; otherwise, $\omega = 2$. The following section discusses the proposed attention mechanism for better degradation feature extraction for the engine system.

3.5. Proposed Attention Mechanism

The attention mechanism enables a model to narrow its focus on critical regions of the selected feature space. It operates by paying more attention to subsets of the data to obtain more optimal scores. The attention mechanism is summarized in three parts, as presented in Figure 6.

Figure 6. Proposed attention mechanism.

The proposed attention scheme modifies the method through which attention weights are calculated. Unlike conventional attention processes, this one employs the sigmoid activation function (10) rather than the SoftMax function. Because the SoftMax function normalizes the weights, it reduces the likelihood that more than one variable is relevant for prediction, as is frequently the case in a multivariate time series. This stage enables the model's attention mechanism to pick out the degradation characteristics more effectively.

$$u_i = \tanh(w_a h_i + b_w) \quad (9)$$

$$\alpha_i = \text{sigmoid}(v_a \cdot u_i) \quad (10)$$

$$v_j = \sum_{i=1}^{i=k} h_i \cdot \alpha_i \quad (11)$$

3.6. Time Window Technique

In multivariate time series-based issues such as RUL prediction, temporal sequence data usually include more information than multivariate data points taken at a single time step. Thus, the processing of time sequences has a great deal of promise in terms of improving prediction performance. In this study, we used a time frame to prepare the data to take advantage of multivariate temporal information.

Hence, N_{tw} represents the time window's size. All sensor data from the previous TW are compiled into a high-dimensional feature vector and utilized as the network's inputs at each time step. Figure 7 illustrates a normalized data sample from the 14 chosen sensors with a time window size of 35, concerning a single-engine unit in the training sub-dataset FD001.

Figure 7. Time window approach to the normalized sensor data.

For the RUL target label, a piecewise linear function is used, as in [42], which is defined as

$$Rul = \begin{cases} Rul, & \text{if } Rul \leq Rul_{\max} \\ Rul_{\max}, & \text{if } Rul > Rul_{\max} \end{cases} \quad (12)$$

where $Rul_{\max}$ is a preset value. $Rul_{\max}$ was set to 150 cycles for each subset of data, as in [39,43]. According to the experimental analysis, m was 35, and l was 1. FD001 had training samples of 17,731 and testing samples of 100, because only the most recent measurements of the test sets were used. The effectiveness of the piecewise linear function on this prediction problem has been confirmed in the literature [13,14,36]. Moreover, the processed label values were smoothed. Figure 8 shows the piecewise RUL target function of an engine unit which the full-time cycle is 130.

Figure 8. Piecewise linear RUL target function.

Lastly, Figure 1 depicts the proposed prognostic experimental approach. First, the FD001 subset of the data was pre-processed by selecting 14 raw sensor measurements and normalizing the accompanying data to fall within the range of $[-1,1]$. Next, the training and testing datasets were created, with each sample providing information about the time sequence within the length of time frame N_{tw} . Hence, the normalized data prepared in 2-dimensional format were directly fed into the attention-based DCNN model as inputs. As a result, hand-crafted signal processing features, such as skewness, kurtosis, and so on were unnecessary. Thus, the suggested method requires no prior knowledge of prognostics or signal processing. Moreover, we used a randomized search method to find the best optimal hyperparameters over a vast hyperparameter space. Randomized hyperparameter

search provides improved hyperparameters for the proposed DCNN model structure with a limited computing budget and faster convergence speed. The attention-based DCNN model's effectiveness is demonstrated in the following section.

4. Experimental Results and Discussion

This section summarizes the experimental findings and discusses their significance. In the first section, the experimental results are discussed. In the second section, the TW effects and the proposed model's training time are examined. Finally, a comparative analysis with literature is provided in the last section.

4.1. Experimental Results

After repeating the experiment ten times, the proposed algorithm's training parameters were tuned to obtain the best score value, shown in Equation (10), and the test set's lowest root mean square error, shown in Equation (11).

$$S = \begin{cases} \sum_{i=1}^N e^{-\frac{d_i}{13}} - 1, d_i < 0 \\ \sum_{i=1}^N e^{-\frac{d_i}{10}} - 1, d_i \geq 0 \end{cases} \quad (13)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N d_i^2} \quad (14)$$

According to the PHM Data challenge in 2008 [54], an asymmetric scoring function penalizes latent predictions $d_i \geq 0$ more severely than early predictions $d_i < 0$. This is for maintenance reasons. Predictions made too late may cause maintenance activities to be delayed, whereas predictions made too early may not be hazardous but consume more maintenance resources. Figure 9 demonstrates that the RMSE and score functions are sparser towards higher values than they are towards zero, confirming the results' validity.

Figure 9. Comparison between RMSE and scoring functions for the testing set results.

The proposed attention-based DCNN candidate model for RUL prediction was constructed, and the configuration was specified, including the number of hidden layers, as well as the number and the length of convolution filters. The parameters used in this experiment are presented in Table 3. The attention-based DCNN model received normalized training data as inputs and produced as an output labeled RUL values for the training

samples. The back-propagation learning method was employed to update the network's weights, and the Adam technique was employed simultaneously with mini-batches. The samples were randomly separated into numerous mini-batches of 512 samples for each training epoch and loaded into the training system. Following that, the network information, i.e., the weights in each layer, was optimized using the mini-mean batch loss function. It should be mentioned that the batch size selection affects the performance of network training [54]. Based on the trial results, a batch size of 512 samples was determined to be appropriate and was utilized in all of the case studies in this study.

Table 3. Training parameters of the proposed model.

Batch Size	Dropout Size	Epoch Number	Iteration per Epoch	Maximum Alteration	Num Hidden Units	Activation
512	0.5	20	32	640	1000	RELU

Moreover, a variable learning rate was used. The initial learning rate was 0.003 for the first 20 epochs of optimization. Following that, a learning rate of 0.0009 was utilized to ensure consistent convergence for the remaining 12 epochs. By default, the maximum number of training epochs is 640 for the attention-based DCNN candidate model.

The time window size is an essential factor affecting the prediction accuracy of the proposed model. Figure 10 shows the effect of the time window size on the model performance. The prediction results of RUL are affected by the amount of historical information. As shown in Figure 10, increasing the time window size can improve the prediction accuracy of the RUL of the engine. Note that the selected time window is determined by the number of the shortest cycle of the engine test set. Therefore, the time window sizes (Ls) of the FD001 and FD002 data sets were 30, and those of the FD003 and FD004 data sets were 35. Furthermore, we trained the proposed attention-based DCNN model 10 separate times to exclude the effects of random disturbances and to take the average of the results. The key parameters of the proposed model are summarized in Table 3.

Figure 10. Effect of time window size on the performance of the model.

In the attention-based DCNN model, the purpose of the pooling layer is to merge similar features into one using non-linear down-sampling functions and to speed up the calculation. The max-pooling layer is the most used pooling layer. The inputs of the pooling layer are the feature map from the previous layers, and the outputs are the maximum of a local patch of the inputs. However, an experiment was conducted to test and verify the effectiveness of the proposed DCNN model with and without the pooling layer, as shown in Table 4. Therefore, the experimental findings confirm that the attention-based DCNN model architecture without the pooling layer achieved a better result compared to the attention-based DCNN model architecture with the pooling layer, with a difference in RMSE error of about 6.42, where the model with the pooling layer obtained an RMSE of 21.34 and the model without the pooling layer obtained an RMSE of 14.92.

Table 4. The result of the model error with or without the pooling layer.

Pooling Layer	RMSE
With	21.34
Without	14.92

The turbofan engine degradation simulation data used in this study were numerical data, and the dimension of the raw feature was relatively low. Although the pooling operation improves the computing efficiency, some useful information is filtered in this prognostic approach. Table 4 shows the different prediction effects with and without the pooling layer in the model. The network structure without the pooling layer showed better results.

Figure 11 shows the RMSE of the attention-based DCNN model during the network training, with the graph showing that the more number alterations and epochs, the more the model accuracy improved. Therefore, the number of alterations was set to 32 per epoch. The maximum alteration of 640 was observed during the training, with a learning rate of 0.0009. Therefore, the effect of increasing the number of convolutional layers is presented in Figure 12, showing that the attention-based DCNN model achieved the lowest RMSE with five convolutional layers.

Figure 11. Effect of the number of epochs and number of alterations per epoch on RMSE and model learning. **Figure 12.** The effect of increasing the number of convolutional layers in the attention-based DCNN model.

4.2. Case Study of Turbofan Engine System

Four verification tests were conducted on the four subsets of the C-MAPSS dataset to verify the effectiveness of the attention-based DCNN model. Each subset of the data had different operating conditions. When the operating conditions are more complex, the RUL prediction is more challenging. Figure 13 shows the RUL prediction results of two verification tests performed on FD001 and FD002 (engine 73 and engine 39, randomly

selected cases). Figure 14 shows the other two evaluation tests conducted on the FD003 and FD004 subset of data. The prediction results of the proposed model were more accurate in relation to the engine degeneration. This is because the model can extract more failure features from the sensor data with increasing degradation. The safety of the system can be improved by accurately predicting the RUL near the stage of the engine failures.

Figure 13. Two cases of lifetime RUL predictions for the tested engine units: (a) RUL prediction of a randomly selected case in the FD001 subset of data; (b) RUL prediction of a randomly selected case in the FD002 subset of data.

Figure 14. Two cases of lifetime RUL predictions for the tested engine units: (a) RUL prediction of a randomly selected case in the FD003 subset of data; (b) RUL prediction of a randomly selected case in the FD004 subset of data.

In the context of aerospace industry risk management, deep-learning-based RUL prediction can assist managers in assessing the likelihood of a system failure prior to a maintenance window. Therefore, maintenance time is fixed in large-scale manufacturing, and several forms of equipment are maintained in a single maintenance session. It is not feasible or cost-effective to maintain all machines in a single window. As a result, the manager must pick which equipment will be serviced during the scheduled maintenance window. A density chart of RUL prediction error can be generated using the RUL model-based deep learning process. As illustrated in Figure 15, the manager can predict the likelihood of a machine failing before the next maintenance window. The management team has the option of adding a machine to the present maintenance list or deferring it until the next maintenance window.

Additionally, the proposed approach shows high robustness and generalization ability, and it can be used in practice as an industrial condition-based maintenance strategy for several manufacturing industries.

Figure 15. For maintenance, the RUL error distribution and confidence interval.

4.3. Comparison with Literature

This section compares the proposed attention-based DCNN predictor with state-of-the-art methods. In the literature, different DL methods have been used to predict RUL using the C-MAPSS benchmark dataset. Table 5 shows a comparison of the proposed deep candidate model with related literature contributions. The comparison is only demonstrated for the available metrics, but essentially, it conveys the promising results of the proposed attention-based DCNN predictor. The results confirmed that the attention-based DCNN model surpasses the other methods for predicting RUL on the entire benchmark independent testing data set. Only one study obtained a better result in the FD003 and FD004 subset of data.

Table 5. Comparison of the proposed method with related literature contributions.

Prediction Model	C-MAPSS				
	Measure	FD001	FD002	FD003	FD004
Proposed attention-based DCNN Predictor	RMSE—	11.81—	18.34—	13.08—	19.88—
	Score	223.0	2550	280.5	2982.31
CNN-XGB [39]	RMSE—	12.61—	19.61—	13.01—	19.41—
	Score	224.73	2525.99	279.36	2930.65
MODBNE [36]	RMSE—	15.04—	25.05—	12.51—	28.66—
	Score	334.23	5585.34	6557.62	6557.62
Echo State Network with Kalman Filter [55]	RMSE—	63.46—	—	—	—
	Score	-	-	-	-
ANN-EN [56]	RMSE—	14.39—	29.09—	15.42—	34.74—
	Score	337	-	533	-
MLP [13]	RMSE—	37.56—	80.03—	37.56—	77.36—
	Score	17972	780280	17409	5616600
Deep CNN [13]	RMSE—	18.45—	30.29—	19.81—	29.16—
	Score	1286.7	13570	1596.2	7886.4
DW-RNN [57]	RMSE—	22.52—	25.90—	18.75—	24.44—
	Score	N/A	N/A	N/A	N/A
MTL-RNN [57]	RMSE—	21.47—	25.78—	17.98—	22.82—
	Score	N/A	N/A	N/A	N/A
DCNN [14]	RMSE—	12.61—	22.36—	12.64—	23.31—
	Score	273.7	10412.0	284.1	12466

Table 5 shows that our proposed attention-based DCNN model has outperformed all the previous models in the literature. Based on the experimental findings, it was observed that increasing the TW results in improving the RUL prediction accuracy. The proposed attention-based DCNN model predicts the RUL of turbofan engines with high accuracy and without the requirement to comprehend the engine construction or failure mechanism and without the need for professional knowledge and experience. It simplifies the modeling process and can serve as a decision-making tool for aircraft engine maintenance and health management.

Additionally, the visual processing of sensor signals using time and frequency domains has achieved excellent results, thus showing its superiority in the diagnosis and examination of rotary machines and resolving the gap between experts at different levels. Further research and popularization will be of great significance for diagnosing such a complex turbofan engine without prior knowledge of system degradation. It could significantly reduce the incidence rate of system failure and improve RUL estimations and industry maintenance strategies.

5. Conclusions

In this paper, a data-driven method-based deep learning approach was proposed to predict the remaining useful life (RUL) of a turbofan engine. Deep learning tends to give decision-makers new insights into their operations, real-time performance indicators, and costs. In this study we aimed to accurately predict the remaining useful life of the turbofan engine, which is significant for improving the reliability and safety of turbofan engine systems. Therefore, a time window technique was adopted to prepare the samples of raw data to fit directly into the proposed model. The dropout method was used to relieve the overfitting issue during the training of the model. The attention mechanism was integrated with DCNN structure to mine useful degradation features from complex historical data. The proposed model's superiority and effectiveness were verified using the C-MAPSS benchmark dataset. The experimental results showed a minimal error between the estimated and the true RUL value in the testing subset of data of the engine units. In addition, the selected time window size significantly improved the prediction performance of the model.

Additionally, during the experiment, it was observed that as the degree of degradation increases, the prediction results are more accurate. Thus, while the proposed approach obtains good experiment results, future architecture optimization is necessary. As with all empirical research, this study has significant limitations. For instance, the model can be further enhanced in future works by increasing the number of convolutional nuclei and hidden neurons in the fully connected layer. Furthermore, it is well-known that in measurements, there are uncommon, inconsistent observations that are outnumbered by most of the other observations, referred to as anomalies. Lastly, since the raw vibration signals are used directly as the input, this diagnostic model requires a more complicated network structure to verify the correctness of the results, resulting in a high calculation load. Thus, a deep hybrid learning model and further signal pre-processing implementations will be investigated to eliminate duplicate information and obtain fault characteristics.

Author Contributions: Conceptualization, A.M., S.M.T. and H.A.; methodology, A.M. and S.M.T.; software, A.M.; validation, A.M. and S.M.F.; formal analysis, S.M.T. and A.M.; data curation, A.M. and S.M.T.; writing—original draft preparation, A.M.; writing—review and editing, S.M.T., H.A. and S.M.F.; visualization, A.M. supervision, S.M.T. and H.A.; project administration, A.M. and S.M.T.; funding acquisition, S.M.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research/paper was fully supported by Universiti Teknologi PETRONAS, under the Yayasan Universiti Teknologi PETRONAS (YUTP) Fundamental Research Grant Scheme (YUTP-015LC0-123).

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: The dataset used in this study available on NASA repository and it is called Commercial Modular Aero-Propulsion System Simulation (C-MAPPS) dataset (<https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/>), (accessed on 1 September 2021)).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Chen, X.; Wang, S.; Qiao, B.; Chen, Q. Basic research on machinery fault diagnostics: Past, present, and future trends. *Front. Mech. Eng.* **2018**, *13*, 264–291. [CrossRef]
- Peng, C.; Chen, Y.; Chen, Q.; Tang, Z.; Li, L.; Gui, W. A Remaining Useful Life Prognosis of Turbofan Engine Using Temporal and Spatial Feature Fusion. *Sensors* **2021**, *21*, 418. [CrossRef] [PubMed]
- Wang, L.; Zhang, Z.; Long, H.; Xu, J.; Liu, R. Wind turbine gearbox failure identification with deep neural networks. *IEEE Trans. Ind. Inform.* **2016**, *13*, 1360–1368. [CrossRef]
- Zhao, G.S.; Wu, S.S.; Rong, H.J. A Multi-Source Statistics Data-Driven Method for Remaining Useful Life Prediction of Aircraft Engine. *J. Xi'an Jiaotong Univ.* **2017**, *51*, 150–155.
- Zhang, Z.; Si, X.; Hu, C.; Lei, Y. Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods. *Eur. J. Oper. Res.* **2018**, *271*, 775–796. [CrossRef]
- Zschech, P.; Bernien, J.; Heinrich, K. Towards a Taxonomic Benchmarking Framework for Predictive Maintenance: The Case of NASA's Turbofan Degradation. In Proceedings of the Fortieth International Conference on Information Systems (ICIS 2019), Munich, Germany, 15–18 December 2018; 2018; pp. 1–15.
- Muneer, A.; Taib, S.M.; Fati, S.M.; Balogun, A.O.; Aziz, I.A. A Hybrid Deep Learning-Based Unsupervised Anomaly Detection in High Dimensional Data. *Comput. Mater. Contin.* **2021**, *71*. [CrossRef]
- Wei, J.; Bai, P.; Qin, D.; Lim, T.C.; Yang, P.; Zhang, H. Study on vibration characteristics of fan shaft of geared turbofan engine with sudden imbalance caused by blade off. *J. Vib. Acoust.* **2018**, *140*, 041010. [CrossRef]
- Si, X.S.; Wang, W.; Hu, C.H.; Zhou, D.H. Remaining useful life estimation—a review on the statistical data driven approaches. *Eur. J. Oper. Res.* **2011**, *213*, 1–14. [CrossRef]
- Ahmadzadeh, F.; Lundberg, J. Remaining useful life estimation. *Int. J. Syst. Assur. Eng. Manag.* **2014**, *5*, 461–474. [CrossRef]
- Xie, Z.; Du, S.; Lv, J.; Deng, Y.; Jia, S. A hybrid prognostics deep learning model for remaining useful life prediction. *Electronics* **2021**, *10*, 39. [CrossRef]
- Wen, L.; Dong, Y.; Gao, L. A new ensemble residual convolutional neural network for remaining useful life estimation. *Math. Biosci. Eng.* **2019**, *16*, 862–880. [CrossRef]
- Babu, G.S.; Zhao, P.; Li, X.L. Deep convolutional neural network-based regression approach for estimation of remaining useful life. In Proceedings of the International Conference on Database Systems for Advanced Applications, Dallas, TX, USA, 16–19 April 2016; Springer: Cham, Switzerland, 2016; pp. 214–228.
- Li, X.; Ding, Q.; Sun, J.Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliab. Eng. Syst. Saf.* **2018**, *172*, 1–11. [CrossRef]
- Hinchi, A.Z.; Tkouat, M. Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network. *Procedia Comput. Sci.* **2018**, *127*, 123–132. [CrossRef]
- Agrawal, S.; Sarkar, S.; Srivastava, G.; Maddikunta, P.K.R.; Gadekallu, T.R. Genetically optimized prediction of remaining useful life. *Sustain. Comput. Inform. Syst.* **2021**, *31*, 100565.
- da Costa, P.R.D.O.; Akcay, A.; Zhang, Y.; Kaymak, U. Attention and long short-term memory network for remaining useful lifetime predictions of turbofan engine degradation. *Int. J. Progn. Health Manag.* **2019**, *10*, 034.
- Ghorbani, S.; Salahshoor, K. Estimating Remaining Useful Life of Turbofan Engine Using Data-Level Fusion and Feature-Level Fusion. *J. Fail. Anal. Prev.* **2020**, *20*, 323–332. [CrossRef]
- Sun, H.; Guo, Y.; Zhao, W. Fault detection for aircraft turbofan engine using a modified moving window KPCA. *IEEE Access* **2020**, *8*, 166541–166552. [CrossRef]
- Elasha, F.; Shanbr, S.; Li, X.; Mba, D. Prognosis of a wind turbine gearbox bearing using supervised machine learning. *Sensors* **2019**, *19*, 3092. [CrossRef]
- Hong, S.; Zhou, Z.; Zio, E.; Hong, K. Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method. *Digit. Signal. Process.* **2014**, *27*, 159–166. [CrossRef]
- Yang, B.; Liu, R.; Zio, E. Remaining useful life prediction based on a double-convolutional neural network architecture. *IEEE Trans. Ind. Electron.* **2019**, *66*, 9521–9530. [CrossRef]
- Ren, L.; Sun, Y.; Wang, H.; Zhang, L. Prediction of bearing remaining useful life with deep convolution neural network. *IEEE Access* **2018**, *6*, 13041–13049. [CrossRef]
- Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 1–74. [CrossRef] [PubMed]

25. Hou, G.; Xu, S.; Zhou, N.; Yang, L.; Fu, Q. Remaining useful life estimation using deep convolutional generative adversarial networks based on an autoencoder scheme. *Comput. Intell. Neurosci.* **2020**, *2020*, 1–14. [[CrossRef](#)] [[PubMed](#)]
26. Chen, Z.; Wu, M.; Zhao, R.; Guretno, F.; Yan, R.; Li, X. Machine remaining useful life prediction via an attention-based deep learning approach. *IEEE Trans. Ind. Electron.* **2020**, *68*, 2521–2531. [[CrossRef](#)]
27. Song, F.; Ai, Z.; Zhang, H.; You, I.; Li, S. Smart Collaborative Balancing for Dependable Network Components in Cyber-Physical Systems. *IEEE Trans. Ind. Inform.* **2020**, *17*, 6916–6924. [[CrossRef](#)]
28. Hoepfner, D.W.; Krupp, W.E. Prediction of component life by application of fatigue crack growth knowledge. *Eng. Fract. Mech.* **1974**, *6*, 47–70. [[CrossRef](#)]
29. Jiang, Y.Y.; Zeng, W.W.; Shen, J.J.; Chu, J. Prediction of remaining useful life of lithium-ion battery based on convex optimization life parameter degradation mechanism model. *Proc. CSU EPSA* **2019**, *31*, 23–28.
30. Gao, T.; Li, Y.; Huang, X.; Wang, C. Data-Driven Method for Predicting Remaining Useful Life of Bearing Based on Bayesian Theory. *Sensors* **2021**, *21*, 182. [[CrossRef](#)] [[PubMed](#)]
31. Le Son, K.; Fouladirad, M.; Barros, A.; Levrat, E.; Iung, B. Remaining useful life estimation based on stochastic deterioration models: A comparative study. *Reliab. Eng. Syst. Saf.* **2013**, *112*, 165–175. [[CrossRef](#)]
32. Strušnik, D.; Brandl, D.; Schober, H.; Ferčec, J.; Avsec, J. A simulation model of the application of the solar STAF panel heat transfer and noise reduction with and without a transparent plate: A renewable energy review. *Renew. Sustain. Energy Rev.* **2020**, *134*, 110149. [[CrossRef](#)]
33. Yuan, M.; Wu, Y.; Lin, L. Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network. In Proceedings of the 2016 IEEE International Conference on Aircraft Utility Systems (AUS), Beijing, China, 10–12 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 135–140.
34. Zhao, R.; Wang, J.; Yan, R.; Mao, K. Machine health monitoring with LSTM networks. In Proceedings of the 2016 10th International Conference on Sensing Technology (ICST), Beijing, China, 10–12 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6.
35. Liao, L.; Jin, W.; Pavel, R. Enhanced restricted Boltzmann machine with prognosability regularization for prognostics and health assessment. *IEEE Trans. Ind. Electron.* **2016**, *63*, 7076–7083. [[CrossRef](#)]
36. Zhang, C.; Lim, P.; Qin, A.K.; Tan, K.C. Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 2306–2318. [[CrossRef](#)] [[PubMed](#)]
37. Zheng, S.; Ristovski, K.; Farahat, A.; Gupta, C. Long short-term memory network for remaining useful life estimation. In Proceedings of the 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), Dallas, TX, USA, 19–21 June 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 88–95.
38. Zhu, J.; Chen, N.; Peng, W. Estimation of bearing remaining useful life based on multiscale convolutional neural network. *IEEE Trans. Ind. Electron.* **2018**, *66*, 3208–3216. [[CrossRef](#)]
39. Zhang, X.; Xiao, P.; Yang, Y.; Cheng, Y.; Chen, B.; Gao, D.; Liu, W.; Huang, Z. Remaining useful life estimation using CNN-XGB with extended time window. *IEEE Access* **2019**, *7*, 154386–154397. [[CrossRef](#)]
40. Frederick, D.; de Castro, J.; Litt, J. *User's Guide for the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS)*; NASA/ARL: Hanover, MD, USA, 2007.
41. Xiang, S.; Qin, Y.; Luo, J.; Pu, H.; Tang, B. Multicellular LSTM-based deep learning model for aero-engine remaining useful life prediction. *Reliab. Eng. Syst. Saf.* **2021**, *216*, 107927. [[CrossRef](#)]
42. Duan, Y.; Li, H.; He, M.; Zhao, D. A BiGRU Autoencoder Remaining Useful Life Prediction Scheme With Attention Mechanism and Skip Connection. *IEEE Sens. J.* **2021**, *21*, 10905–10914. [[CrossRef](#)]
43. Heimes, F.O. Recurrent neural networks for remaining useful life estimation. In Proceedings of the 2008 International Conference on Prognostics and Health Management, Denver, CO, USA, 6–9 October 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 1–6.
44. Durairajah, V.; Gobe, S.; Muneer, A. Automatic vision based classification system using DNN and SVM classifiers. In Proceedings of the 2018 3rd International Conference on Control, Robotics and Cybernetics (CRC), Penang, Malaysia, 26–28 September 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 6–14.
45. Choi, J.Y.; Lee, B. Ensemble of deep convolutional neural networks with Gabor face representations for face recognition. *IEEE Trans. Image Process.* **2019**, *29*, 3270–3281. [[CrossRef](#)] [[PubMed](#)]
46. Akbar, N.A.; Darmayanti, I.; Fati, S.M.; Muneer, A. Deep Learning of a Pre-trained Language Model's Joke Classifier Using GPT-2. *J. Hunan Univ. Nat. Sci.* **2021**, *48*.
47. Muneer, A.; Fati, S.M.; Fuddah, S. Smart health monitoring system using IoT based smart fitness mirror. *Telkomnika.* **2020**, *18*, 317–331. [[CrossRef](#)]
48. Muneer, A.; Fati, S.M. Efficient and Automated Herbs Classification Approach Based on Shape and Texture Features using Deep Learning. *IEEE Access* **2020**, *8*, 196747–196764. [[CrossRef](#)]
49. Naseer, S.; Ali, R.F.; Muneer, A.; Fati, S.M. IAmideV-deep: Valine amidation site prediction in proteins using deep learning and pseudo amino acid compositions. *Symmetry* **2021**, *13*, 560. [[CrossRef](#)]
50. Naseer, S.; Ali, R.F.; Fati, S.M.; Muneer, A. iNitroY-Deep: Computational Identification of Nitrotyrosine Sites to Supplement Carcinogenesis Studies Using Deep Learning. *IEEE Access* **2021**, *9*, 73624–73640. [[CrossRef](#)]
51. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Into Imaging* **2018**, *9*, 611–629. [[CrossRef](#)] [[PubMed](#)]

-
52. Naseer, S.; Saleem, Y. Enhanced Network Intrusion Detection using Deep Convolutional Neural Networks. *TIIS* **2018**, *12*, 5159–5178.
 53. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
 54. Saxena, A.; Goebel, K.; Simon, D.; Eklund, N. Damage propagation modeling for aircraft engine run-to-failure simulation. In Proceedings of the 2008 International Conference on Prognostics and Health Management, Denver, CO, USA, 6–9 October 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 1–9.
 55. Peng, Y.; Wang, H.; Wang, J.; Liu, D.; Peng, X. A modified echo state network based remaining useful life estimation approach. In Proceedings of the 2012 IEEE Conference on Prognostics and Health Management, Denver, CO, USA, 18–21 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1–7.
 56. Laredo, D.; Chen, Z.; Schütze, O.; Sun, J.Q. A neural network-evolutionary computational framework for remaining useful life estimation of mechanical systems. *Neural Netw.* **2019**, *116*, 178–187. [[CrossRef](#)] [[PubMed](#)]
 57. Aggarwal, K.; Atan, O.; Farahat, A.K.; Zhang, C.; Ristovski, K.; Gupta, C. Two birds with one network: Unifying failure event prediction and time-to-failure modeling. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1308–1317.