Sustainable Machine Intelligence Journal

SUSTAINABLE MACHINE INTELLIGENCE JOURNAL

Journal Homepage: sciencesforce.com/smij

SCIENCES FORCE
International Academic Publisher

Sustain. Mach. Intell. J. Vol. 8 (2024) 14–34

Paper Type: Original Article

A Data-driven Deep Learning Approach for Remaining Useful Life in the ion mill etching Process

Ahmed Darwish [1,*]

[1] Department of Computer Science, Faculty of Computer and Informatics, Zagazig University, Zagazig, 44519, Egypt; adarwish@fci.zu.edu.eg

Received: 05 Feb 2024 **Revised:** 04 May 2024 **Accepted:** 06 Jun 2024 **Published:** 09 Jun 2024

Abstract

Prognostics and Health Management (PHM) is regarded as an essential element in the scope of intelligent manufacturing. Precise forecasting of the remaining useful life (RUL) of an ion mill is crucial in order to enhance the overall efficiency of the ion mill etching (IME) procedure. This paper proposed a Data-driven Deep Learning (DL) framework that integrates a Temporal Convolution Network (TCN), Long Short-Term Memory (LSTM), and self-attention mechanism to improve the accuracy of RUL prediction in the ion mill etching Process. Initially, sensor input data is divided into two parallel paths - one with TCN blocks for capturing long-range dependencies, and the other with LSTM layers for extracting temporal patterns. The outputs from both paths are then merged and input into an LSTM layer for enhanced learning, followed by a self-attention mechanism to highlight important features then fully connected layer for predicting RUL. The efficacy of this suggested model was assessed through the utilization of the 2018 PHM Data Challenge Dataset and juxtaposed against various Deep Learning models to demonstrate its efficacy. The results from the experiments indicate that ATCN-LSTM serves as a robust option for estimating the RUL in the ion mill etching Process as it outperformed all other models that were compared. The source code is publicly accessible at https://github.com/ion-mill-etching-Process.

Keywords: Remaining Useful Life; Prognostics and Health Management; Long Short-term Memory Network; Temporal Convolutional Network; Attention Mechanism.

1 | Introduction

Prognostics and Health Management (PHM) is considered a vital component within the realm of smart manufacturing [1]. PHM aims to predict the remaining useful life (RUL) of a specific system and formulate the optimal approach for managing its health [2]. PHM is a computational framework that explores the realm of physical knowledge related to the functioning and upkeep of structures, systems, and components [3]. In the realm of semiconductor manufacturing, the presence of sophisticated machinery involves a complex interconnection of various elements, which presents a significant obstacle to predicting system-level performance. In order to mitigate the risk of abrupt equipment breakdown due to the degradation of components, minimize maintenance costs, and uphold competitiveness in the industry, companies are progressively turning to advanced PHM technology to methodically oversee the operational status of production machinery. The prediction of RUL offers insights into the duration until failure and is crucial for

Corresponding Author: adarwish@fci.zu.edu.eg

https://doi.org/10.61356/SMIJ.2024.8288

guiding maintenance choices prior to a breakdown, ultimately diminishing uncertainties [2]. Ion mill etching (IME) is a semiconductor manufacturing method that is both accurate and effective [4]. During this procedure, an unreactive gas is aimed at the wafer by passing through various grids with distinct voltage variances. The wafer is positioned on a revolving platform and can be inclined towards the ion beam. The etching procedure commences upon the opening of the shutter, with the ion beam eliminating material from the wafer [5]. Wafers, which are silicon chips utilized in the production of integrated circuits and photovoltaic items, are undergoing increasingly intricate processing due to the swift advancement of both the integrated circuit and photovoltaic industries [6] Ion etching technology is capable of altering the surface morphology of a product at the micron scale [7]. As a result, the utilization of IME machines has become indispensable in the realm of wafer processing. The configuration of these IME machines can be observed in Figure 1. To avoid overheating, a helium water system known as Flowcool is employed to lower the temperature of the wafer. Nonetheless, this particular system is susceptible to malfunctions that may result in unforeseen periods of inactivity and increased maintenance expenses. Consequently, it is crucial to anticipate and pinpoint potential malfunctions in order to guarantee prompt maintenance and enhance operational effectiveness.

Various research methodologies have been developed by scholars in recent times to predict the RUL of ion mill etching. These strategies are commonly classified into two main groups: physics-based methods and data-driven methods [8-12]. In the approach based on physics, existing scientific knowledge is utilized to develop physical models that illustrate the deterioration of mechanical machinery [13]. To construct a physical prognostic model, the initial stage involves utilizing a set of dynamic ordinary or partial differential equations to elucidate the behaviors of the system [14]. The disparities observed between the real values and the results generated by the behavioral model are referred to as residuals. If a residual exceeds a specified threshold, it signifies the presence of a fault [15]. Sadabadi et al. [16] developed a sophisticated single-particle model that includes enhanced parameters for predicting the RUL of Electric Vehicle (EV) batteries using charging data. This concept provides a fundamental explanation for the deterioration of batteries caused by electrochemical activities taking place inside them. Gai et al. [17] proposed a method to estimate the fatigue lifespan by utilizing contact stress based on the fatigue theoretical design methodology. This approach allows for the calculation of the maximum stress at the point of contact, given the provided radial and axial forces applied to the bearing. This calculation helps in determining the lifetime of the bearing due to contact fatigue by referring to the contact fatigue life curve. Physical models are able to provide an accurate and analytical representation of system behavior due to being developed with a deep comprehension of the system. Additionally, physical models help elucidate the physical mechanisms of failure. The processes of IME entail intricate interplays among various particles and surfaces, with the dynamics being significantly influenced by the unique properties of the materials and the processing parameters. Consequently, creating physics-driven models to characterize the degradation patterns, mechanisms, and failure modes poses a formidable challenge. However, this approach needs to demonstrate more performance in creating physical representations for various complex systems [18]. Furthermore, the process of constructing highly accurate physical models through trial and error to align with experimental findings can be a time-consuming endeavor lasting several years. Data-driven approaches involve the creation of predictive models by utilizing past data collected from sensors installed on equipment and the corresponding measurements. The utilization of a data-driven methodology is adept at achieving generalization without the need for specialized expertise. By employing data-driven techniques, the connection between sensor data and the deterioration of a system can be revealed [19], showcasing a strong ability to generalize and reduce the dependence on empirical knowledge [20]. Data-driven approaches for predicting RUL have the potential to reduce dependence on traditional engineering knowledge and generate accurate predictions through the examination of temporal data patterns. This technique proves valuable in the continuous monitoring of equipment in real-time, mitigating the risk of critical malfunctions, and improving forthcoming manufacturing operations [11].

This research assesses the RUL of intricate systems through the application of data-driven approaches. These methods include both Deep Learning (DL) and Machine Learning (ML) methodologies. ML methodologies have evolved into a robust instrument in multiple facets of our daily routines, empowering computers to assimilate information from data without the need for explicit programming and to consistently enhance their efficiency. The capability of ML to derive insights from data and execute tasks automatically is revolutionizing our lifestyle, professional activities, and technological interactions. As the field develops, we can predict even more substantial impacts on our global society. ML methodologies can exploit vast quantities of sensor data, operational parameters, and historical maintenance records. This methodological data-driven approach enables machine learning algorithms to comprehend intricate connections among various variables that impact the condition and decline of machinery. Extraction of complex features is carried out across time, frequency, and time-frequency domains [21], and the selection of features must also be executed to guarantee the optimal performance of conventional machine learning models. Sigaud et al. [22] utilize artificial neural networks (ANN) to estimate the Remaining Useful Life (RUL) by leveraging models trained on inspection data. While the multilayer perceptron (MLP) with a single hidden layer is theoretically capable of fitting any continuous function, it falls short in addressing sequence learning challenges. This limitation arises from the fact that original MLP networks can solely map from current input to output vectors, lacking the capability to effectively capture changes within sequences [23]. Zhang et al. [24] employed the relevance vector machine (RVM) in combination with the differential evolution (DE) algorithm to forecast the RUL of batteries by utilizing denoised data produced through the wavelet denoising algorithm. Berghout et al. [25] detailed the crucial steps involved in forecasting the remaining operational longevity through the utilization of machine learning techniques. They systematically examined the prospective advantages and challenges that may arise in the future. Mejia et al. [26] conducted wavelet packet decomposition to extract coefficients from the original sensor data, utilizing a combination of Gaussians Hidden Markov Model (HMM) for evaluating the present operational status of the machinery. Furthermore, the model was applied for the estimation of RUL and the corresponding confidence level. Many of these techniques necessitate significant feature engineering to identify important features. Furthermore, the moderate complexity of these models constrains their ability to accurately process data and demonstrate strong generalization skills.

Deep learning is a significant subset in the field of machine learning and has triggered substantial changes across multiple facets of our lives. It integrates artificial neural networks with multiple layers for data processing that simulates the cognitive functions of the human brain. The progress in deep learning methodologies has been remarkable in light of their generality. Feature engineering is not required for deep learning algorithms since they possess the ability to autonomously extract feature representations. DL is distinguished by its more intricate network architecture, enabling the extraction of deteriorating features from historical sensor data of monitored equipment [27]. Consequently, DL exhibits superior efficiency in handling high-dimensional and unorganized data types like images and time series data. Notably, DL models outperform other ML techniques in the realm of predictive maintenance [11]. Convolutional neural networks (CNNs) integrated with additional structures have exhibited enhanced feature extraction capabilities in predicting RUL as a result of their expanded receptive field and efficient computational processes [28]. CNN is employed for extracting local characteristics from the input data and merging these characteristics to produce advanced features through a series of convolutional and pooling layers stacked on top of each other [29]. The effectiveness of CNN has been shown in applications related to predictive maintenance [30-32]. The adjustment required for enhancing predictive maintenance often involves a straightforward process of transitioning the input kernel from a 2-dimensional filter to a 1-dimensional filter for processing the numerical data within a time series dataset. Nonetheless, in order to enhance pattern recognition capabilities, additional filters must be utilized to detect a greater number of distinct patterns [33]. Recurrent neural networks (RNN) examine the temporal relationships within an input sequence at a specific time point, leveraging the knowledge stored in the hidden units pertaining to past observations of the sequence. Catelani et al. [34] integrated the RNN with a filtering-based technique for predicting RUL. The enhancement of the RNN's efficacy is ensured in this methodology by employing Genetic Algorithms (GA). RNNs encounter two potential limitations when employing gradient-based backpropagation for the purpose of modifying the weights of the network. The issue arises as the number of hidden layers (or, in this particular scenario, the loopback) grows, leading neural networks to potentially confront either the challenge of the exploding gradient or the diminishing gradient, particularly in the context of extended-term forecasting [35]. Long short-term memory (LSTM) models incorporate gate control mechanisms to address the limitations of traditional RNNs, specifically in managing memory retention and forgetting processes [36]. Jianjing et al. [37] introduced an LSTM neural network, in which state cells are incorporated to retain long-term memory. Wang et al. [38] introduced a bidirectional LSTM model incorporating an attention mechanism to carry out RUL regression for lithium batteries. Gated Recurrent Unit (GRU) bears a resemblance to LSTM in its utilization of gating mechanisms to manage long-term dependencies; however, the primary distinction between them lies in their respective gates [39]. Chen et al. [40] introduced a methodology consisting of a two-part procedure that incorporates Kernel Principal Component Analysis (KPCA) for feature extraction, followed by the utilization of a GRU-based architecture for the prediction of RUL.

CNNs are frequently employed for prediction due to their efficient feature extraction capabilities; however, they encounter challenges when dealing with long-term dependencies [41]. In response, TCNs were developed as an enhanced version of CNNs with a specific emphasis on sequence modeling. TCNs utilize 1D convolutions in contrast to 2D convolutions, as the latter extract features from multiple sensors simultaneously, leading to increased noise with each feature extraction [42]. Wang et al. [41] introduced a TCN which incorporates two distinct convolutional techniques aimed at enhancing the capability to capture local relationships and enhancing accuracy: causal convolutions, which enable the examination of preceding linear steps in the sequence, and dilated convolutions, which allow for the observation of a broader segment of the entire sequence. Transformers employ a multi-head self-attention mechanism to capture long-range dependencies within a sequence irrespective of the distance between elements. This strategy enhances the model's ability to handle longer sequences and eliminates the need for employing recurrence or convolution techniques [43]. Ma et al. [44] developed a modified version of the G-Transformer model framework which incorporates the encoder component from conventional Transformer models. This adaptation is specifically designed for sampling and extracting features in the context of natural language processing for PM.

Although there have been notable recent improvements in data-driven prediction techniques, there are two distinct and challenging obstacles when it comes to constructing models employing condition monitoring (CM) data from the IME process. 1) The distribution of valuable information in the IME procedures is unequal. The records collected from various sensors provide deterioration information related to different physical characteristics of the system. The physical features have a considerable impact on the behavior of different failure types. As a result, the data collected in the IME process shows an unequal distribution of effective information across time. 2) The deterioration patterns that are a part of this process are still not completely understood. Within the framework of IME, certain errors can manifest suddenly, while other flaws might develop gradually. Moreover, the latency duration of various sensors in identifying symptoms might differ and is not well-defined, resulting in an inadequate representation of the deterioration characteristics noted during this procedure.

To address these obstacles, we introduce a novel data-driven deep learning approach aimed at predicting the RUL within the IME process. A novel technique referred to as the integrated temporal convolution with long short-term memory (ATCN-LSTM) network is suggested to extract spatial characteristics and capture temporal dependencies from sensor data for the prediction of RUL. Temporal convolutional networks (TCNs) employ one-dimensional causal filters and dilated convolutions for feature extraction from raw time series data, demonstrating significant memory capabilities for long time series. The LSTM component is employed to understand temporal dependencies from TCN output and extract temporal information from input data. Additionally, an attention layer is introduced to address the issue of inadequate learning in LSTM and RNN models when dealing with lengthy sequences. Finally, fully connected layers are utilized to link LSTM representations to the RUL label. Our experimental validation demonstrates that the suggested approach is capable of efficiently extracting pertinent features from data, identifying degradation patterns within the IME process, and generating reliable RUL predictions. The significant contributions of this study can be outlined as follows:

1) A novel deep learning architecture utilizes a multi-branch strategy to understand temporal relationships in time series data. This is achieved through the integration of a temporal convolution network (TCN), long short-term memory (LSTM), and self-attention mechanism, which efficiently captures deterioration patterns in the IME process.

2) The validity of the model was confirmed through the utilization of the dataset from the 2018 PHM data challenge [6], demonstrating its superior performance in predicting RUL compared to other models.

The subsequent sections of this paper are structured as follows. Section 2 elaborates on the proposed methodology. This is followed by the exposition of Experimental settings in Section 3. Section 4 discusses the Results and discussion. Lastly, Section 5 houses the conclusions of this study.

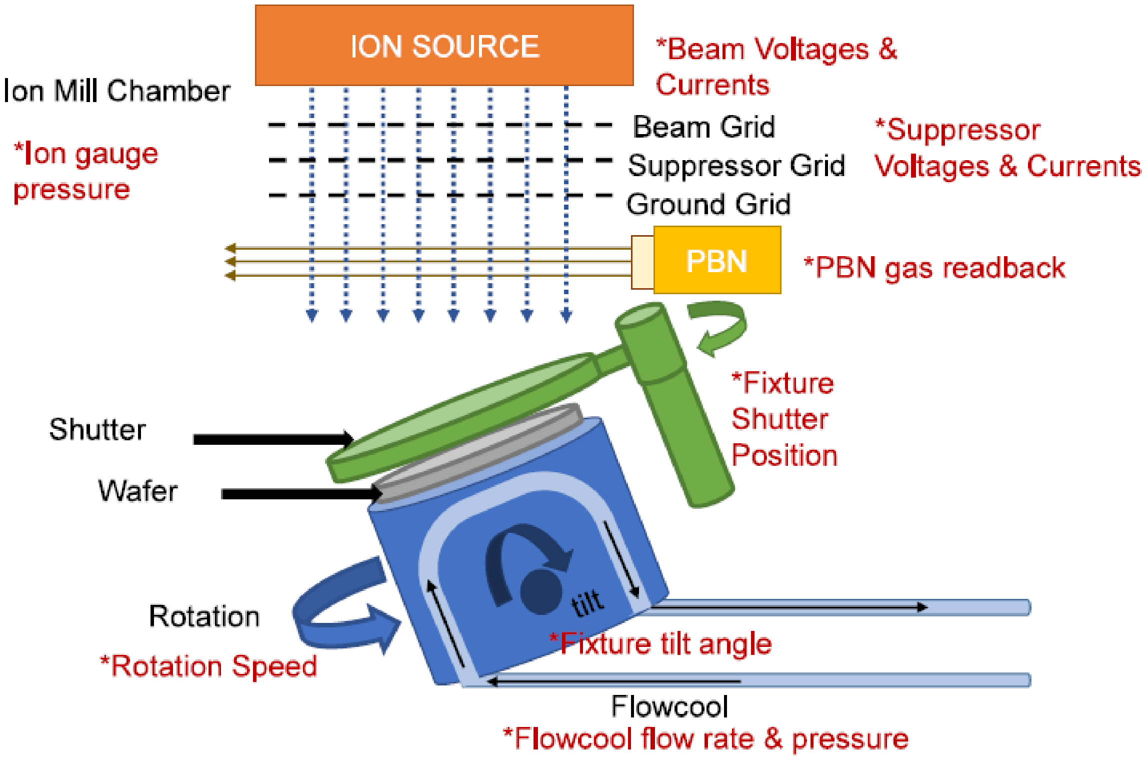

Figure 1. An ion mill etching apparatus and its associated parameters **[15].**

2 | The Proposed DL Model

2.1 | TCN-LSTM Components

2.1.1 | Temporal Convolution Network (TCN)

TCNs have a greater capacity to capture distant interdependencies within temporal data sequences compared to traditional RNNs such as LSTM or GRU models. The superior performance of TCNs stems from their utilization of Causal convolutions, which augment the effective receptive field exponentially as the depth increases. The TCN is a neural network design specifically tailored for processing sequential data, which incorporates causal convolution inspired by the Wavenet model [45]. This unique feature enables TCNs to discern patterns across lengthier sequences without imposing a substantial rise in computational overhead. TCNs represent an advancement in CNNs that specifically emphasize sequence modeling. TCNs are designed to employ 1D convolutions in place of 2D convolutions, as the latter extract characteristics from multiple sensors simultaneously, thereby introducing extra noise with each feature extraction [42]. 1D-temporal convolutions exclusively capture the characteristics pertinent to the temporal relationships among separate sensors, which are then utilized in forming extensive short-term temporal dependencies [42]. TCNs distinguish themselves from their standard CNN equivalent through two additional features: firstly, by incorporating convolutions to avoid temporal leakage between sets, and secondly, by adopting the RNN framework which permits input sequences of varying lengths and produces output sequences of identical lengths [46]. A causal relationship within the convolutional layers is employed to encode past information in time series data. Figure 2 and Figure 3 illustrate the general convolution operation and the causal convolution operation, respectively. The distinction between general convolution and causal convolution lies in the fact that causal convolution exclusively takes into account past data.

To enhance the long-term memory capacity of TCN, dilations are incorporated along with causal convolutions to support autoregressive forecasting. Additionally, a residual connection is introduced to facilitate the stacking of multiple layers while preventing model overfitting. The employment of dilation-causal convolution operation in TCN replaces the conventional use of general convolution. As illustrated in Figure 4, the dilation causal convolution operation is depicted. When compared to causal convolution with an equivalent convolution kernel size, dilated convolution offers a larger receptive field. To ensure the stability of a deep neural network, a residual block with a shortcut connection is integrated into the TCN [47]. Spatial dropout and weight normalization are commonly applied following each dilated-causal convolution within the residual block of TCN for regularization purposes, as depicted in Figure 5. Furthermore, the input X of the residual block is combined with the output Y using a 1x1 convolution layer to guarantee that the elementwise addition maintains the same width in TCN.

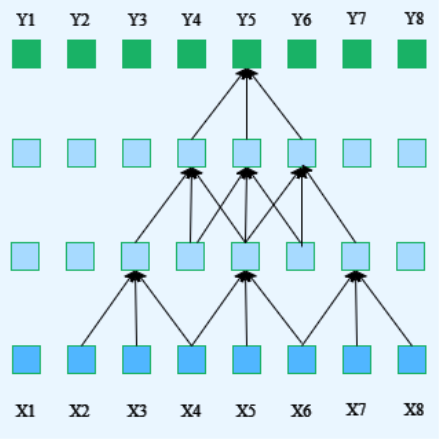

Figure 2. General Convolution.

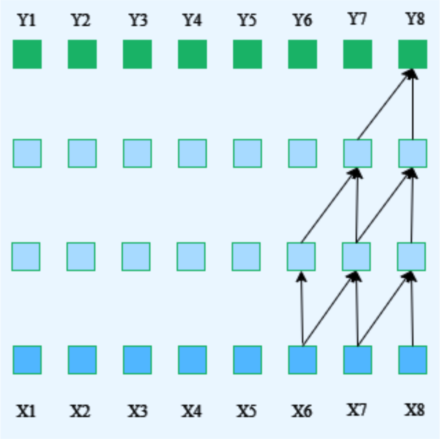

Figure 3. Causal Convolution.

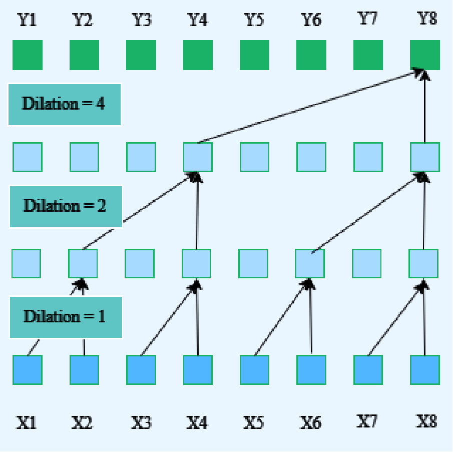

Figure 4. Dilation-causal convolution operation.

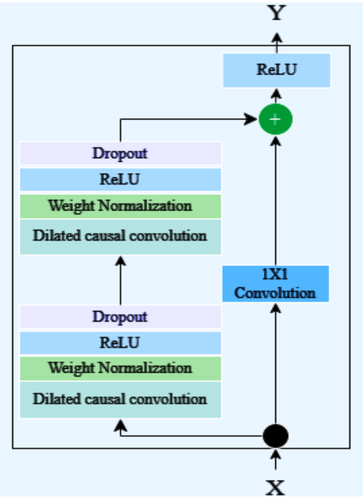

Figure 5. Residual block of TCN.

2.1.2 | Long Short-Term Memory (LSTM)

LSTM introduces a novel variant of the RNN architecture, specifically designed to address the challenge of vanishing gradient to effectively manage the prediction of long-term dependencies [36]. This particular model introduces a complex memory unit that allows for the retention of information over lengthy sequences, making it well-suited for tasks requiring long-term dependencies. The LSTM cell architecture is shown in Figure 6. The LSTM cell consists of the forget gate (f_t), input gate (i_t), and output gate (o_t) are the three essential elements that jointly control the information flow and manage the interactions within the network. The forget gate f_t plays a crucial role in deciding the information to be omitted from the prior cell state. Through the evaluation of the present input and the preceding hidden state, the forget gate produces an output value that spans from 0 (representing total forgetfulness) to 1 (representing complete retention). Its formulation is formally expressed in equation 1.

$$f_t = \sigma\left(W_f\, x_t + U_f\, h_{t-1} + b_f\right) \tag{1}$$

Where the symbol σ indicates the sigmoid function, t signifies the time step, x_t denotes the input feature at time t, h_{t-1} signifies the output hidden state from the previous time sample, the parameters W_f, U_f, b_f are optimized during the training process. The input gate i_t plays a crucial role in the selection of new information to be stored in the cell state. Through the utilization of the current input and the previous hidden state, it produces an output value within the range of 0 to 1. Additionally, it creates a new candidate value that is aimed at being integrated into the cell state. This process is formally described by equation 2.

$$i_t = \sigma(W_i\, x_t + U_i\, h_{t-1} + b_i) \tag{2}$$

The parameters W_i, U_i, b_i are optimized during the training process. The output gate o_t controls the determination of which information will be conveyed as the concealed state of the present LSTM cell, utilizing input from the current time step and the preceding hidden state. This gate produces a numeric value between 0 and 1, and its calculation is specified by equation 3.

$$o_t = \sigma(W_o\, x_t + U_o\, h_{t-1} + b_o) \tag{3}$$

The parameters W_o, U_o, b_o are optimized during the training process. The candidate value c'_t represents new information that could potentially be added to the cell state during the current time step (t). It is produced by the input gate, considering both the current input and the previous hidden state. The calculation of this value is formally described by equation 4.

$$c'_t = tanh(W_a\, x_t + U_a\, h_{t-1} + b_a) \tag{4}$$

The parameters W_a, U_a, b_a are optimized during the training process. The c_t value, which denotes the unit state at time t, is determined mathematically by equation 5. Following this, the h_t value, representing the hidden state at time t, is then calculated using mathematical equation 6.

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t \tag{5}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{6}$$

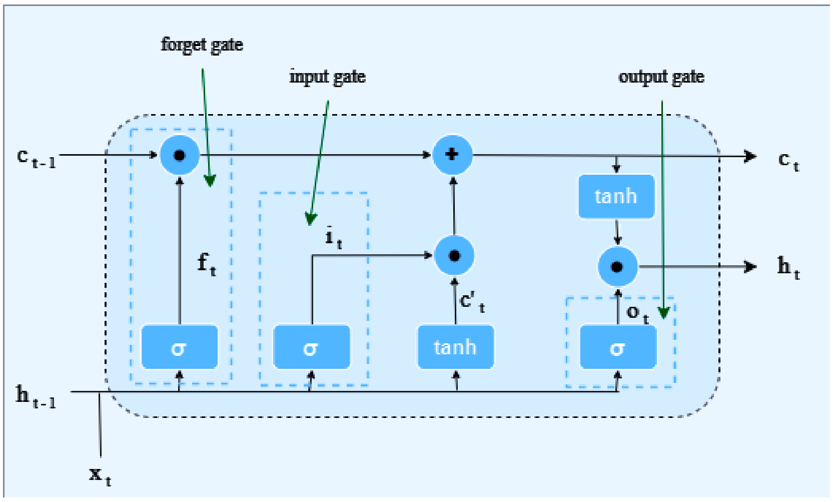

Figure 6. LSTM cell architecture.

2.1.3 | Self-attention Mechanism

Traditional time series models commonly face challenges in capturing intricate relationships in the data, particularly long-term dependencies. The utilization of self-attention enables the identification of dependencies across various time steps, facilitating the effective learning and representation of temporal relationships. In contrast to sequential data processing approaches, self-attention empowers each time step to consider all other time steps in the sequence. This empowers the model to determine the most pertinent past values for future predictions. Time series data frequently showcase extended dependencies, where the value at a specific time step is influenced by distant historical observations. Self-attention mechanisms, notably those incorporating dilated convolutions or multi-head attention, are adept at capturing such extended dependencies, thereby enabling the model to make forecasts based on a broader contextual understanding. The utilization of the attention mechanism allows for the integration of data gathered from every time step of the concealed state through the allocation of weights to the most relevant information in the sequence. The operation of the attention mechanism involves three vectors: Queries (Q), Keys (K), and Values (V), which stem from the transformation of the input X into distinct feature spaces. The query vector (Q) represents the central point of the model at every time step. The key vector (K) encapsulates the available information at each time step in the sequence, while the value vector (V) holds the factual data content for each time step. The computation of the vectors Q, K, and V is performed in the following manner:

$$Q = X \times W_q \tag{7}$$

$$K = X \times W_k \tag{8}$$

$$V = X \times W_v \tag{9}$$

where W_q, W_k, and W_v are trainable parameters, and X indicates the input tensor. Upon obtaining Q and K, a score is calculated by performing a scaled dot product between the query Q and each K across the sequence, divided by the $\sqrt{d_k}$ after the inner product, where $\sqrt{d_k}$ signifies the dimension of K. as shown in Equation 10. Following this, the scores are subjected to a softmax function, which converts them into a probability distribution referred to as the attention weights (A) in Equation 11. These weights indicate the relative importance of each time step in the sequence for the current prediction (determined by the query vector). Subsequently, the attention weights (A) are employed to allocate weight to the corresponding value vectors (V) as per Equation 12. This procedure results in the creation of a context vector that encapsulates the most relevant information from the entire sequence based on the current focus (query).

$$attention_{scores} = \frac{Q \ . \ K^T}{\sqrt{d_k}} \tag{10}$$

$$A = softmax(attention_{scores}) \tag{11}$$

$$context_{vector} = A \ . \ V \tag{12}$$

2.2 | TCN-LSTM Model

The prediction of the remaining useful life in the ion mill etching Process is commonly recognized as a supervised regression assignment where information from several sensors is employed to train and evaluate different DL models. This study proposed a novel data-driven DL model named ATCN-LSTM, which incorporates TCN, LSTM, and a self-attention mechanism to predict the RUL in the ion mill etching Process. The ATCN-LSTM model is designed to process input sensor data through two distinct parallel paths. The initial path comprises two TCN blocks that aim to capture long-range dependencies present in temporal data sequences. Conversely, the second parallel path involves two LSTM layers that specialize in extracting temporal patterns from the input data. Subsequently, the outputs from both paths are merged and fed into an LSTM layer to enhance the learning of temporal patterns from the combined features, followed by a self-attention mechanism to emphasize relevant features while disregarding less significant ones. The outcome of this intricate process is then passed through a fully connected layer to make accurate predictions. Additionally, the pseudocode for the proposed model is detailed in Algorithm 1, while the architectural flowchart is visually represented in Figure 7 to provide a comprehensive overview of its structure and functionality.

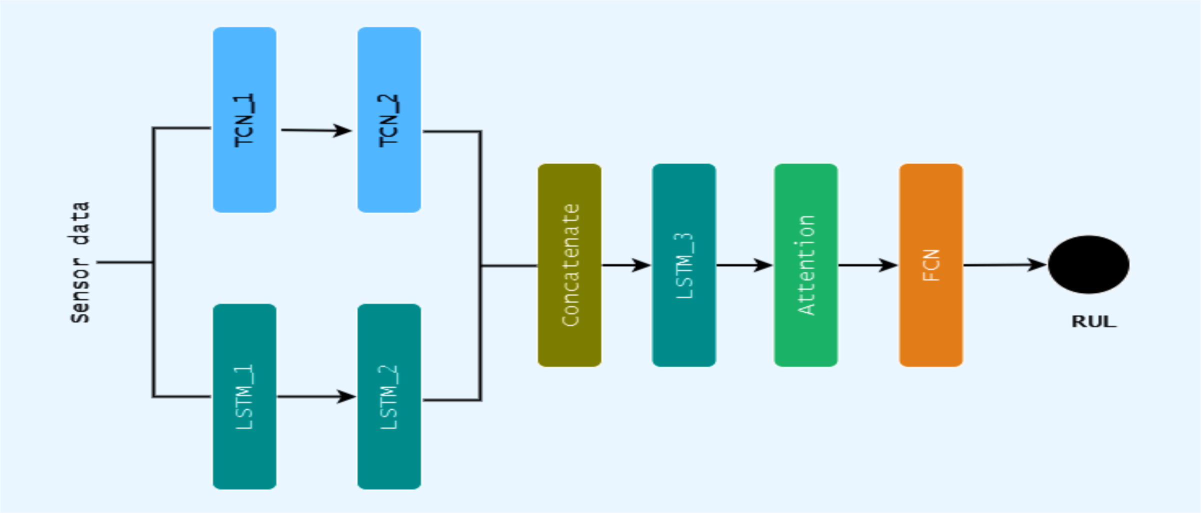

Figure 7. Flowchart of the proposed ATCN-LSTM.

Algorithm 1 Pseudo-code of ATCN-LSTM

```
Input: Input data ($D$), batch size ($Bs$), maximum epoch ($T$), and learning rate ($lr$)
Output: loss ($MSE$), $RMSE$
Conducting the preprocessing step
/* Create the proposed ATCN-LSTM model */
Input: Construct an input layer to receive the input data
/* Feature extraction and temporal learning based on the TCN and LSTM */
/* First Parallel Path*/
P1: Create TCN block with [1, 2, 4] dilation, 32 kernels, 4 kernel size, $ReLU$ activation function, and a dropout rate of 0.4 to take the data from the input layer.
P1: Add TCN block with [1, 2] dilation, 16 kernels, 8 kernel size, $ReLU$ activation function, and dropout rate of 0.4 to P1.
/* Second Parallel Path*/
P2: Create an LSTM layer with 128 cells and a $Tanh$ activation function to take the data from the input layer.
P2: Add LSTM layer with 128 cells and $Tanh$ activation function to P2.
/* Concatenation stage */
x: Concatenate ([P1, P2])
x: Add an LSTM layer with 64 units and $Tanh$ activation function to x.
x: Add a Self-attention mechanism to x.
/* Prediction Block */
x: Add a dense layer with 16 nodes and $ReLU$ activation functions to x.
x: Add a dense layer with 8 nodes and $ReLU$ activation functions to x.
x: Add a dense layer with 1 node to x.
/* Optimization process */
$N = Size(D)/Bs$ /* Estimate the number of batches */
$t = 0$, Current epoch
while $t < T$
    $i = 0$, the current batch size.
    while $i < N$
        Compute the Score function using the $ith$ batch.
        Update the weights based on the Adam to optimize the MSE function.
        $i = i + 1$
    end while.
    $t = t + 1$
end while
```

As delineated in Algorithm 1, the suggested model undertakes the reception of the input data for preprocessing purposes to eliminate various issues, such as outliers and predominant features that may adversely affect the model's performance. Following this stage, the input layer processes this data and directs it toward two distinct paths concurrently: the initial path encompasses two TCN blocks designed to capture intricate long-range dependencies inherent in temporal data sequences. Each TCN block is intricately constructed with dilations and causal convolutions to facilitate autoregressive forecasting. Moreover, a residual connection is seamlessly integrated to ease the incorporation of multiple layers while concurrently preventing the occurrence of model overfitting. The residual block is formulated with dilated causal convolution intertwined with weight normalization, Rectified Linear Unit (ReLU) activation function, and a dropout layer strategically incorporated to mitigate the number of trainable parameters, thereby mitigating the overfitting concern and enhancing the overall generalization capacity of the proposed model. Additionally, the input variable X of the residual block is melded with the output variable Y through the utilization of a 1x1 convolution layer, ensuring that the elementwise addition sustains uniform width within the TCN structure. The first TCN block is characterized by dilations of [1, 2, 4], 32 kernels, a kernel size of 4, and a dropout rate set at 0.4. Conversely, the second TCN block features dilations of [1, 2], 16 kernels, a kernel size of 8, and an equivalent dropout rate of 0.4. The alternative path is comprised of two LSTM layers equipped with 128 neurons each and a hyperbolic tangent (Tanh) activation function tailored to assimilate temporal intricacies embedded within the input data sequence.

The concatenated results from the two paths are passed through a concatenation layer before being directed to an additional LSTM layer equipped with 64 units and the Tanh activation function. Subsequently, a self-attention mechanism is applied to highlight important features while downplaying less crucial ones, enhancing the overall understanding of the data. To conclude the process, the outcome of this particular layer is inputted into three fully connected (FC) layers: the initial two layers containing 16 and 8 neurons, respectively, are dedicated to providing a more refined representation of the input sequences, while the third layer, consisting of 1 neuron, is responsible for predicting the remaining relevant aspects in the ion mill etching procedure.

3 | Experimental Settings

3.1 | 2018 PHM Data Challenge Dataset

The 2018 PHM Data Challenge Dataset represents a real-world dataset that investigates the fault behavior of multiple ion mill etching tools utilized in an IME process [6]. This dataset comprises twenty tools equipped with a variety of sensors for process monitoring, as depicted in Figure 1. Ions are produced from an ion source and propelled by an electric field across a series grid at specific voltages. The ion beam impacts the surface of the wafer to eliminate the material. The wafer is positioned on a rotating fixture and inclined at various angles towards the incoming ion beam. A particle beam neutralizer system is utilized to regulate the quantity and shape of the ion beam at the wafer surface. The treated wafers are cooled using a cooling gas flow (helium gas) and water system. The dataset comprises 24 feature variables that are captured at 4-second intervals, encompassing 5 categorical features (such as wafer ID, tool ID, and recipe) and 19 numerical features (including voltage, current, pressure, and flow rate). The details of the dataset are listed in Table 1. The various run-to-failure cycles within each tool are consolidated based on the type of failure. Effective prediction of potential failures enables proactive scheduling of equipment maintenance to prevent unexpected downtime and uphold overall equipment efficiency in the etching process. The dataset also provides information on the timing of three types of failures: Flowcool Pressure Dropped Below Limit (F1), Flowcool Pressure Too High Check Flowcool Pump (F2), and Flowcool leak (F3) [6]. Considering the variations in the operational and consumptive states of the equipment, it is essential to develop individual models for each device. This study focuses on predicting the RUL for the 01_M02 equipment as a case in point.

Table 1. 2018 PHM Data Challenge Dataset Sensors description.

ID	Name	Description	Type
S1	Time	time	Numerical
S2	Tool	tool id	Categorical
S3	Stage	processing stage of the wafer	Categorical
S4	Lot	wafer id	Categorical
S5	Runnum	number of times the tool has been run	Numerical
S6	Recipe	describes tool settings used to process wafer	Categorical
S7	Recipe_step	the process step of a recipe	Categorical
S8	IONGAUGEPRESSURE	pressure reading for the main process chamber when under vacuum	Numerical
S9	ETCHBEAMVOLTAGE	the voltage potential applied to the beam plate of the grid assembly	Numerical
S10	ETCHBEAMCURRENT	ion current impacting the beam grid determining the amount of ions accelerated through the grid assembly to the wafer	Numerical
S11	ETCHSUPPRESSORVOLTAGE	voltage potential applied to the suppressor plate of the grid assembly	Numerical
S12	ETCHSUPPRESSORCURRENT	ion current impacting the suppressor grid plate	Numerical
S13	FLOWCOOLFLOWRATE	rate of flow of helium through the flow cool circuit, controlled by the mass flow controller	Numerical

S14	FLOWCOOLPRESSURE	resulting in helium pressure in the flow cool circuit	Numerical
S15	ETCHGASCHANNEL1READBACK	rate of flow of argon into the source assembly in the vacuum chamber	Numerical
S16	ETCHPBNGASREADBACK	rate of flow of argon into the PBN assembly in the chamber	Numerical
S17	FIXTURETILTANGLE	wafer tilt angle setting	Numerical
S18	ROTATIONSPEED	wafer rotation speed setting	Numerical
S19	ACTUALROTATIONANGLE	measure the wafer rotation angle	Numerical
S20	FIXTURESHUTTERPOSITION	open/close shutter setting for wafer shielding	Numerical
S21	ETCHSOURCEUSAGE	counter of use for the grid assembly consumable	Numerical
S22	ETCHAUXSOURCETIMER	counter of the use for the chamber shields consumable	Numerical
S23	ETCHAUX2SOURCETIMER	counter of the use for the chamber shields consumable	Numerical
S24	ACTUALSTEPDURATION	measured time duration for a particular step	Numerical

3.2 | Data Preprocessing

During the data preprocessing phase, the initial step entails eliminating samples containing missing values or outliers. The operational status of the IME machine greatly depends on the position of the shutter [15]. The Fixture Shutter Position parameter is employed to indicate the status of the shutter. A value of '0' for Fixture Shutter Position corresponds to minimum values for Flowcool Pressure and Ion Gauge Pressure. As the shutter position transitions from '0' to '1', both pressures experience a rapid increase. Once the designated pressure levels are achieved, the Fixture Shutter Position shifts to '1', signifying the opening of the shutter and the commencement of the flowcool subsystem. Upon shutter closure, the Fixture Shutter Position reverts to '0' and the pressure rapidly decreases. Hence, data where the Fixture Shutter Position parameter is '1' is chosen to denote the operational state of the ion mill. The average duration of the run-to-fault series across three fault modes is reported as 2.6×10^5, 3.8×10^5, and 7.5×10^4, it is logical to categorize data with RUL less than 500 (approximately half an hour) as fault data, while those with RUL exceeding 5000 (around 5 hours) are deemed normal. The various run-to-failure cycles in each tool are consolidated based on the type of failure. A total of 112 cycles are allocated for F1, 42 cycles for F2, and 34 cycles for F3. This study exclusively utilizes sensor features for model training and assessment. The data about each fault type are divided into training (70%), validation (20%), and testing (10%) datasets in a randomized manner. Due to the diverse data scales of sensor values recorded for different recipes, which may negatively impact model training, all sensor values for each recipe are standardized within the range [0, 1] through min-max normalization using Equation 13. The limited duration of the subsequent may lead to a lack of adequate information. To capture the characteristics of the time series, each subsequence has a length of 500 and a sliding window size of 1.

$$x'_{i,j} = \frac{x_{i,j} - x_{j\,min}}{x_{j\,max} - x_{j\,min}} \tag{13}$$

Where $x_{i,j}$ donates the value of the ith sample, and jth feature, $x_{j\,min}$, $x_{j\,max}$ donates the minimum, and maximum values in the jth feature, respectively.

Table 2. Experimental Analysis of the Influence of hyperparameter on prediction results.

		value	F1	F2	F3
TCN_1	dilation order	[1, 2]	677.82	612.64	651.81
		[1, 2, 4]	**597.35**	**405.72**	**533.45**
		[1, 2, 4, 8]	836.71	932.27	799.39
	Number of kernels	8	1324.81	974.33	1250.28
		16	841.25	636.42	815.77
		32	**597.35**	**405.72**	**533.45**
		64	764.76	621.91	730.31
	Kernel size	2	705. 41	599.77	771.73
		4	**597.35**	**405.72**	**533.45**
		8	688.44	553.55	699.76
TCN_2	dilation order	**[1, 2]**	**597.35**	**405.72**	**533.45**
		[1, 2, 4]	699.11	521.34	674.27
		[1, 2, 4, 8]	914.17	1014.35	954.61
	Number of kernels	8	788.51	804.66	732.16
		16	**597.35**	**405.72**	**533.45**
		32	685.11	671.62	658.24
		64	841.96	845.75	801.40
	Kernel size	2	750.14	697.37	791.17
		4	625.95	623.49	640.87
		8	**597.35**	**405.72**	**533.45**
LSTM_1	Number of cells	64	785.34	615.87	764.96
		128	**597.35**	**405.72**	**533.45**
		256	685.18	503.05	621.22
LSTM_2	Number of cells	64	733.12	868.68	794.66
		128	**597.35**	**405.72**	**533.45**
		256	699.27	601.99	615.94
LSTM_3	Number of cells	**64**	**597.35**	**405.72**	**533.45**
		128	752.49	601.91	721.26
		256	942.11	775.17	916.37
Dense_1	Number of nodes	8	701.47	535.67	699.17
		16	**597.35**	**405.72**	**533.45**
		32	654.78	495.18	604.53
Dense_2	Number of nodes	**8**	**597.35**	**405.72**	**533.45**
		16	684.87	554.29	666.18
		32	736.64	604.51	699.59
Learning rate	Learning rate vale	0.0001	964.19	912.88	906.33
		0.001	**597.35**	**405.72**	**533.45**
		0.002	655.99	601.76	605.94
		0.01	1130.82	1164.23	1210.64

3.3 |Evaluation Metrics

in this study, the Adam optimization algorithm [48] and mean square error (MSE) loss, calculated using equation 14, are utilized to optimize the network parameters. The introduction of root mean square error (RMSE) is employed as a measure to assess the suggested model by comparing RMSE values between actual and predicted labels of each instance in the dataset. The RMSE value is computed mathematically as outlined in equation 15, where N denotes the total number of samples, and y_i and y'_i represent the actual and predicted labels of the ith sample, respectively. Both MSE and RMSE need to be minimized to improve the accuracy in predicting the RUL in the Ion Mill Etching Process.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - y'_i\right)^2 \tag{14}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - y'_i\right)^2} \tag{15}$$

4 | Results and Discussion

4.1 | Hyper Parameter Tuning

The ATCN-LSTM model proposed in this study encompasses numerous hyper-parameters such as the dilation order, number of kernels, size of kernels in each TCN block, number of cells in each LSTM layer, as well as the learning rate. These parameters need to be accurately defined to enhance the model's effectiveness and lower the RMSE. Therefore, a sequence of experiments has been conducted in the course of this study, examining various configurations for each variable to determine the most effective values that result in a significant improvement in the performance of the model, as illustrated in Table 2. The influence of those hyper-parameters on the performance of the ATCN-LSTM method is displayed from Figure 8 to Figure 19. For example, the efficiency of the model is notably affected by the quantity of cells present in every LSTM layer. As a result, several trials were conducted to determine the optimal number of cells for each LSTM layer, spanning from 64, 128, to 256. The influence of the number of cells can be observed in Figures 14, 15, and 16. The learning rate plays a vital role as a hyperparameter when training deep learning models. It dictates the magnitude of the increments made during the optimization process to modify the model's parameters towards minimizing the loss function. So multiple experiments were conducted to determine the most suitable learning rate within the range of 0.0001, 0.001, 0.002, and 0.01, with findings indicating that the optimal learning rate is 0.001. The impact of learning rate values is depicted in Figure 19.

Table 3. The ATCN-LSTM hyperparameters.

Parameter	value
TCN_1 dilation order	[1, 2, 4]
TCN_1 number of kernels	32
TCN_1 kernel size	4
TCN_2 dilation order	[1, 2]
TCN_2 number of kernels	16
TCN_2 kernel size	8
LSTM_1 number of cells	128
LSTM_2 number of cells	128
LSTM_3 number of cells	64
Dropout rate	0.4
Learning rate	0.001
Max no. of epoch	1000
Loss	MSE
Optimizer	Adam

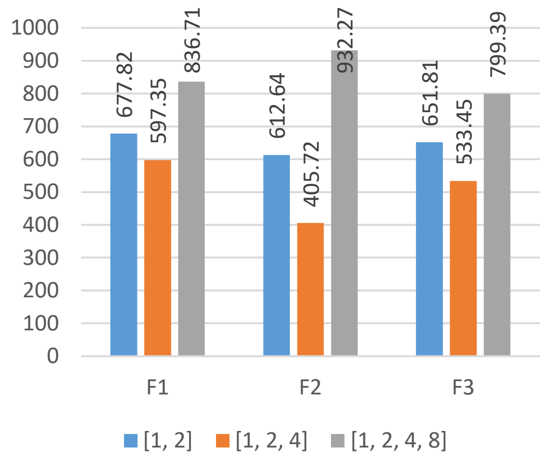

Figure 8. The influence of dilation order of TCN_1 through experiments.

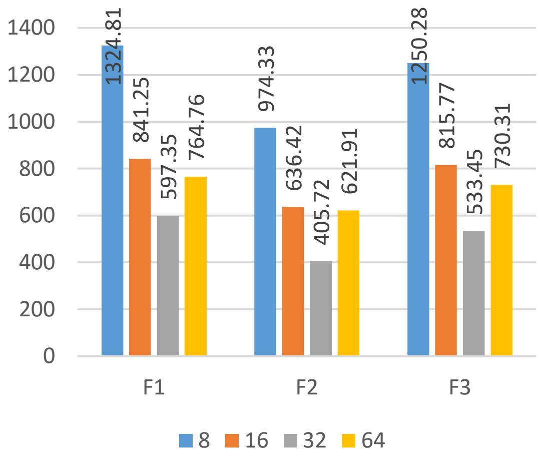

Figure 9. The influence of the number of kernels of TCN_1 through experiments.

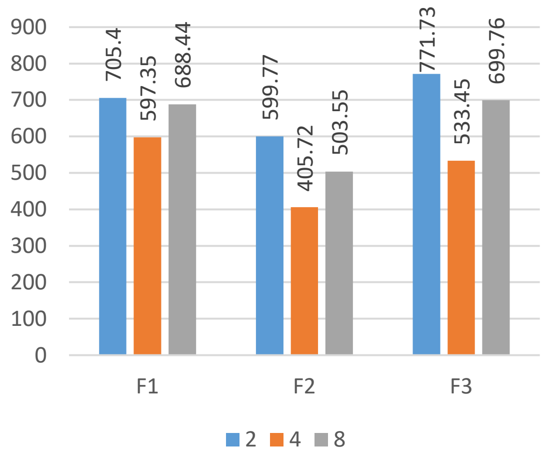

Figure 10. The influence of kernel size of TCN_1 through experiments.

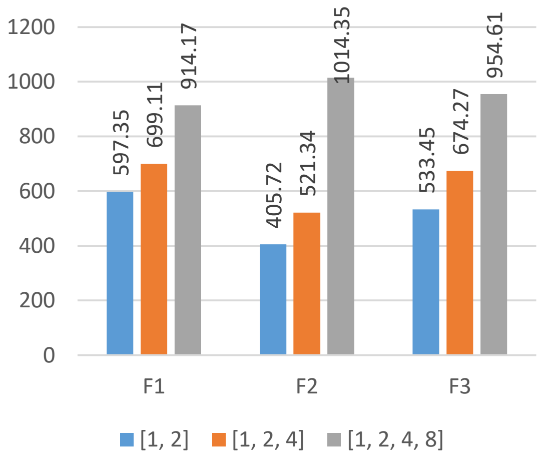

Figure 11. The influence of dilation order of TCN_2 through experiments.

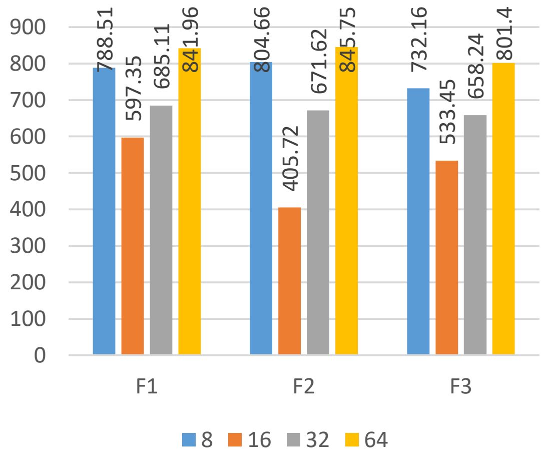

Figure 12. The influence of the number of kernels of TCN_2 through experiments.

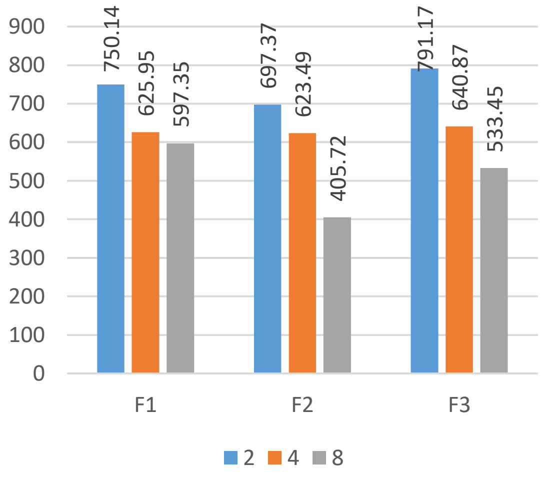

Figure 13. The influence of kernel size of TCN_2 through experiments.

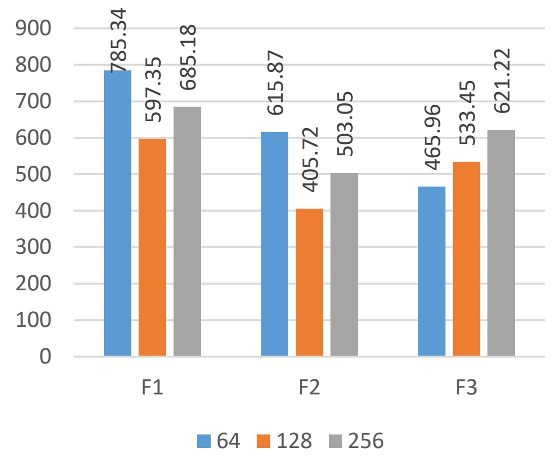

Figure 14. The influence of the number of cells of LSTM_1 through experiments.

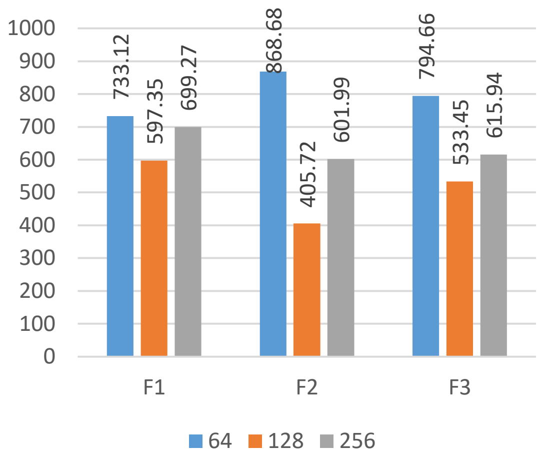

Figure 15. The influence of the number of cells of LSTM_2 through experiments.

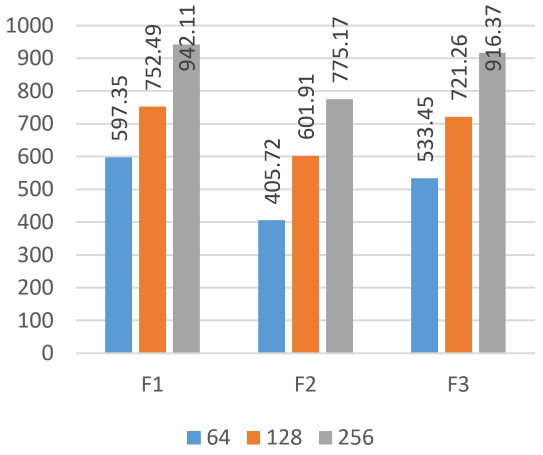

Figure 16. The influence of the number of cells of LSTM_3 through experiments.

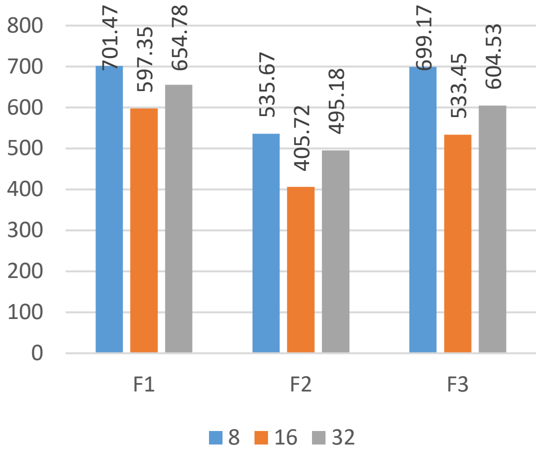

Figure 17. The influence of the number of nodes of Dense_1 through experiments.

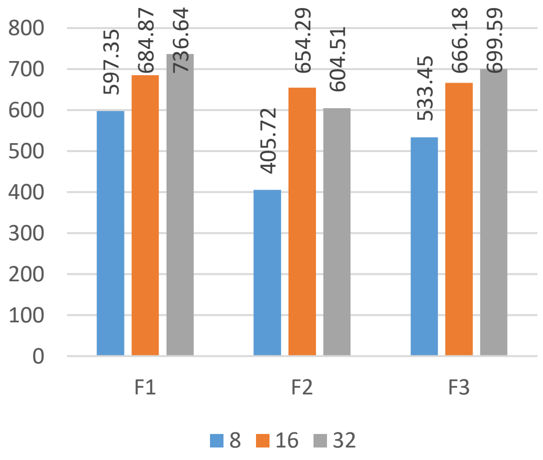

Figure 18. The influence of the number of nodes of Dense_2 through experiments.

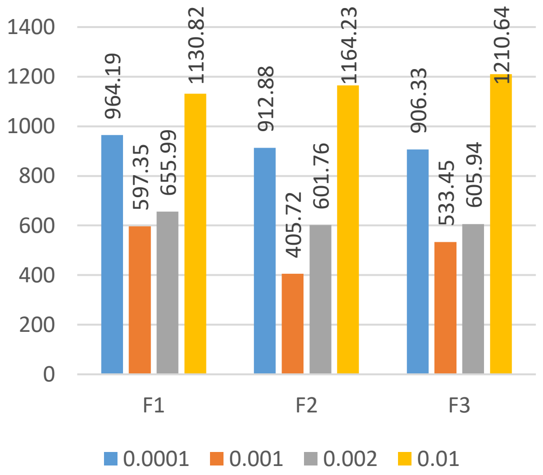

Figure 19. The influence of learning rate values through experiments.

4.2 | Comparison Results

In this particular section, the results achieved by ATCN-LSTM for the F1, F2, and F3 scenarios within the 2018 PHM Data Challenge Dataset are contrasted with competing models to demonstrate their efficacy and productivity. These results are quantified by the RMSE metric to showcase the efficiency of the models in minimizing the difference between the anticipated and desired RUL. The outcomes of the proposed model F1, F2, and F3 conditions from the 2018 PHM Data Challenge Dataset are compared with several competing models such as RFR [49], MLP [49], LSTM [49], TCLSTM [50], DW-GRU [15], DW-GRU-FCs [15], and HF-MS-MBTransformer [51]. The superior performance of the system is demonstrated through the presentation of results, which can be observed in the RMSE values provided in Table 4.

Table 4. Performance comparison of three fault modes.

	F1	F2	F3
RFR	5476	5567	5294
MLP	5196	4113	5004
LSTM	1469	2557	1877
TCLSTM	601.47	748.12	541.24
DW-GRU	1014	489	686
DW-GRU-FCs	998	409	703
HF-MS-MBTransformer	646.42	798.29	691.31
Proposed method	**597.35**	**405.72**	**533.45**

The superior results are emphasized in bold format. The table illustrated indicates that ATCN-LSTM surpasses all other models examined in terms of the RMSE across three specified scenarios, achieving RMSE values of 597.35, 741.72, and 533.45 for F1, F2, and F3, respectively. Upon comparison of our findings with the top results achieved by various cited models, our proposed model demonstrates a reduction in RMSE of 0.6%, 0.7%, and 1.4% for the 2018 PHM Data Challenge Dataset under conditions F1, F2, and F3, respectively. The proposed model is viewed as a robust option for addressing this issue as it has the potential to perform the most effectively based on the RMSE metric, which assigns the same level of importance to predictions made early on as well as those made later. To visually demonstrate the effectiveness of the suggested model, Figure 20 has been incorporated to present the RMSE values derived from various algorithms for F1, F2, and F3 conditions.

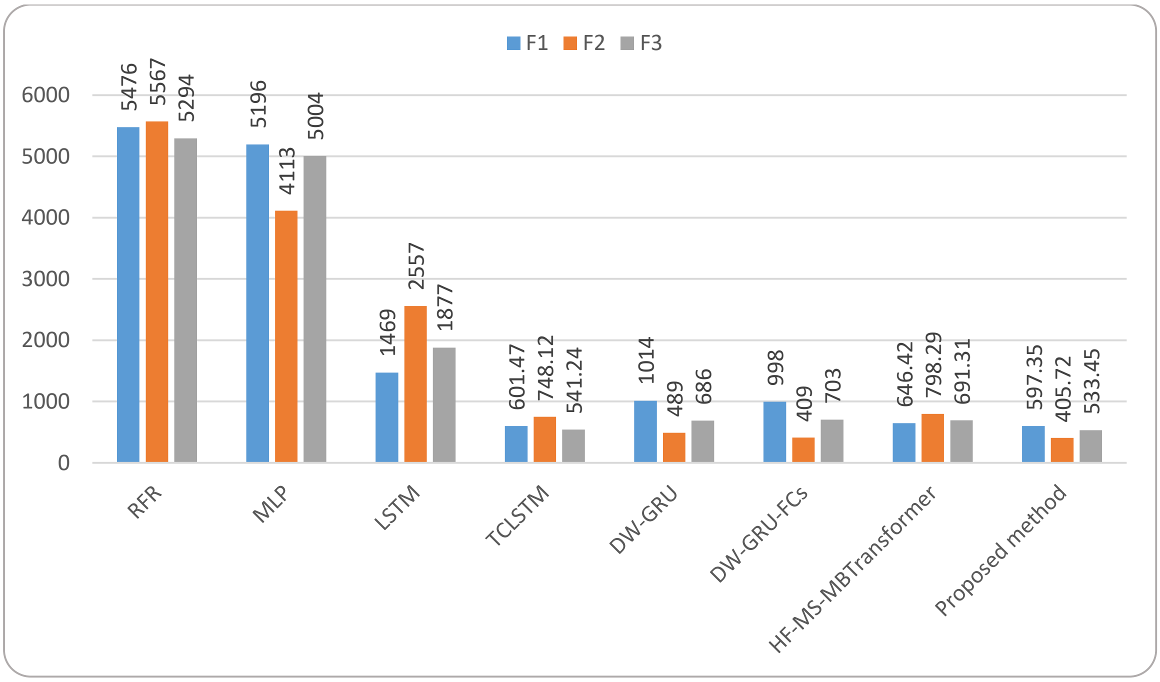

Figure 20. illustrates the representation of RMSE values acquired from different models.

4.3 | Ablation Study Results

To assess the effects of individual components within the ATCN-LSTM, the suggested approach is utilized to carry out ablation experiments. Therefore, this section introduces six experiments that will be compared with the ATCN-LSTM, all carried out in the same conditions. These experiments involve the utilization of various neural network architectures, including an LSTM model without any attention mechanism, a TCN model also without any attention mechanism, an integrated model combining LSTM and TCN (referred to as TCN-LSTM), an integrated model combining LSTM with a self-attention mechanism (A-LSTM), a model that combines TCN with a self-attention mechanism (A-TCN), and a model that integrates LSTM with TCN and a self-attention mechanism (ATCN-LSTM). Moreover, the results of the experiments are thoroughly presented in Table 5 and visually represented in Figure 21. The ATCN-LSTM demonstrates better results in the RMSE metric compared to LSTM, TCN, TCN-LSTM, A-LSTM, and A-TCN across the F1, F2, and F3 conditions. Under the first condition (F1), the RMSE of the proposed technique ATCN-LSTM is measured at 597.35, showcasing improvements of 28.5%, 41.7%, 38.3%, 20.8%, and 39.4% over TCN, LSTM, TCN-LSTM, A-TCN, and A-LSTM, respectively. Similarly, under the second condition (F2), the RMSE of ATCN-LSTM stands at 405.72, demonstrating enhancements of 50%, 65.6%, 49.1%, 30.9%, and 58.2% in comparison to TCN, LSTM, TCN-LSTM, A-TCN, and A-LSTM, respectively. Finally, for the third condition (F3), the RMSE of ATCN-LSTM is recorded at 533.45, revealing improvements of 38.3%, 52.1%, 46.6%, 30%, and 46.8% over TCN, LSTM, TCN-LSTM, A-TCN, and A-LSTM, respectively.

Table 5. Performance comparison of the ablation study.

	F1	F2	F3
TCN	836.19	812.84	865.27
LSTM	1024.66	1180.41	1113.28
TCN-LSTM	968.56	797.21	999.65
A-TCN	754.36	587.48	762.91
A-LSTM	985.24	950.73	1002.94
ATCN-LSTM	**597.35**	**405.72**	**533.45**

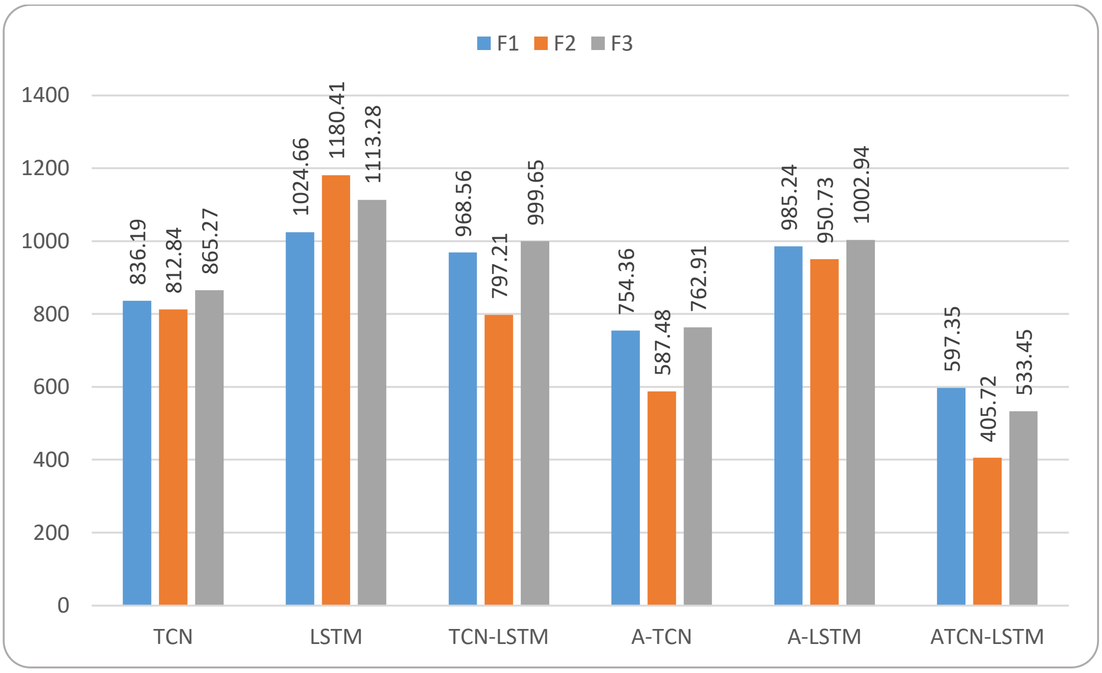

Figure 21. The representation of RMSE values acquired through ablation experiments.

5 | Conclusion

This study introduced a novel approach named ATCN-LSTM for predicting the Remaining Useful Life (RUL) in the ion mill etching process. The ATCN-LSTM model integrates TCN, LSTM, and a self-attention mechanism, with TCN utilized to capture extended dependencies in time series data. LSTMs are employed to retain patterns across lengthy sequences, enabling the model to grasp intricate temporal relationships in temporal data. Furthermore, an attention mechanism is incorporated to align input and output sequences by considering the context or importance of the input sequence. The RUL predictions are produced through a Fully Connected (FC) layer. The experimental evaluation was conducted using the PHM 2018 Challenge dataset. Comparative analysis with leading models in the field demonstrated that our proposed model achieved a decrease in RMSE of .6%, .7%, and 1.4% for the three dataset conditions, respectively, to capture extended dependencies in temporal data sequences, Flowcool Pressure Too High Check Flowcool Pump, and Flowcool Pressure Too High Check Flowcool Pump.

Acknowledgments

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Funding

This research was conducted without external funding support.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Vogl, G. W., Weiss, B. A., & Helu, M. (2019). A review of diagnostic and prognostic capabilities and best practices for manufacturing. Journal of Intelligent Manufacturing, 30, 79-95. https://doi.org/10.1007/s10845-016-1228-8

[2] Umeda, S., Tamaki, K., Sumiya, M., & Kamaji, Y. (2021). Planned maintenance schedule update method for predictive maintenance of semiconductor plasma etcher. IEEE Transactions on Semiconductor Manufacturing, 34(3), 296-300. https://doi.org/10.1109/TSM.2021.3071487

[3] Cao, G. (2023). Remaining Useful Life prediction of Aircraft Engines Using DCNN-BiLSTM with K-means Feature Selection. Paper presented at the International Symposium on Artificial Intelligence and Robotics. https://doi.org/10.1007/978-981-99-9109-9_35

[4] He, A., & Jin, X. (2019). Failure detection and remaining life estimation for ion mill etching process through deep-learning based multimodal data fusion. Journal of Manufacturing Science and Engineering, 141(10), 101008. https://doi.org/10.1115/1.4044248

[5] Vishnu, T., Gupta, P., Malhotra, P., Vig, L., & Shroff, G. (2018). Recurrent neural networks for online remaining useful life estimation in ion mill etching system. Paper presented at the Proceedings of the Annual Conference of the PHM Society, Philadelphia, PA, USA.

[6] Anon. (2018). 2018 phm data challenge. Retrieved from: https://www.phmsociety.org/events/conference/phm/18

[7] Stauss, W., & Lizotte, T. (2018). The cutting edge of ion beam etch and thin film technology. http:/www. microfabnh. com/ion_beam_etch_design_guide. php, Accessed July, 31, 2019.

[8] Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2014). Machine learning for predictive maintenance: A multiple classifier approach. IEEE Transactions on Industrial Informatics, 11(3), 812-820. https://doi.org/10.1109/TII.2014.2349359

[9] Khelif, R., Chebel-Morello, B., Malinowski, S., Laajili, E., Fnaiech, F., & Zerhouni, N. (2016). Direct remaining useful life estimation based on support vector regression. IEEE Transactions on Industrial Electronics, 64(3), 2276-2285. https://doi.org/10.1109/TIE.2016.2623260

[10] Xia, T., Dong, Y., Xiao, L., Du, S., Pan, E., & Xi, L. (2018). Recent advances in prognostics and health management for advanced manufacturing paradigms. Reliability Engineering & System Safety, 178, 255-268. https://doi.org/10.1016/j.ress.2018.06.021

[11] Zhang, W., Yang, D., & Wang, H. (2019). Data-driven methods for predictive maintenance of industrial equipment: A survey. IEEE systems journal, 13(3), 2213-2227. https://doi.org/10.1109/JSYST.2019.2905565

[12] Jiang, Y., & Yin, S. (2018). Recent advances in key-performance-indicator oriented prognosis and diagnosis with a MATLAB toolbox: DB-KIT. IEEE Transactions on Industrial Informatics, 15(5), 2849-2858. https://doi.org/10.1109/TII.2018.2875067

[13] Cui, L., Wang, X., Wang, H., & Ma, J. (2019). Research on remaining useful life prediction of rolling element bearings based on time-varying Kalman filter. IEEE Transactions on Instrumentation and Measurement, 69(6), 2858-2867. https://doi.org/10.1109/TIM.2019.2924509

[14] Sikorska, J. Z., Hodkiewicz, M., & Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. Mechanical systems and signal processing, 25(5), 1803-1836. https://doi.org/10.1016/j.ymssp.2010.11.018

[15] Wu, S., Jiang, Y., Luo, H., & Yin, S. (2021). Remaining useful life prediction for ion etching machine cooling system using deep recurrent neural network-based approaches. Control Engineering Practice, 109, 104748. https://doi.org/10.1016/j.conengprac.2021.104748

[16] Sadabadi, K. K., Jin, X., & Rizzoni, G. (2021). Prediction of remaining useful life for a composite electrode lithium ion battery cell using an electrochemical model to estimate the state of health. Journal of Power Sources, 481, 228861. https://doi.org/10.1016/j.jpowsour.2020.228861

[17] Cai, S., & Zhang, G. (2017). Fatigue life prediction of high-speed railway bearing based on contact stress. Paper presented at the 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM). https://doi.org/10.1109/ICCIS.2017.8274854

[18] Sun, J., Zhang, X., & Wang, J. (2023). Lightweight bidirectional long short-term memory based on automated model pruning with application to bearing remaining useful life prediction. Engineering Applications of Artificial Intelligence, 118, 105662. https://doi.org/10.1016/j.engappai.2022.105662

[19] Wu, J., Hu, K., Cheng, Y., Zhu, H., Shao, X., & Wang, Y. (2020). Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network. ISA transactions, 97, 241-250. https://doi.org/10.1016/j.isatra.2019.07.004

[20] Fu, S., Zhang, Y., Lin, L., Zhao, M., & Zhong, S.-s. (2021). Deep residual LSTM with domain-invariance for remaining useful life prediction across domains. Reliability Engineering & System Safety, 216, 108012. https://doi.org/10.1016/j.ress.2021.108012

[21] Gao, Z., Cecati, C., & Ding, S. X. (2015). A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. IEEE Transactions on Industrial Electronics, 62(6), 3757-3767. https://doi.org/10.1109/TIE.2015.2417501

[22] Sigaud, O., & Droniou, A. (2015). Towards deep developmental learning. IEEE Transactions on Cognitive and Developmental Systems, 8(2), 99-114. https://doi.org/10.1109/TAMD.2015.2496248

[23] Zhang, J., Jiang, Y., Luo, H., & Yin, S. (2021). Prediction of material removal rate in chemical mechanical polishing via residual convolutional neural network. Control Engineering Practice, 107, 104673. https://doi.org/10.1016/j.conengprac.2020.104673

[24] Zhang, C., He, Y., Yuan, L., Xiang, S., & Wang, J. (2015). Prognostics of lithium-ion batteries based on wavelet denoising and DE-RVM. Computational intelligence and neuroscience, 2015, 14-14. https://doi.org/10.1155/2015/918305

[25] Berghout, T., & Benbouzid, M. (2022). A systematic guide for predicting remaining useful life with machine learning. Electronics, 11(7), 1125. https://doi.org/10.3390/electronics11071125

[26] Ali, J. B., Chebel-Morello, B., Saidi, L., Malinowski, S., & Fnaiech, F. (2015). Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. Mechanical systems and signal processing, 56, 150-172. https://doi.org/10.1016/j.ymssp.2014.10.014

[27] Wu, J.-Y., Wu, M., Chen, Z., Li, X.-L., & Yan, R. (2021). Degradation-aware remaining useful life prediction with LSTM autoencoder. IEEE Transactions on Instrumentation and Measurement, 70, 1-10. https://doi.org/10.1109/TIM.2021.3055788

[28] Tziolas, T., Papageorgiou, K., Theodosiou, T., Papageorgiou, E., Mastos, T., & Papadopoulos, A. (2022). Autoencoders for anomaly detection in an industrial multivariate time series dataset. Engineering Proceedings, 18(1), 23. https://doi.org/10.3390/engproc2022018023

[29] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324. https://doi.org/10.1109/5.726791

[30] Li, X., Zhang, W., & Ding, Q. (2019). Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. Reliability Engineering & System Safety, 182, 208-218. https://doi.org/10.1016/j.ress.2018.11.011

[31] Zhu, J., Chen, N., & Peng, W. (2018). Estimation of bearing remaining useful life based on multiscale convolutional neural network. IEEE Transactions on Industrial Electronics, 66(4), 3208-3216. https://doi.org/10.1109/TIE.2018.2844856

[32] Ma, M., & Mao, Z. (2020). Deep-convolution-based LSTM network for remaining useful life prediction. IEEE Transactions on Industrial Informatics, 17(3), 1658-1667. https://doi.org/10.1109/TII.2020.2991796

[33] Hong, C. W., Lee, C., Lee, K., Ko, M.-S., Kim, D. E., & Hur, K. (2020). Remaining useful life prognosis for turbofan engine using explainable deep neural networks with dimensionality reduction. Sensors, 20(22), 6626. https://doi.org/10.3390/s20226626

[34] Catelani, M., Ciani, L., Fantacci, R., Patrizi, G., & Picano, B. (2021). Remaining useful life estimation for prognostics of lithium-ion batteries based on recurrent neural network. IEEE Transactions on Instrumentation and Measurement, 70, 1-11. https://doi.org/10.1109/TIM.2021.3111009

[35] Graves, A., & Graves, A. (2012). Long short-term memory. Supervised sequence labelling with recurrent neural networks, 37-45. https://doi.org/10.1007/978-3-642-24797-2_4

[36] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[37] Zhang, J., Wang, P., Yan, R., & Gao, R. X. (2018). Long short-term memory for machine remaining life prediction. Journal of manufacturing systems, 48, 78-86. https://doi.org/10.1016/j.jmsy.2018.05.011

[38] Wang, F.-K., Amogne, Z. E., Chou, J.-H., & Tseng, C. (2022). Online remaining useful life prediction of lithium-ion batteries using bidirectional long short-term memory with attention mechanism. Energy, 254, 124344. https://doi.org/10.1016/j.energy.2022.124344

[39] Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259. https://doi.org/10.48550/arXiv.1409.1259

[40] Chen, J., Jing, H., Chang, Y., & Liu, Q. (2019). Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process. Reliability Engineering & System Safety, 185, 372-382. https://doi.org/10.1016/j.ress.2019.01.006

[41] Wang, H.-K., Cheng, Y., & Song, K. (2021). Remaining useful life estimation of aircraft engines using a joint deep learning model based on TCNN and transformer. Computational intelligence and neuroscience, 2021. https://doi.org/10.1155/2021/5185938

[42] Jayasinghe, L., Samarasinghe, T., Yuenv, C., Low, J. C. N., & Ge, S. S. (2019). Temporal convolutional memory networks for remaining useful life estimation of industrial machinery. Paper presented at the 2019 ieee international conference on industrial technology (icit). https://doi.org/10.1109/ICIT.2019.8754956

[43] Zhang, Z., Song, W., & Li, Q. (2022). Dual-aspect self-attention based on transformer for remaining useful life prediction. IEEE Transactions on Instrumentation and Measurement, 71, 1-11. https://doi.org/10.1109/TIM.2022.3160561

[44] Ma, Q., Zhang, M., Xu, Y., Song, J., & Zhang, T. (2021). Remaining useful life estimation for turbofan engine with transformer-based deep architecture. Paper presented at the 2021 26th international conference on automation and computing (ICAC). https://doi.org/10.23919/ICAC50006.2021.9594150

[45] Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., . . . Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. arXiv preprint arXiv:1609.03499. https://doi.org/10.48550/arXiv.1609.03499

[46] Liu, C., Zhang, L., & Wu, C. (2019). Direct remaining useful life prediction for rolling bearing using temporal convolutional networks. Paper presented at the 2019 IEEE Symposium series on computational intelligence (SSCI). https://doi.org/10.1109/SSCI44817.2019.9003163

[47] Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271. https://doi.org/10.48550/arXiv.1803.01271

[48] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980

[49] Huang, W., Khorasgani, H., Gupta, C., Farahat, A., & Zheng, S. (2018). Remaining useful life estimation for systems with abrupt failures. Paper presented at the Annual conference of the PHM society. September.

[50] Hsu, C.-Y., Lu, Y.-W., & Yan, J.-H. (2022). Temporal convolution-based long-short term memory network with attention mechanism for remaining useful life prediction. IEEE Transactions on Semiconductor Manufacturing, 35(2), 220-228. https://doi.org/10.1109/TSM.2022.3164578

[51] Yuan, Z., & Wang, R. (2023). Multi-Scale and Multi-Branch Transformer Network for Remaining Useful Life Prediction in Ion Mill Etching Process. IEEE Transactions on Semiconductor Manufacturing. https://doi.org/10.1109/TSM.2023.3324057

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Sciences Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.