





Article

# Long-Term Lifetime Prediction of Power MOSFET Devices Based on LSTM and GRU Algorithms

Mesfin Seid Ibrahim <sup>1,2,\*</sup>, Waseem Abbas <sup>1</sup>, Muhammad Waseem <sup>1,3</sup>, Chang Lu <sup>1</sup>, Hiu Hung Lee <sup>1</sup>, Jiajie Fan <sup>4,5,6</sup> and Ka-Hong Loo <sup>1,3,\*</sup>

- <sup>1</sup> Centre for Advances in Reliability and Safety, New Territories, Hong Kong
  - <sup>2</sup> Kombolcha Institute of Technology, Wollo University, Kombolcha P.O. Box 208, Ethiopia
  - <sup>3</sup> Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
  - <sup>4</sup> Institute of Future Lighting, Academy for Engineering & Technology, Fudan University, Shanghai 200433, China
  - <sup>5</sup> Shanghai Engineering Technology Research Center for SiC Power Device, Fudan University, Shanghai 200433, China
  - <sup>6</sup> Institute of Wide Bandgap Semiconductor Materials and Devices, Research Institute of Fudan University in Ningbo, Fudan University, Ningbo 315336, China
- \* Correspondence: mesfin.ibrahim@connect.polyu.hk (M.S.I.); kh.loo@polyu.edu.hk (K.-H.L.)

**Abstract:** Predicting the long-term lifetime of power MOSFET devices plays a central role in the prevention of unprecedented failures for power MOSFETs used in safety-critical applications. The various traditional model-based approaches and statistical and filtering algorithms for prognostics have limitations in terms of handling the dynamic nature of failure precursor degradation data for these devices. In this paper, a prognostic model based on LSTM and GRU is developed that aims at estimating the long-term lifetime of discrete power MOSFETs using dominant failure precursor degradation data. An accelerated power cycling test has been designed and executed to collect failure precursor data. For this purpose, commercially available power MOSFETs passed through power cycling tests at different temperature swing conditions and potential failure precursor data were collected using an automated curve tracer after certain intervals. The on-state resistance degradation data identified as one of the dominant failure precursors and potential aging precursors has been analyzed using RNN, LSTM, and GRU-based algorithms. The LSTM and GRU models have been found to be superior compared to RNN, with MAPE of 0.9%, 0.78%, and 1.72% for MOSFET 1; 0.90%, 0.66%, and 0.6% for MOSFET 5; and 1.05%, 0.9%, and 0.78%, for MOSFET 9, respectively, predicted at 40,000 cycles. In addition, the robustness of these methods is examined using training data at 24,000 and 54,000 cycles of starting points and is able to predict the long-term lifetime accurately, as evaluated by MAPE, MSE, and RMSE metrics. In general, the prediction results showed that the prognostics algorithms developed were trained to provide effective, accurate, and useful lifetime predictions and were found to address the reliability concerns of power MOSFET devices for practical applications.

**Keywords:** LSTM; GRU; power cycling; power MOSFETs; long-term lifetime prediction; failure precursors

**MSC:** 62N05; 90B25



**Citation:** Ibrahim, M.S.; Abbas, W.; Waseem, M.; Lu, C.; Lee, H.H.; Fan, J.; Loo, K.-H. Long-Term Lifetime Prediction of Power MOSFET Devices Based on LSTM and GRU Algorithms. *Mathematics* **2023**, *11*, 3283. <https://doi.org/10.3390/math11153283>

Academic Editors: Rui Peng, Kaiye Gao and Wen Zhang

Received: 15 June 2023  
Revised: 20 July 2023  
Accepted: 23 July 2023  
Published: 26 July 2023



**Copyright:** © 2023 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

Power electronics technology has shown tremendous advancements as an essential component of power conversion and control processes to meet specific demands. Metal Oxide Field Effect Transistors (MOSFETs) and Insulated Gate Bipolar Transistors (IGBTs) have been used in a wide spectrum of power electronics systems. The application area ranges from automotive and locomotive to avionics and aerospace and other space missions as well as safety-critical operations. The proper functioning and health monitoring of these

devices are of utmost importance to avoiding downtime, human safety, and the prevention of catastrophic failures [1].

The discrete power devices or modules in power electronic devices will be exposed to various stresses including electrical loading, thermal stresses, and mechanical vibration as well as humidity and chemical stresses as operating or environmental conditions. These stresses cause degradation or catastrophic failure, which affects the proper functioning and hence the lifetime and reliability of electrical and electronic products and systems. This has been the main driving force behind improving the reliability of power devices. As part of this, the reliability requirement of safety and mission-critical applications such as automotive, locomotive, and airplane applications oblige almost zero-defect or zero-failure tolerance. Regardless of the efforts that significantly reduced the failure rates of power devices, reliability remains the central focus in many application areas [2]. The demand for uninterrupted and consistent power delivery, system availability, and the safety of these critical applications has led to the need for advanced reliability assessment techniques, and accurate prognostics have been instigated by [3,4]. This helps to optimize design parameters and quality characteristics, which ultimately helps to minimize failure tolerances and establish maintenance strategies [5,6].

Prognostics and health management (PHM) is an engineering method enabling the diagnostics and prognostics of products and systems based on in situ health monitoring, offline degradation patterns, and the identification of failure modes and mechanisms. There are multiple failure modes of power MOSFETs that can be grouped as package structure failure and precursor drift failure. Each failure mode can be caused by one or multiple failure mechanisms. According to the literature, the common failure mechanisms in power devices can be either chip-related (intrinsic) or package-related (extrinsic) [1,5].

Chip-related failures occur during the body diode and gate oxide degradation, while package-related failures happen during bond wire failures (liftoff or fracture) and solder layers degradation due to wear-out (degradation) or overstress [2,7]. In fact, the package-related failures are due to thermomechanical stress caused by a mismatch in the coefficient of thermal expansion of multilayer material. This mismatch, in turn, occurs when power MOSFETs are subjected to thermal cycling stress. As the reliability of such power semiconductor devices is greatly affected due to various failure mechanisms, lifetime prediction has become a critical issue in avoiding catastrophic failures by taking early warning measures. The lifetime prediction of power devices demands understanding the dominant failure modes and mechanisms. Consequently, the first and most important task in the lifetime prediction of power devices is the extraction of failure precursor data. Thereupon, the extraction of dominant failure precursor parameters leads to the selection of an appropriate prognostic method [1]. The main failure precursors often used for power MOSFETs are on-state resistance ( $R_{ds(on)}$ ), threshold voltage ( $V_{th}$ ), and junction temperature ( $T_j$ ), while collector-emitter voltage ( $V_{CE}$ ) and collector-emitter current ( $I_{CE}$ ) [8–11] are used for IGBTs.

The studies on PHM for power devices can be divided into three categories: physics-based, data-driven, and fusion prognostic approaches. Physics-based models, also known as model-based approaches, employ mathematical models or equations developed based on the first principle of damage or degradation mechanisms observed with experiments. These models utilize knowledge about the system's or product's lifecycle loading and failure mechanisms to perform reliability modeling and lifetime assessment [12,13]. For example, model-based approaches make use of junction temperature swings and cycles for failure data based on empirical or analytical lifetime models such as the Coffin–Manson model, the Bayerer model, the Norris–Landzberg model, and so on in the lifetime estimation of power devices [1,14].

In model-based approaches, filtering algorithms such as Kalman filtering (KF) and particle filtering (PF) as well as Bayesian methods are usually employed to estimate model parameters recursively by using measured data. Dusmez et al. [15] used the KF algorithm for the lifetime estimation of thermally stressed MOSFETs, with an assumption of a linear system with Gaussian noise, which is often idealistic. Patil et al. [9] used the particle

filter algorithm in the lifetime prediction of IGBTs with emitter–collector current data. Wu et al. [16] employed an improved PF algorithm on a thermally overstressed MOSFET dataset. Although there is some progress in the model-based prognostics for power MOSFETs, physical models that describe the evolution of degradation are not accurate, as the future operating condition is usually uncertain. In addition, it is challenging to formulate a physical model that describes the degradation of power MOSFETs working under complex and dynamic systems. Also, the popular methods for model parameter estimation, such as PF, have the drawback of particle degeneracy, and KF assumes ideal system linearity with Gaussian noise.

On the other hand, data-driven approaches implement algorithms and methods that recognize patterns from large amounts of experimental or simulation data to drive empirical degradation models. These methods do not require an explicit mathematical model to describe the evolution of power MOSFET degradation. Recently, artificial intelligence (AI) and machine learning algorithms benefited the power electronics sector due to their immense potential in anomaly detection, diagnostics, and the prognostics of semiconductor device degradation [17]. There are very few assumptions made about the underlying principles governing the lifetime of power devices, as machine learning predictive algorithms are trained with raw data performance indicators or precursors. Pugalenth, et al. [18] employed a feed-forward neural network to model the reliability of power converters using run-to-failure data. Although they are able to predict the future lifetime, vanilla neural networks are not designed for sequential data and are less accurate. Zhao et al. [19] derived composite precursor parameters and applied a genetic programming algorithm in the lifetime prediction of power devices. Similarly, fusion prognostic approaches can be employed in lifetime prediction, as they leverage the advantages of both data-driven and model-driven methods.

In general, classical model-based approaches and statistical and filtering algorithms have limitations in terms of handling the dynamic failure precursor (performance data) degradation nature of power devices. Although there are some studies on lifetime prediction, they are not often on the prognostics of power MOSFETs based on degradation data collected under realistic working conditions and do not utilize advanced prognostic approaches. Recently, the recurrent neural network (RNN) has been used to overcome these shortcomings. The notion of time step introduced in RNNs makes it suitable for sequential learning [20], including natural language processing, machine translation, image captioning, genomic analysis, and so on [21]. Nevertheless, RNN also suffers from the problem of gradient exploding or vanishing, which led to the introduction of the long short-term memory (LSTM) algorithm. LSTM networks are a popular variant of RNN that are able to address the shortcomings of traditional RNN due to their long-term memory capacity. With their immense ability to predict long-term degradation dependency, deep learning algorithms such as LSTM and GRU have been widely used for time-series data prediction in LEDs, batteries, and other applications [22,23].

In this paper, a prognostic method based on LSTM and its variant GRU algorithm is developed to estimate the lifetime of power devices. The study mainly focuses on the long-term lifetime prediction of power MOSFETs using failure precursor data. An accelerated aging test based on power cycling is conducted to gather electrical failure precursor parameters for the power device test samples. The accelerated aging test designed in this work is based on a power cycling test where power devices are supplied with electrical power and exposed to a certain period of ON/OFF, which is suitable for mimicking real application conditions. The on-state resistance degradation precursor will be used as the long-term time-series data to train and validate the proposed LSTM and GRU models. The failure precursors have been partitioned as training and testing sets for proposed neural network structures to prevent information leakage in the developed model.

The remaining part of this paper is organized as follows: Section 2 describes the prognostic model developed based on the deep learning algorithms RNN, LSTM, and GRU. The experimental design and setup employed to gather the accelerated aging degradation

data are described in Section 3. In Section 4, the experimental results, the details of data analysis, as well as results and discussions are presented, and conclusions are drawn in Section 5.

## 2. Description of the Proposed Theory and Methodology

In this section, an overview of the proposed methodologies will be discussed. The proposed models and algorithms for modeling the degradation of power MOSFET device test samples and deep learning-based reliability assessment and lifetime prediction are introduced. The prognostic model developed based on the LSTM and GRU algorithms as well as the model regularization and prediction accuracy metrics for the performance degradation of power MOSFET test samples are presented. Figure 1 shows an overview of the experimental design and setup for data collection, the prognostic model with LSTM/GRU networks, and the long-term lifetime prediction and evaluation methodology.

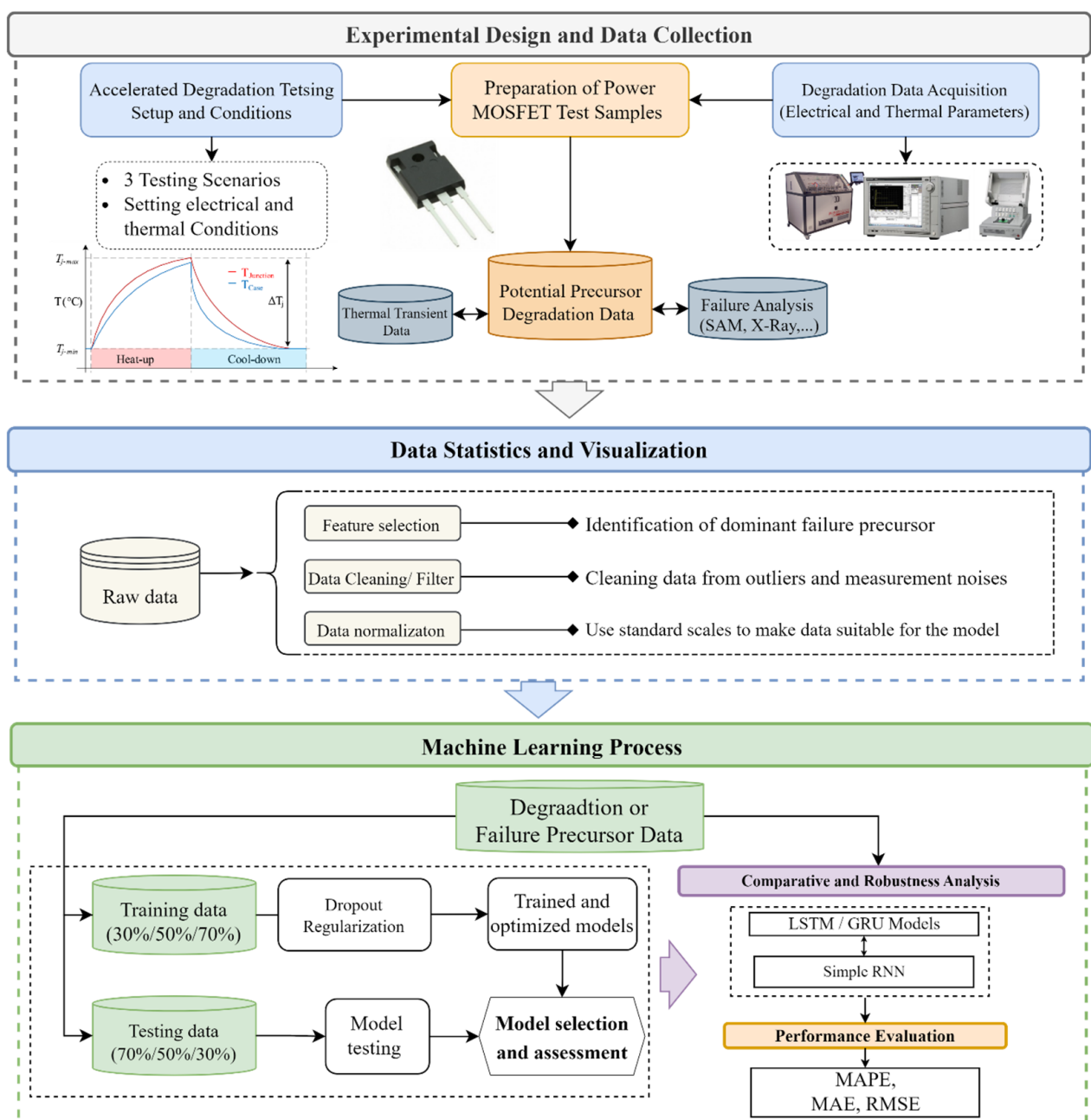


Figure 1. An Overview of LSTM- and GRU-based Reliability Assessment Methodology.

There are a few considerations that have to be addressed in the application of LSTM for the long-term lifetime prediction of power MOSFETs. The first consideration is an appropriate choice of training algorithms that are suitable for deep learning, as some of the classical stochastic gradient descent- or batch gradient descent-based training algorithms tend to be slower to converge to optimal solutions. The second issue is handling the problem of overfitting, as it is one of the drawbacks of deep neural networks in general and the LSTM algorithm in particular.

Thus, a prognostic model based on the LSTM and GRU long-term lifetime prediction algorithms is developed to address these challenges by incorporating three elements. These are the LSTM/GRU algorithm architectures and network parameter training using the Adam optimization method as well as the dropout to prevent overfitting problems encountered in deep learning. Adaptive moment optimization (Adam) is an efficient optimization algorithm that uses estimates of a gradient’s first and second moments to adapt the learning rate for each neural network weight. This algorithm is often used instead of the traditional stochastic gradient descent procedure, as it requires low memory and combines the advantages of its predecessor, the adaptive gradient algorithm (AdaGrad), and root mean square propagation (RMSProp) [24].

2.1. An Overview of RNN, LSTM, and GRU

As computing power grows, machine learning algorithms have become an essential part of many industrial and business sectors. The application of machine learning algorithms has shown tremendous growth in the diagnostics, prognostics, anomaly detection, and general reliability assessment of power electronic components and systems. This is because machine learning algorithms are able to overcome the drawbacks of traditional model-based and statistical approaches in handling uncertainties in terms of noise factors, unknown failure mechanisms, as well as dynamic environmental and loading conditions [17].

2.1.1. Standard Recurrent Neural Networks (RNN) Architecture

In classical multilayer perceptron (MLP), also known as feed-forward neural networks, information flows in one direction (i.e., forward propagation), which makes it not suitable for time series data. Recurrent neural networks are variants of neural networks that are suitable for sequence learning, as they allow for the use of the previous layer output as the input to the current state in order to predict the future layer output [20]. This cell architecture enables RNNs to have a state and thus a memory which is used to capture information operations in previous states. A typical architecture for an RNN unrolled in time  $t$  in the estimation of performance parameters (such as  $R_{dson}$ ) for power MOSFETs is shown in Figure 2.

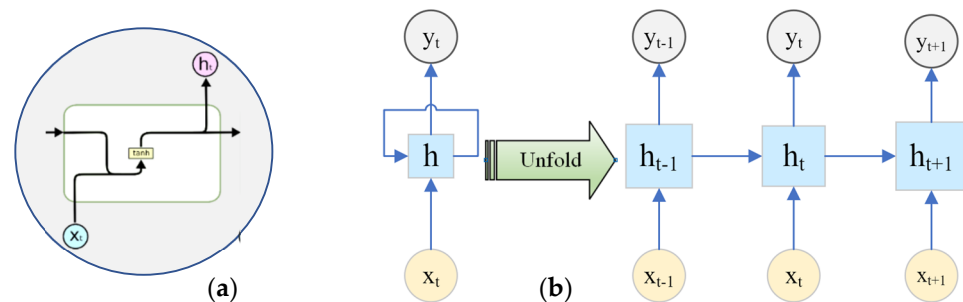


Figure 2. (a) RNN Cell Architecture/Internal structure of a cell; (b) An unrolled RNN Structure.

It can be noticed that the recurrent neuron is fed not only information from  $x_t$  from the current time step  $t_i$  but also a hidden state  $h_{t-1}$  from the previous time step  $t_{i-1}$ . Then,

an output  $y_t$  will be generated at the output layer along with an updated hidden state  $h_t$ . This process can generally be described mathematically, as follows:

$$h_t = f(W_h x_t + U_h h_{t-1} + b_h) \tag{1}$$

$$y_t = f(W_y h_t + b_y) \tag{2}$$

where  $f(\cdot)$  is the activation function,  $W_h$ ,  $U_h$ , and  $W_y$  are the weight matrices between the hidden and input layer, the hidden layer and itself, as well as the output and hidden layers, respectively, at adjacent time steps. Similarly, the vectors  $b_h$  and  $b_y$  are bias parameters added at hidden and output layers that enable the nodes to learn an offset [20]. Theoretically, RNNs are thought to have the capability of handling long-term dependencies. Practically, however, RNNs are able to look back 5–10 time steps [21], as they suffer from vanishing or exploding gradients during the back-propagation of error signals. The LSTMs are specifically designed to overcome such problems and are suitable for sequential time series modeling, as discussed in the next section.

### 2.1.2. Long Short-Term Memory (LSTM) Algorithm Architecture

Long short-term memory (LSTM) is a special variant of RNN used in the area of deep learning. Introduced by Hochreiter et al. [25], the LSTM is a powerful algorithm for overcoming exploding and vanishing gradient problems observed with simple RNNs. Due to this, LSTMs are found to be suitable for sequential time-series data analysis with long-term dependencies. The network architecture of an LSTM cell is shown in Figure 3a. It can be observed that an LSTM cell has three sigmoid ( $\sigma$ ) activation functions which control the flow of information and protect the input ( $i_t$ ), forget ( $f_t$ ), and output ( $o_t$ ) gates. These sigmoid gates are activated by inputs from the current input layer  $x_t$  as well as from the hidden layer  $h_{t-1}$  at the previous time step, which enables the LSTM algorithm to function as intended. The cell state (also called the internal state or memory),  $C_t$ , is used to preserve information at the current time.

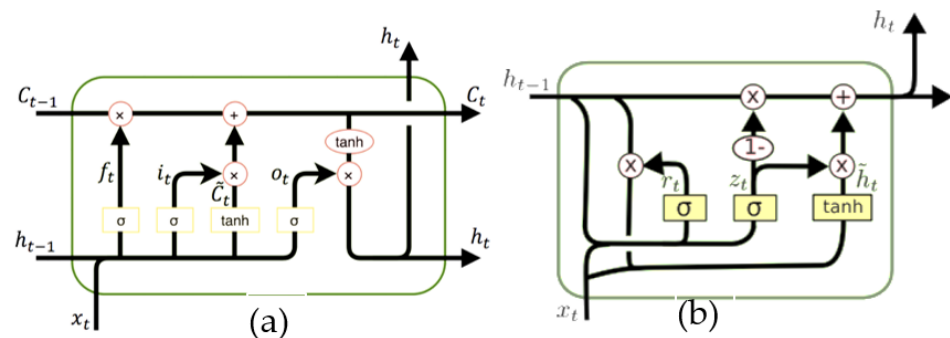


Figure 3. Internal structure/architecture of (a) an LSTM Cell and (b) a GRU Cell.

The mathematical formulation for each gate and the operation of the LSTM algorithm are described in this section. The first step is to identify which information to remember and, on the other hand, to choose the information that has to be forgotten. This is executed by the forget gate, which uses the sigmoid function. The sigmoid function takes the inputs  $x_t$  and  $h_{t-1}$  and outputs values between 0 and 1, where 0 represents discarding everything and 1 represents preserving everything from the previous cell state.

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{3}$$

The second step is to decide what information to store in the cell state, which is determined by the input gate layer with a sigmoid function, and candidate information can be saved as  $\tilde{C}_t$  at new cell states as follows:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{4}$$

$$\tilde{C}_t = \tanh(W_C x_t + U_C h_{t-1} + b_C) \tag{5}$$

Combining Equations (3) to (5), the values of old cell states  $C_{t-1}$  can be updated as the current cell state or cell memory  $C_t$  and calculated as follows:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{6}$$

Third, the output gate layer determines what information to output as  $o_t$  with the sigmoid function and output hidden state  $h_t$ , as follows:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{7}$$

$$h_t = \tanh(C_t) \odot o_t \tag{8}$$

In these equations,  $W_i, U_i; W_o, U_o$  and  $W_f, U_f$  are weight vectors for the input, output, and foregate gate layers, respectively, and  $b_i, b_o,$  and  $b_f$  are the biases for the input, output, and forget gates between the current time  $t$  and previous time  $t - 1$  in the LSTM network. The symbol  $\odot$  is the unit by the element-wise product, also known as the Hadamard product. In addition,  $\sigma$  and  $\tanh$  are sigmoid and hyperbolic tangent activation functions that map values between 0 and 1 and between  $-1$  and 1, respectively. The values for the  $\sigma$  and  $\tanh$  functions are calculated by using:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \text{ and } \tanh = \frac{1 - e^{-2x}}{1 + e^{-2x}} \tag{9}$$

The standard LSTM and GRU algorithms have Bi-LSTM and Bi-GRU architecture variants, which are the variants of the standard networks; they have forward and backward propagation. This paper will not implement the Bi-LSTM and Bi-GRU variants in the deep learning algorithm for sequential data modeling.

### 2.1.3. Gated Recurrent Unit (GRU)

The gated recurrent unit (GRU) is another variation of an LSTM that has a simplified neural network compared to LSTM. GRU was first introduced by Cho et al. [26] for machine translation. It has been widely adopted in the prognostic application for degradation modeling [22]. The GRUs are considered less computationally expensive, as they use fewer training parameters, resulting in faster convergence and a good choice for smaller datasets. On the other hand, LSTMs may work better for larger datasets, as they retain more temporal information. The cell structure of GRU has only two gates (i.e., a forget gate and an update gate), with no separate output gate, as shown in Figure 3b, and the equations are as follows:

$$r_t = \sigma(W_r h_{t-1} + U_r x_t + b_r) \tag{10}$$

$$z_t = \sigma(W_z h_{t-1} + U_z x_t + b_z) \tag{11}$$

$$\tilde{h}_t = \tanh(W(r_t \odot h_{t-1} + Ux_t + b_h)) \tag{12}$$

$$h_t = z_t \odot \tilde{h}_t + 1 - z_t \odot h_{t-1} \tag{13}$$

Here,  $r_t$  and  $z_t$  are the reset gate and update gate, respectively, while  $x_t$  is the current input for the cell at time  $t$ ;  $h_{t-1}$  denotes the previous output of the hidden layer; and  $\sigma$  is the sigmoid function.  $\tilde{h}_t$  is the state candidate that uses the tanh function to scale the data for the current state  $h_t$ . The network architectures of the deep RNN, LSTM, as well as GRU can be represented as shown in Figure 4.

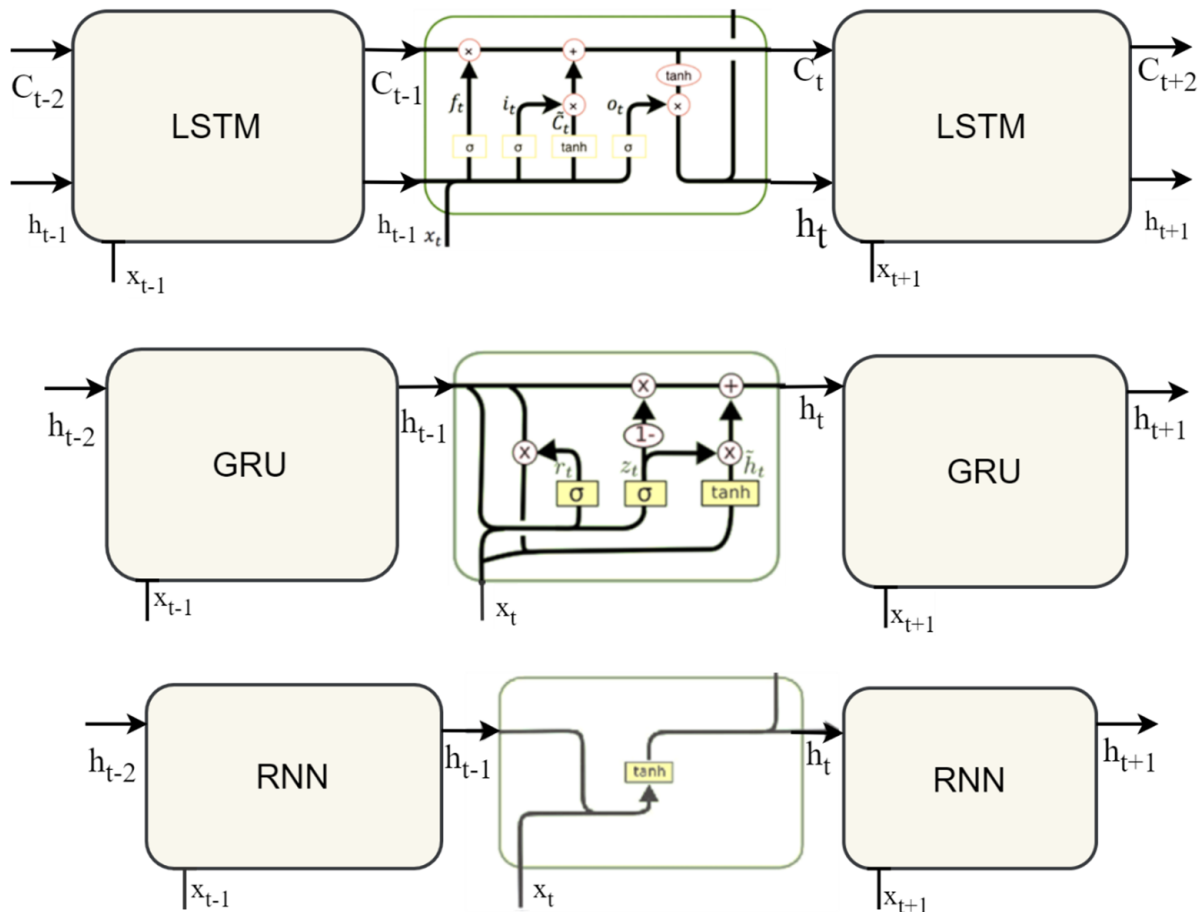


Figure 4. Schematic Diagram of the Deep RNN, LSTM, and GRU Cell Networks.

### 2.2. Training Algorithm for LSTM and Variants

A proper selection of a training algorithm is an important part of learning with neural networks. The gradient descent or loss function has been a widely used optimization algorithm. In addition, the LSTM algorithm has been trained using Root Mean Squared Propagation (RMSProp) and the Adaptive Gradient (AdaGrad) version of gradient descent. Whereas the former is an extension of the gradient descent or loss function, the latter uses a decaying average of partial gradients in the adaptation of the step size for each parameter. In this paper, Adam (adaptive moment estimation) optimization is used, as it helps to realize the advantages of both RMSProp and AdaGrad and is hence considered the popular optimization algorithm for most machine learning approaches.

### 2.3. Model Regularization to Overcome Overfitting

One of the common drawbacks of training neural network algorithms such as LSTM is overfitting. The accuracy of a model will be in doubt, even if it shows overfitting, as the model is trying to capture all the noise and outliers available in a dataset. In general, overfitting is a scenario where the LSTM or GRU model tries to learn from the details along with the noise in the data and tries to fit each data point on the curve, while the model curves may not correspond to the patterns in the new data. To overcome overfitting,



regularization is employed so that it prevents models from either overfitting or underfitting by discouraging the learning of a more complicated or flexible model. Mainly, three types of methods, namely, Lasso (L1), Ridge (L2), and dropout regularization, have been used to prevent overfitting. A combination of L1 and L2, known as Elastic Net regression (L1/L2 regularization), can also be applied. L1 regularization sums up the absolute value of weights to the network, and L2 regularization adds the sum of all squares of the weights in the network. Recently, dropout regularization [27] has been found to be effective for neural networks in preventing overfitting and, as a result, has been used in the prognostics of batteries using LSTM [28].

#### 2.4. Prediction of Long-Term Lifetime

Due to the thermal, electrical, mechanical, and humidity stress that discrete or module devices are exposed to, their failure mechanism leads to degradation or device failure. As a common phenomenon, the degradation failure mechanism dominates, and in the power electronics industry, the failure threshold of power MOSFETs is when the on-state resistance ( $\Delta R_{ds(on)}$ ) reaches up to 17% [1] of its initial or pristine state value. Other performance parameters such as the threshold voltage ( $V_{th}$ ) and drain current ( $I_d$ ) are also considered failure precursors when the threshold values reach 20% [7] and five times ( $5\times$ ) [5] of their initial values, respectively [29]. Thus, for power MOSFETs, the RUL can be estimated when a certain failure precursor crosses the failure threshold.

$$RUL = T_{EOL} - T_{pred} \quad (14)$$

where  $T_{EOL}$  is about a 17% increment from the pristine state  $R_{dson}$  value, considering on-state resistance as the main failure precursor (degradation parameter);  $T_{pred}$  is the time when the prediction started considering a certain portion (30%, 50%, or 70% of training data) from degradation trend data.

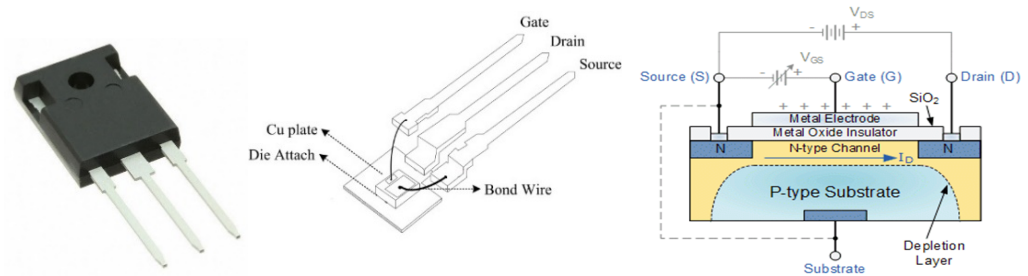
The initial on-state resistance value at a pristine state is 37.137 mohm for DUT #1, 36.733 mohm for DUT #5, and 37.231 mohm for DUT #9; the respective failure threshold will be 44.56 mohm, 42.98 mohm, and 43.56 mohm. However, the degradation value shows that the failure precursor values of the devices have not crossed their failure threshold, and thus, the prediction metrics of MAE, MSE, and RMSE will be suitable for long-term lifetime prediction.

### 3. Experimental Setup and Data Collection

In this section, an overview of the experimental design, and setup for gathering failure precursor parameters or degradation data and assessing the reliability of power devices is presented. The experimental design mainly focuses on the determination of accelerated power cycling conditions and the identification of potential failure precursors suitable for degradation analysis and non-destructive failure analysis. This will be of benefit in determining the main failure precursors that can be used for long-term lifetime prediction, which ultimately helps to decide appropriate maintenance activities and enhance the quality and reliability of such power devices.

#### 3.1. Description of Experimental Test Samples

The test sample considered in this study is a commercial power MOSFET with a power rating of 600 V and a maximum current capacity of 49 A. The packaging is type TO-247, which is designed for high-voltage applications with super-junction (SJ), which is characterized by fast switching and has low conduction and switching losses. The pictorial, schematic, as well as cross-sectional views of the test sample power MOSFET device are presented in Figure 5.



**Figure 5.** Test Samples' Internal Structure and Package Level Structure.

The main performance parameters of the power MOSFET test samples are given in Table 1, as follows.

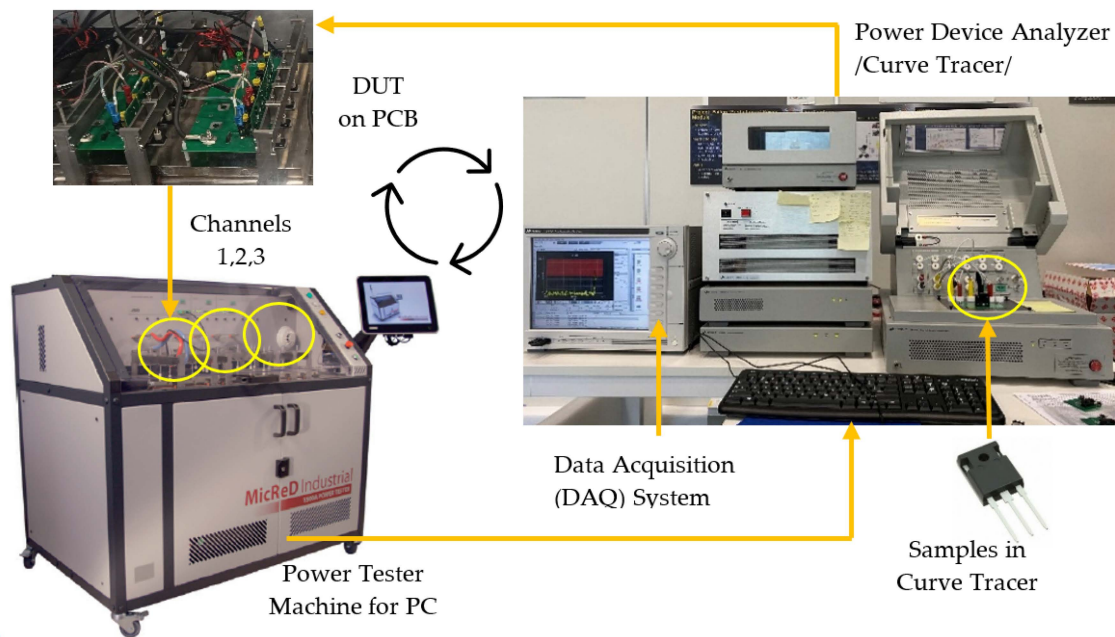
**Table 1.** Basic Performance Parameters of Test Samples.

Parameters	Description of Parameters and Values
$V_{ds} @ T_{jmax}$	650 V
Pulsed drain current $I_{D,pulse}$	272 A
Continuous drain current (ID)	77.5 A @ $T_C = 25\text{ }^\circ\text{C}$ 49 A @ $T_C = 100\text{ }^\circ\text{C}$
$E_{oss} @ 400\text{ V}$	22 $\mu\text{J}$
Power Dissipation, $P_{tot}$	481 W @ $T_C = 25\text{ }^\circ\text{C}$
Body diode $d_i/d_t$	300 A/ $\mu\text{s}$

### 3.2. Experiment Design, Setup, and Data Collection

In this paper, an accelerated degradation test based on power cycling was conducted on power MOSFETs, aimed at assessing the long-term lifetime of these devices at a discrete level by investigating the impact of long-term thermal and electrical stresses on failure precursor parameters or the degradation of performance. Comprehensive failure precursor data are collected, which help to identify dominant parameters that can be explored in the remaining useful life estimation of these devices at the accelerated test condition.

In this experiment, twelve samples of power MOSFETs from the same batch were prepared in three groups with different junction temperature ( $T_j$ ) swing scenarios. Each degradation testing scenario consisted of four test samples, where each group was set for a  $T_j$  swing of  $45\text{ }^\circ\text{C}$ ,  $100\text{ }^\circ\text{C}$ , and  $110\text{ }^\circ\text{C}$ , in which the  $T_{jmin}$  and  $T_{jmax}$  range from  $40\text{ }^\circ\text{C}$  to  $85\text{ }^\circ\text{C}$ ,  $25\text{ }^\circ\text{C}$  to  $125\text{ }^\circ\text{C}$ , and  $25\text{ }^\circ\text{C}$  to  $135\text{ }^\circ\text{C}$ , respectively. The design of the experiment for the different  $T_j$  swing scenarios is based on the various working environments of power MOSFETs, which can be related to real-world conditions. The test samples in all groups were aged under a normal ambient temperature of  $23 \pm 2\text{ }^\circ\text{C}$  for a total of 77,600 cycles. The various static and dynamic electrical characteristics, which are considered as failure precursors, were collected every 400 cycles, where a single cycle covers a time of 45 secs ON (heating) and 90 secs OFF (cooling) with a power tester (MicReD PowerTester 1500 A). In general, the experiment had two phases: accelerated aging and failure precursor parameters testing, which continued until sufficient degradation data were obtained. An overview of the overall experimental design for an accelerated degradation test, the experimental setup, and the data collection procedure is shown in Figure 6.



**Figure 6.** Experimental design and setup for data collection.

In general, the experimental design parameters, testing conditions, and a brief summary of the test duration, number of cycles, as well as failure precursors are shown in Table 2.

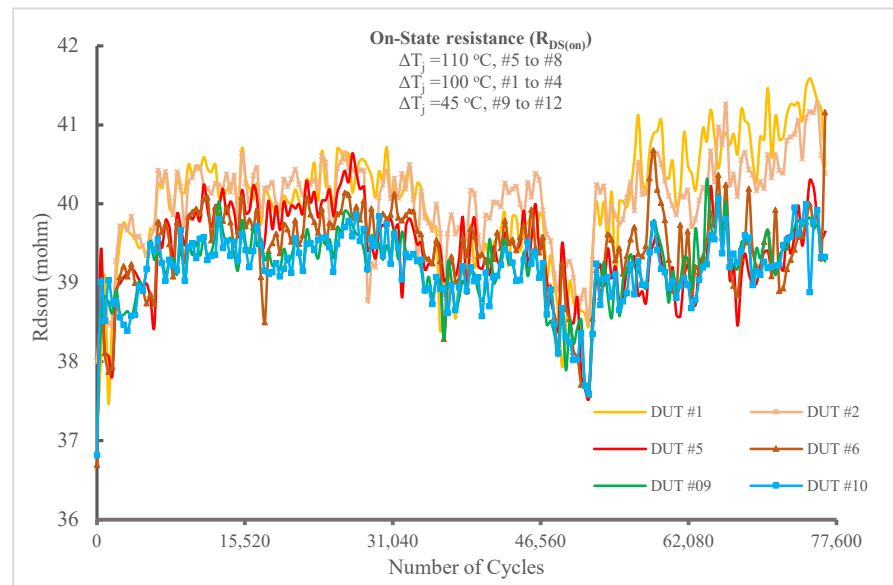
**Table 2.** Test conditions for the experiment.

Terms	Parameters	Test Conditions
Testing duration	Number of cycles/hours	77,600 cycles
Testing cycle	ON/OFF time	135 s (45 s on and 90 s off)
Interval of precursor data collection	On-state resistance ( $R_{ds(on)}$ )	every 400 cycles
	Threshold Voltage ( $V_{gs(th)}$ )	
	Body Diode Voltage ( $V_{sd}$ )	
	Drain Current ( $I_{dss}$ )	
Testing conditions (input electrical and thermal parameters)	Capacitance ( $C_{iss}, C_{oss}, C_{rss}$ )	Current: $\leq 49$ A rated current Voltage: 8 V for 4 MOSFETs for each scenario 14.2 V supplied for the PCB
	Scenario 1: $T_j = 40$ °C to 85 °C Scenario 2: $T_j = 25$ °C to 125 °C Scenario 3: $T_j = 25$ °C to 125 °C	
Temperature	Ambient	$T_c = 22 \pm 3$ °C

### 3.3. Failure Precursors Data Collection

The test samples are removed from the Powertester every 400 cycles and plugged into the curve tracer, as shown in Figure 6, for parameter measurement. A total of nine electrical parameters were measured during the power cycling-based accelerated degradation test including the on-state resistance ( $R_{ds(on)}$ ), threshold voltage ( $V_{th}$ ), body diode forward voltage ( $V_{sd}$ ), breakdown voltage ( $V_{br}$ ), drain current ( $I_{dss}$ ), drain-source on-state voltage ( $V_{dson}$ ), input, output, and reverse transfer capacitances ( $C_{iss}$ ,  $C_{oss}$ , and  $C_{rss}$ ), using a power device analyzer (Keysight B1505 A). A change in a failure precursor parameter can be attributed to various failure mechanisms that lead to a certain failure mode when they cross a specified failure threshold that varies based on a specific application. More details on the failure precursors of power MOSFETs will be reported in a separate study. The failure

precursor chosen for this study is the on-state resistance, and the results for test samples 1, 2, 5, 6, 9, and 10 are shown in Figure 7.



**Figure 7.** Variation in the actual on-state resistance ( $R_{DS(on)}$ ) value with respect to the power cycling time. The values passed through normalizations for analysis during the prediction process.

A specific failure precursor parameter can be used to develop failure models that can capture the degradation trend. For instance, the change in on-state resistance ( $R_{ds(on)}$ ) demonstrates the presence of die degradation and bond-wire lift-off, which could be caused by a high electric field and thermal runaway [1]. Some failure modes can be caused by a shift in one or multiple failure precursors, which makes the degradation modeling of power devices challenging. The failure precursors will show either an increasing, decreasing, or constant trend depending on the power-MOSFET response to dynamic thermal and electrical stresses.

It can be noticed that the on-state resistance values non-monotonically increase as the aging time increases, unlike other failure precursor parameters. This demonstrates the presence of die degradation and bond-wire lift-off, which could be caused by a high electric field and thermal runaway. This is potentially attributed to an increase in the drain-to-source voltage ( $V_{DS}$ ). The increasing trend of on-state resistance for aged power MOSFETs has also been reported in [8,30]. There is an obvious unit-to-unit variability among samples under the same scenario, which may arise from manufacturing imperfections. On the other hand, the degree of degradation shows that samples with higher temperature swings (110 °C and 100 °C) have shown a higher degree of on-state resistance variation compared with samples with a lower  $T_j$  (45 °C). From the comprehensive accelerated test, the on-state resistance data are found to be representative of the degradation pattern in power MOSFETs when exposed to long-term power cycling tests or applications. Based on this long-term time series precursor data, the LSTM and GRU neural network can be used to predict the future lifetime of power MOSFETs.

### 3.4. Data Preprocessing and Evaluation Metrics

In order to use the appropriate degradation data and apply the proposed algorithm to long-term lifetime prediction, data preparation and preprocessing should be conducted. The degradation data have passed through preliminary screening, the elimination of outliers (checking data with  $3\text{-}\sigma$  in a normal distribution), the removal of noise, as well as normalization. The probability plot is used to assess whether the failure precursor data can follow the normal distribution. Although it is not a requirement for the data to be within a  $\pm 3\sigma$  range, it will help the neural network learn and perform better.

In this study, the simple moving average filter (MAF) is used to avoid noise in the original data using Formula (15). In this equation,  $n$  is the total number of measured values,  $k$  is the window size,  $x_i$  is the original value, and  $x_k$  is the data after filtering. Using a value of  $k = 3$ , the filtered data used in the process of model training and prediction are plotted, as shown in Figure 8.

$$x_k = \frac{1}{k} \sum_{i=1-k+1}^n x_i \tag{15}$$

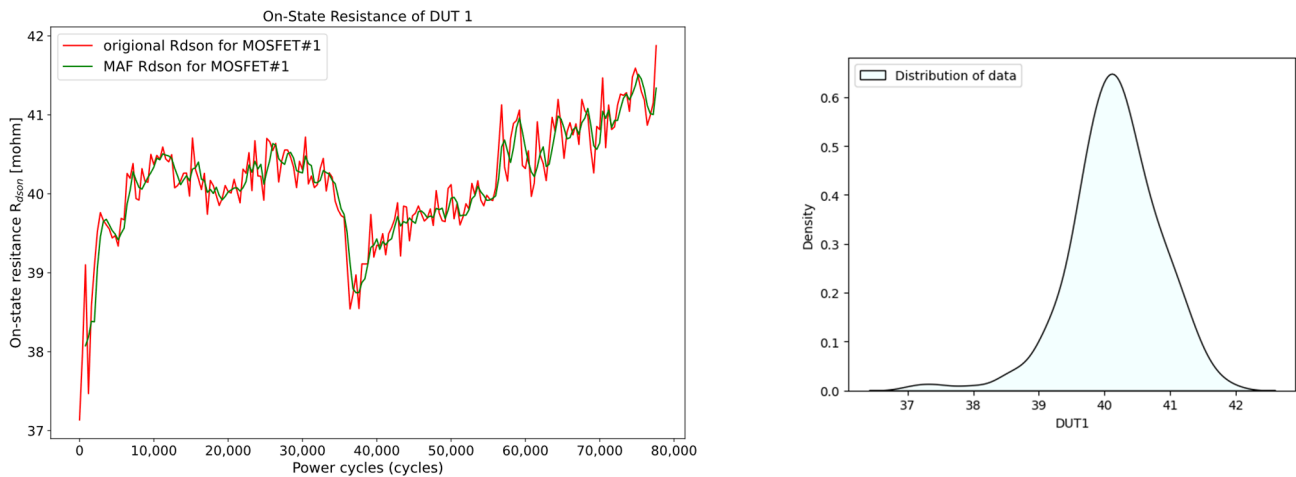


Figure 8. Comparative plot of original and MAF on-state resistance data.

In addition to noise removal, the input data have been normalized between the values of  $[0, 1]$  to make them suitable for the algorithm using the following equation.

$$x_N = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{16}$$

where  $x_N$  is the normalized value,  $x_{min}$  is the minimum value, and  $x_{max}$  is the maximum value.

The relative prediction performance of the proposed algorithms compared to other variants of deep learning methods can be compared quantitatively using three accuracy metrics or indexes including the mean average error/mean absolute percentage error (MAE/MAPE), mean squared error (MSE), and root mean squared error (RMSE), as given in Equations (17)–(19). The MAPE and RMSE are preferred metrics due to their ability to punish large errors with square roots [31].

$$MAE = \frac{1}{n} |e_i| = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{17}$$

The mean squared error (MSE) is given as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{18}$$

The root mean squared error (RMSE) is also described as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{19}$$

where  $y_i$  and  $\hat{y}_i$  denote real and predicted values, respectively, whereas  $e_i$  represents the error (the difference between  $y_i$  and  $\hat{y}_i$ ) values.

## 4. Results and Discussion

In this section, an analysis of the on-state resistance degradation data for three samples that represent different power cycling conditions is conducted to validate the proposed algorithm. All the tests are implemented with Python 3.9 on a laptop equipped with an AMD Ryzen 7-5800 H processor (16 MB cache, up to 4.4 GHz), 16 G DDR4 3200 memory, and an NVIDIA GeForce RTX 3050 4 GB graphic card. All the proposed algorithms in this paper including simple RNN, LSTM, as well as GRU have similar model arrangements, except for the replacement of each particular algorithm for computation and a fair comparison of the prediction results.

The proposed LSTM/GRU predictive algorithms are trained using the training set and optimized according to their performance on the validation set. The performance of each model at different prediction points was evaluated using only testing data that were not included in the training process to prevent information leakage. The one-step and multi-step prediction of long-term power MOSFET degradation data can be conducted. The one-step ahead prediction is mainly suitable for online failure prediction, as all-failure precursor measurements cannot be easily obtained while working. On the other hand, a prediction of a long-term degradation trend suggested that LSTMs and variants are suitable for reflecting on future degradation phases [32]. Thus, multi-step prediction is used in the proposed LSTM algorithm and compared with the traditional simple RNN and GRU methods.

### 4.1. Data Preprocessing, Parameter Setting, and Model Formation for the Models

The degradation data from power MOSFETs 1, 5, and 9 (selected one sample from each testing condition) are used to validate the proposed algorithm. To make a fair comparison of the different algorithms for validation, the model parameters are kept similar. The configurations of the RNN, LSTM, and GRU models considered are as follows: an Adam optimizer with a default learning rate = 0.001, two hidden layers with 128 and 64 neurons, one output layer, an Adam optimizer, a training loss that the model minimizes, which is a mean squared error, an epoch size = 100, and a batch size = 16, as shown in Table 3.

**Table 3.** Model Architecture and Summary.

Model	Number of Units	Optimizer	Training Loss Function	Dropout	Activation
RNN	(128, 64)	Adam	Mean Squared Error	20% (0.2)	relu
LSTM	(128, 64)	Adam	Mean Squared Error	20% (0.2)	relu
GRU	(128, 64)	Adam	Mean Squared Error	20% (0.2)	relu

### 4.2. Implementation of RNN/LSTM/GRU Models and Prediction Results

After setting up the model parameters, the model is trained with training data, and the future lifetime of power MOSFETs is predicted with three different prediction starting points (cycles). These prediction starting points considered are the on-state resistance data of the first 24,000 cycles (i.e., 30% of the full degradation path), 40,000 cycles (i.e., 50% of the full lifetime), as well as 57,600 cycles (i.e., 70% of the full lifetime). The training and testing partition of the degradation data is performed, as shown in Figure 9.

The long-term lifetime prediction of power MOSFETs with a starting point of 40,000 cycles is carried out first, followed by 24,000 cycles and, finally, 57,600 cycles. This procedure will be helpful in comparing the prediction performance of algorithms with multiple training and testing datasets. After training the proposed algorithms with 40,000 cycles of on-state resistance degradation data, the model losses showed that RNN has faster convergence compared to LSTM and GRU, whereas the training and testing losses are smaller for the advanced model at the end of model training, as depicted in Figure 10.

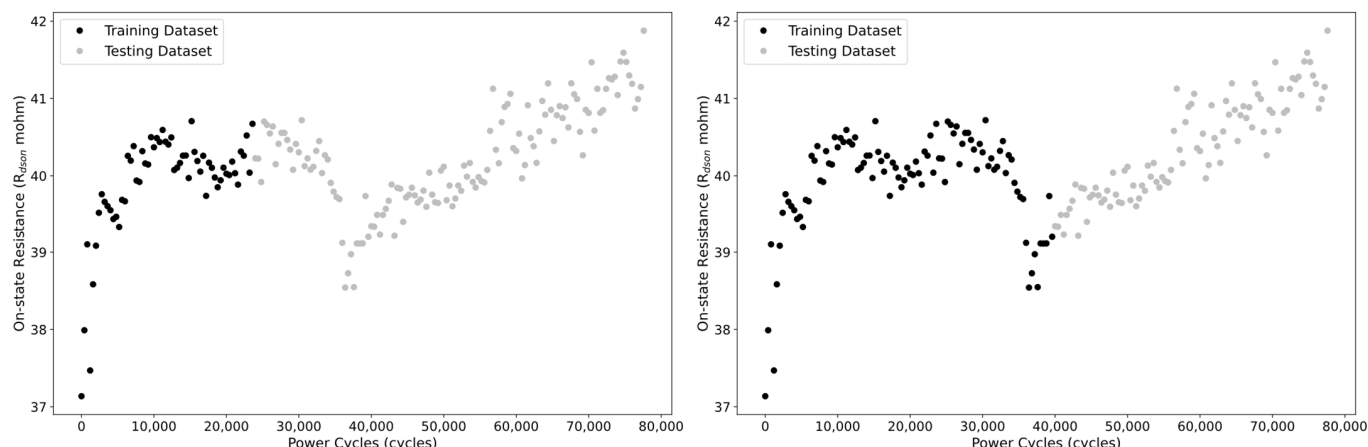


Figure 9. Training and Testing Data Partition for Algorithm Training (30% and 50% as the sample).

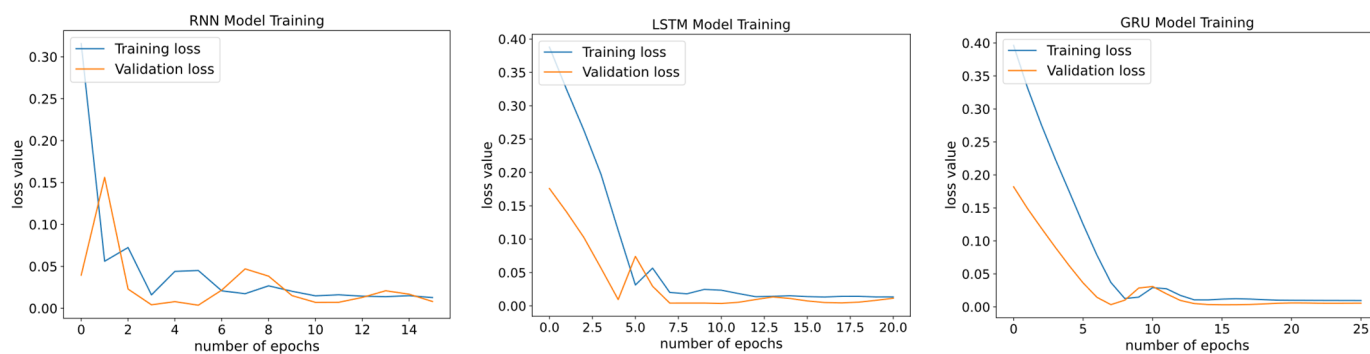


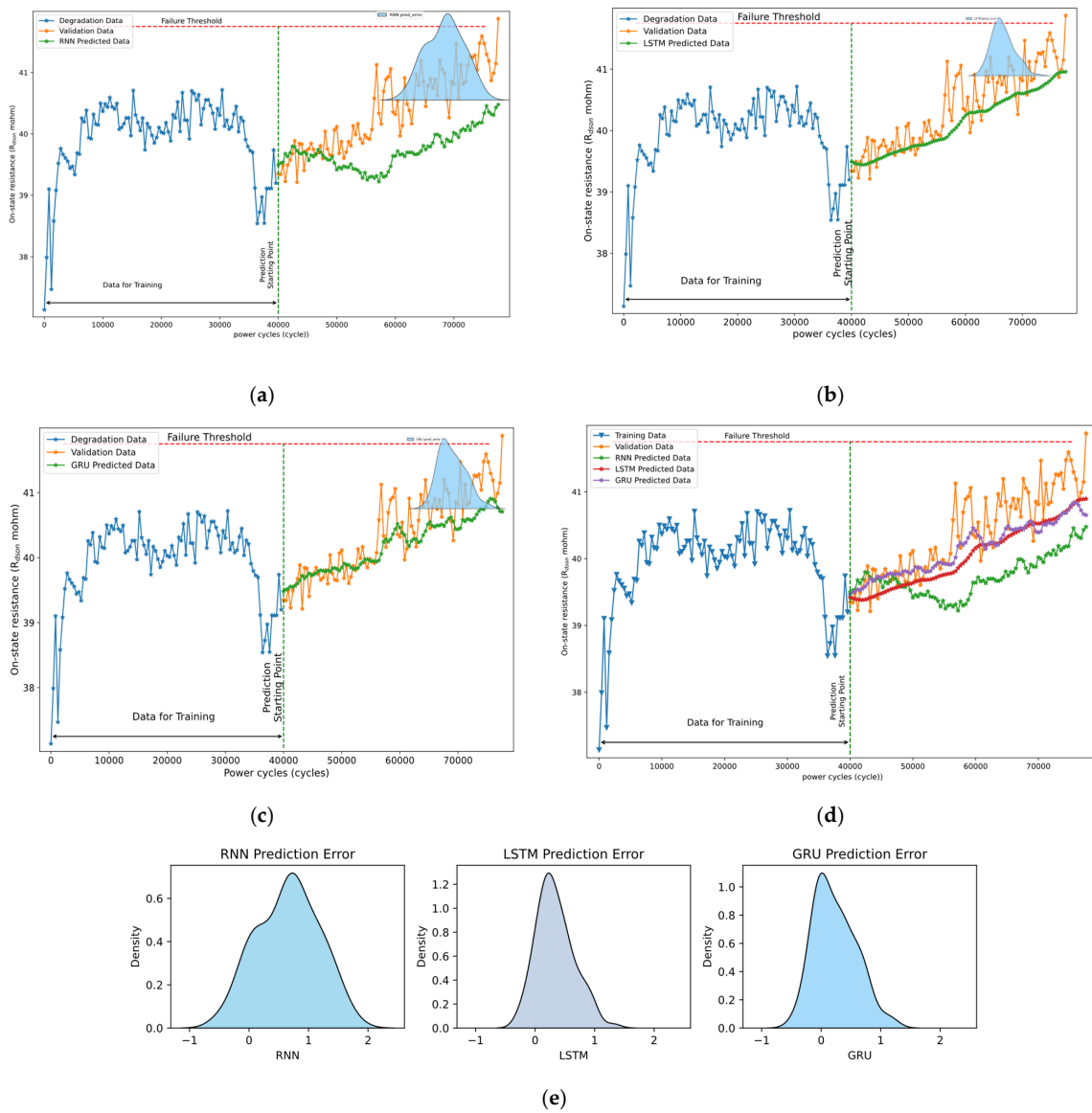
Figure 10. Training and testing losses of RNN, LSTM, and GRU for MOSFET 1.

The multi-step ahead prediction results based on the on-state resistance degradation data for power MOSFET 1 at a starting point of 40,000 cycles (50% training and 50% testing split) using the RNN, LSTM, and GRU algorithms are carried out, as shown in Figure 11.

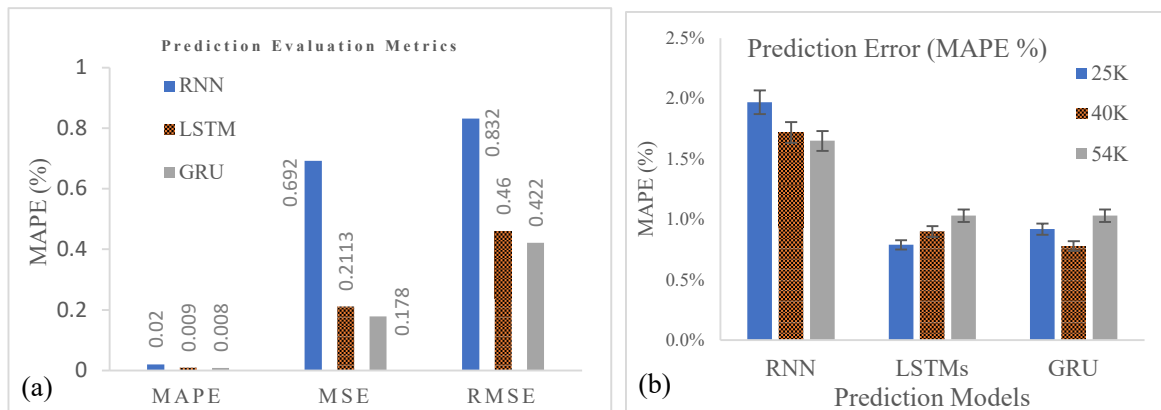
It can be noted that the prediction accuracy of the LSTM and GRU is superior compared to that of the simple RNN algorithm, with an MAPE of 0.9%, 0.78%, and 1.72%, respectively. The prediction metrics results at 40,000 cycles, where the first 50% of the data are used in model training, and at 24,000 as well as 54,000 cycles of training data are shown in Figure 12. It can be noted that the LSTM and GRU performed better for the prediction compared to the RNN as the results of MAPE manifested. Similarly, the distribution of the prediction error from the estimated and testing data is plotted in Figure 11e and confirms the prediction metrics results, with a wide base distribution error for RNN and a narrow distribution for LSTM and GRU.

The difference in the performance of the proposed prognostics algorithms for MOSFET 1 can be easily observed by visualizing the MAPE, MSE, and RMSE of estimates at the different starting points of measured data, as shown in Table 4.

In addition, the long-term lifetime prediction of the MOSFET 5 and 9 test samples from two different conditions of junction temperature swing is performed at a 40,000 starting point. The prediction results along with the distribution of the model prediction error are shown in Figures 13a–d and 14a–d. In the prediction process, the training loss for RNN converged faster compared to that of LSTM and GRU, which may be attributed to its simplified cell architecture, as shown in Figures 15 and 16. Similarly, the on-state resistance prediction error estimated from 40,000 cycles of the starting point of the training and testing data values using the RNN, LSTM, and GRU prediction models is shown in Figure 13e.

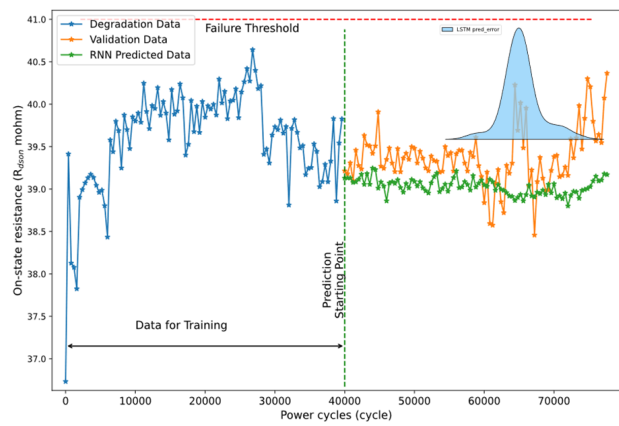


**Figure 11.** Rds(on) prediction result for MOSFET 1 with (a) RNN, (b) LSTM, and (c) GRU; (d) comparison of the three models; (e) prediction error using RNN/LSTM/GRU.

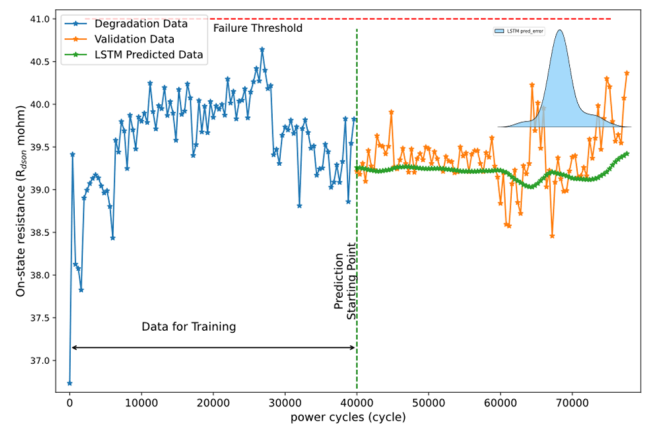


**Figure 12.** Comparison of Prediction Results (a) at 40,000 cycles and (b) at 24,000, 40,000, and 54,000 cycles. Starting Points for MOSFET 1.

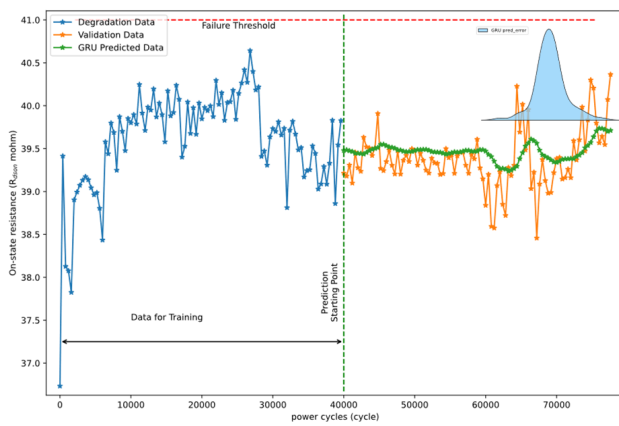




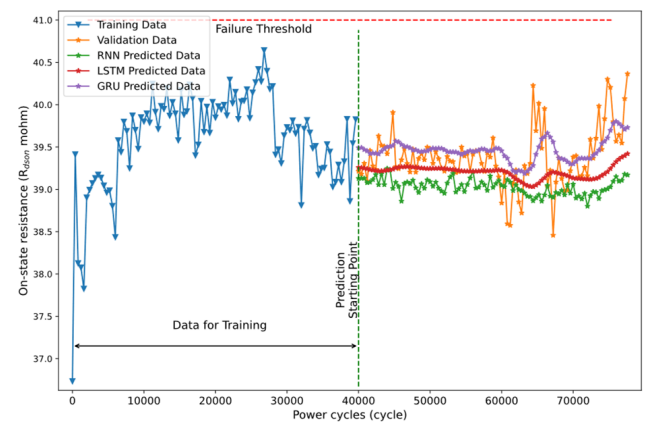
(a)



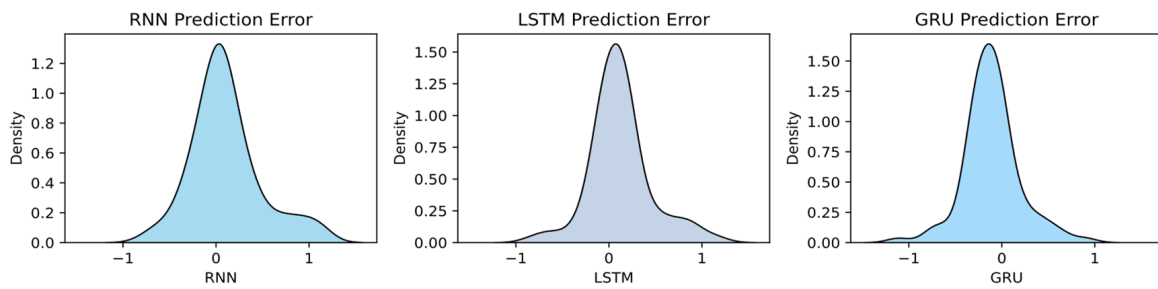
(b)



(c)



(d)



(e)

**Figure 13.** Rds(on) prediction result for MOSFET 5 using (a) RNN, (b) LSTM, and (c) GRU; (d) comparison of the three models; (e) prediction error (AE) using RNN, LSTM, and GRU.

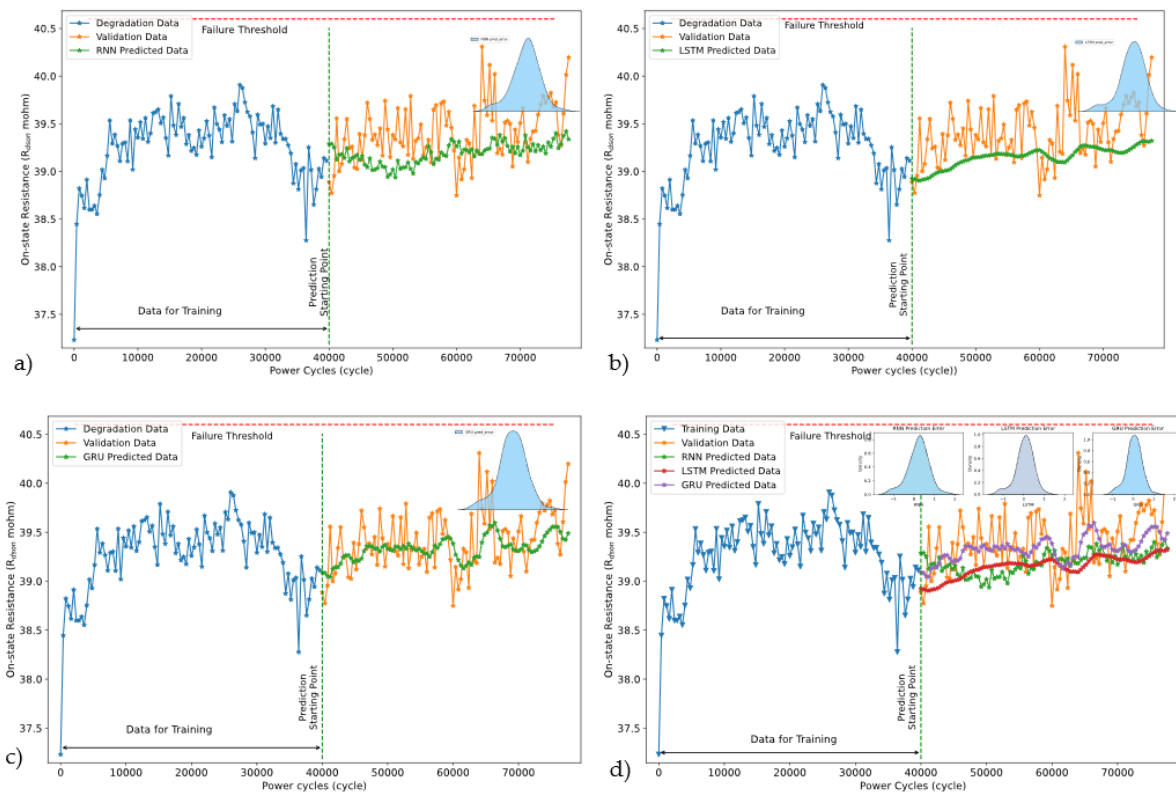


Figure 14. Rds(on) prediction result for MOSFET 9 using (a) RNN, (b) LSTM, and (c) GRU; (d) comparison of the three models with the prediction error.

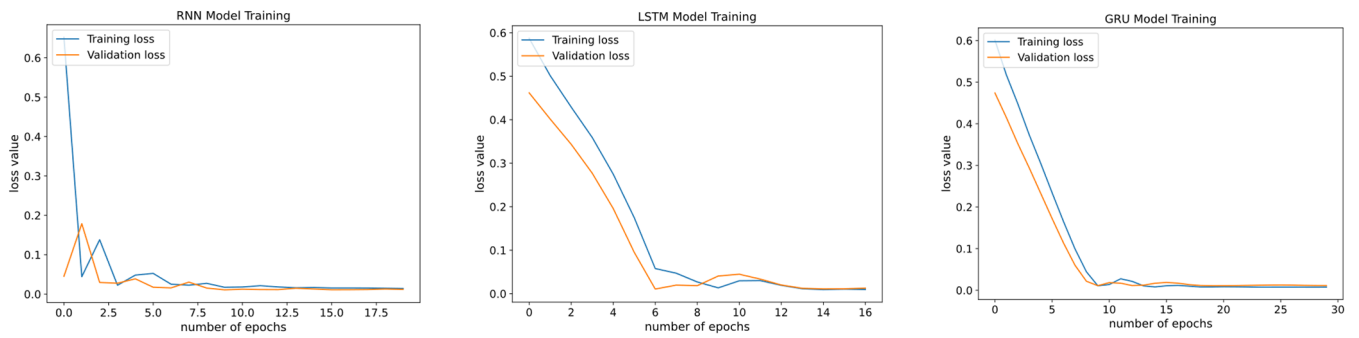


Figure 15. Training and testing losses of RNN, LSTM, and GRU for MOSFET 5.

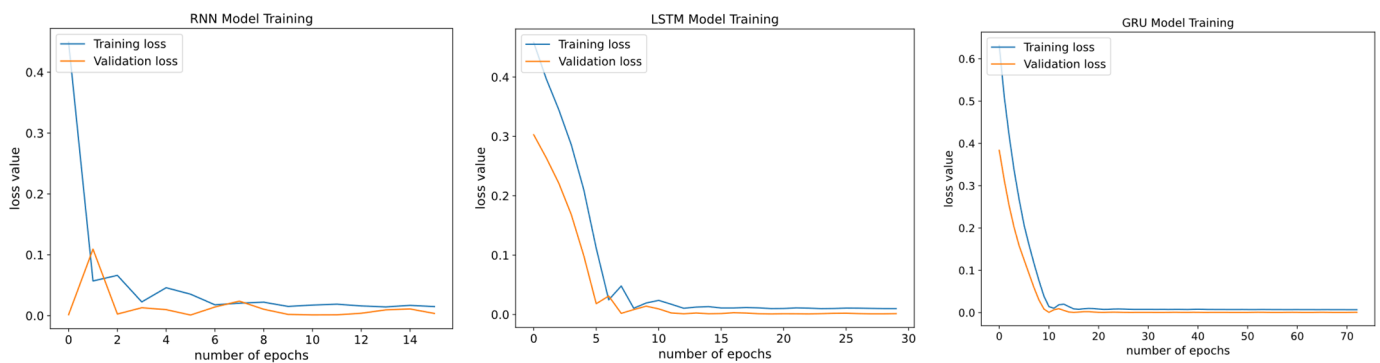


Figure 16. Model training and testing losses of the RNN, LSTM, and GRU models for MOSFET 9.

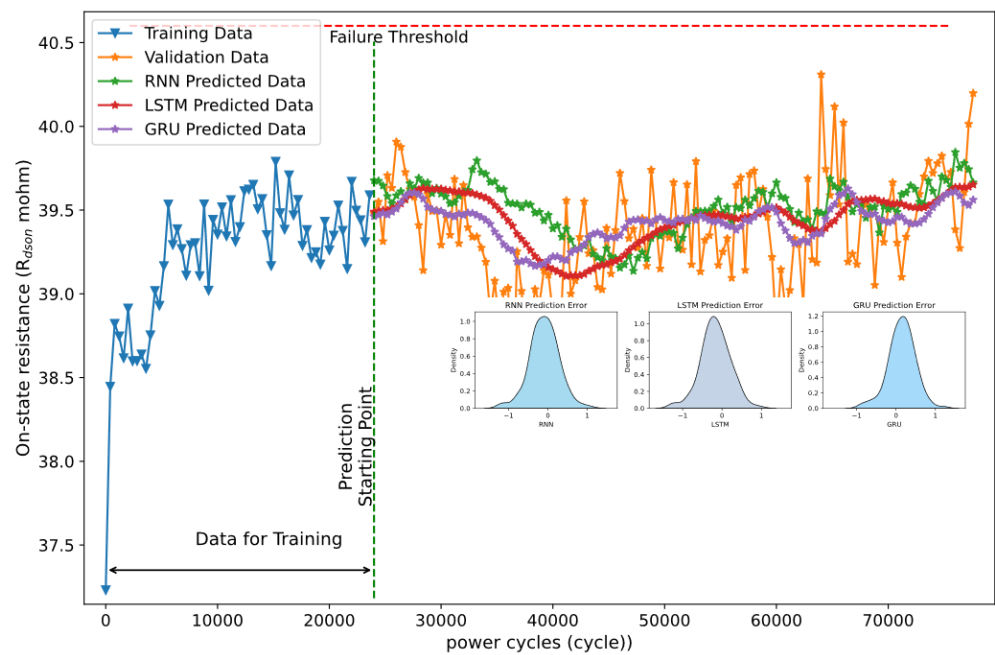
**Table 4.** Prognostic Performance (Prediction Error) Metrics Summary for MOSFET 1.

Model	Starting Points	MAPE	MSE	RMSE
RNN	24,000	0.0197	0.245	0.495
	40,000	0.0172	0.699	0.832
	57,600	0.0165	0.746	0.864
LSTMs	24,000	0.0079	0.203	0.451
	40,000	0.009	0.366	0.459
	57,600	0.0103	0.264	0.514
GRU	24,000	0.0092	0.236	0.485
	40,000	0.0078	0.318	0.422
	57,600	0.0103	0.268	0.518

4.3. Discussion Based on Lifetime Prediction Metrics and Model Robustness

As the prediction of long-term lifetime is conducted using the proposed method on degradation data, the next logical procedure is to evaluate the model prediction error. The model prediction error for such a regression type of the problem is mainly assessed using MAPE, MSE, as well as RMSE. The robustness of the proposed method is tested by using different proportions of training data or measurement cycles of failure precursors. Here, the long-term lifetime of two test samples (MOSFET 9 and 5) is explored at 24,000 and 54,000 cycles of starting points with 30% and 70% training data and 70% and 30% testing data, respectively, to demonstrate the robustness of the proposed algorithms.

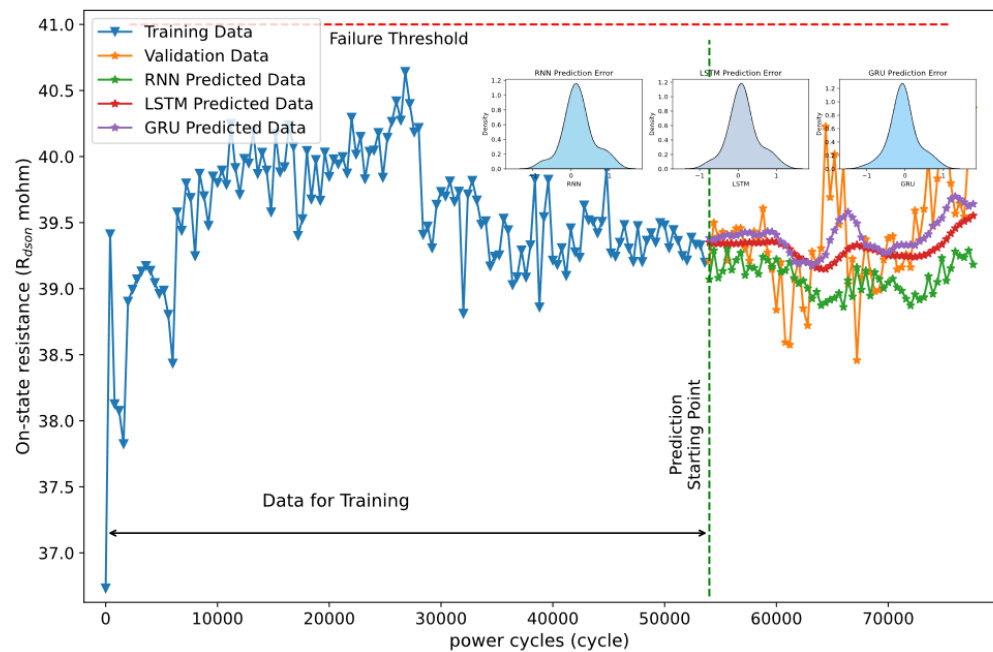
Figure 17 presents the prediction plots and distribution of prediction error at 24,000 cycles of a starting point for MOSFET 9 degradation data. It can be noted that the prediction accuracy has decreased for the three models, as the algorithms used less training data. On the other hand, the RNN model is less affected by the smaller dataset, as its capability is also limited by shorter memories compared to the LSTM and GRU algorithms.



**Figure 17.** Long-term lifetime prediction for MOSFET 9 at 24,000 cycles of starting points with three models and prediction error plots.

Similarly, the prediction plots at 54,000 cycles of the starting point and the distribution of prediction errors for MOSFET 5 degradation data are presented in Figure 18. The results

showed that GRU and LSTM outperformed the simple RNN model. In addition, the overall prediction accuracy of the three models increased as more training on-state resistance degradation data were used, as compared to the fewer training data of 24,000 (30%) and 40,000 (50%) cycles.



**Figure 18.** Long-term lifetime prediction for MOSFET 5 at 54,000 cycles of the starting point. (a) Prediction plots of the three models; (b) distribution of prediction error.

Although the prediction plots are displayed for the randomly selected test samples 1, 5, and 9 due to space limitations, all these prediction results are given as depicted in Tables 4–7. Based on the long-term lifetime prediction, the model performance metrics results with MAPE, MSE, and RMSE for MOSFETs 1, 5, and 9 at 24,000 cycles of (30% of a full lifetime) starting points are shown in Table 5.

**Table 5.** Prediction performance of three models at 24,000 cycles of starting points.

Test Samples	Indices	RNN	LSTM	GRU
MOSFET #1	MAPE	0.0197	0.0079	0.0092
	MSE	0.2452	0.2031	0.2356
	RMSE	0.4952	0.451	0.4854
MOSFET #5	MAPE	0.0090	0.0066	0.0073
	MSE	0.1948	0.1271	0.1309
	RMSE	0.4413	0.3565	0.3618
MOSFET #9	MAPE	0.0075	0.0083	0.0074
	MSE	0.1462	0.1681	0.1363
	RMSE	0.3824	0.4099	0.3692

As shown in Table 6, the MAPE prediction error metrics at 40,000 cycles of the starting point are 1.72%, 0.90%, and 0.78% for MOSFET 1; 0.94%, 0.6%, and 0.6% for MOSFET 5; and 1.05%, 0.91%, and 0.78% for MOSFET 9 using the RNN-, LSTM-, and GRU-based models, respectively. Similarly, the MSE and RMSE results for the model performance metrics are given in the same chart. These results showed that the LSTM- and GRU-based prognostic models performed better compared to the simple RNN-based model. In addition, the

prediction error decreased compared to the previous scenario, where training data of only 24,000 cycles of starting points was used.

**Table 6.** Prediction performance of three models at 40,000 cycles of starting points.

Test Sample	Indices	RNN	LSTM	GRU
MOSFET #1	MAPE	0.0172	0.009	0.0078
	MSE	0.6992	0.3664	0.3178
	RMSE	0.832	0.459	0.422
MOSFET #5	MAPE	0.0090	0.0066	0.0060
	MSE	0.2250	0.1174	0.1017
	RMSE	0.4743	0.3426	0.3189
MOSFET #9	MAPE	0.0105	0.0091	0.0078
	MSE	0.2831	0.2048	0.1569
	RMSE	0.5321	0.4525	0.3961

**Table 7.** Prediction performance of three models at 54,000 cycles of starting points.

Test Sample	Indices	RNN	LSTM	GRU
MOSFET #1	MAPE	0.0165	0.0103	0.0103
	MSE	0.7458	0.2641	0.2678
	RMSE	0.8636	0.5139	0.5175
MOSFET #5	MAPE	0.0103	0.0074	0.0065
	MSE	0.1811	0.1697	0.245
	RMSE	0.4256	0.4119	0.3528
MOSFET #9	MAPE	0.0107	0.0089	0.0065
	MSE	0.2706	0.1952	0.1035
	RMSE	0.5202	0.443	0.3217

Lastly, the MAPE, MSE, and RMSE prediction error metrics at 54,000 cycles of starting points for MOSFET 1, 5, and 9 are given in Table 7. The long-term lifetime prediction results for MOSFET 9 show MAPE results of 1.07%, 0.89%, and 0.65% with RNN, LSTM, and GRU, respectively. These prediction results show that the model’s prediction accuracy increases as it receives more training data compared to 24,000 and 4000 cycles of starting points, and GRU- and LSTM-based models predict better compared to the RNN model.

It is worth noting that the prediction errors with more training data (such as 70%) are smaller and closer to the actual or measured value compared with the models using less training (such as 30%) data, as shown in Tables 5 and 7, respectively. This is interesting, as the prediction uncertainty increases with an increase in the long-term prediction curve and less training data. Overall, the prediction metrics values of the proposed LSTM and GRU methods showed accurate and precise long-term estimation, which shows a reliable multistep-ahead prediction for power MOSFET degradation precursor parameters.

In general, the long-term lifetime prognostics results showed that the proposed algorithms are suitable for dealing with failure precursor degradation analysis problems for power MOSFETs. The LSTM and GRU performed better compared to the simple RNN model for long-term lifetime predictions. It is also worth noting that the convergence speed of RNN and GRU is faster than that of LSTM in model training, which is attributed to the simpler internal cell structure of RNN followed by GRU relative to LSTM networks.

## 5. Conclusions

In this study, data-driven deep learning algorithms based on LSTM and GRU are used to predict the future degradation pattern and, hence, the lifetime of power MOSFET devices. As one of the dominant performance parameters, the on-state resistance failure precursor data of these devices are considered in the implementation of the proposed algorithm. To demonstrate the proposed LSTM and GRU models, the on-state resistance data from an accelerated degradation test based on power cycling were collected at different junction temperature swings of 45 °C, 100 °C, and 110 °C. The adaptive moment estimation (Adam) optimizer is used to update network weights. Dropout regularization (0.2) is employed to prevent overfitting, and a learning rate of 0.0001 is set during data training for the constructed neural network models.

The accuracy of prognostic models based on the RNN, LSTM, and GRU algorithms is evaluated using the MAPE, MSE, and RMSE prediction metrics. The prediction results based on the proposed LSTM and GRU showed an accurate and precise lifetime prediction compared to the classic RNN algorithm. It is also worth noting that the convergence speed of RNN and GRU is faster than that of LSTM in model training, which is attributed to the simpler cell structures. The robustness of the proposed approaches is verified by using 30% and 70% of the measured data for model training in addition to the 50% training and testing setup, which shows the adaptability of the model for power device degradation trends. In general, the LSTM and GRU models are found to be effective for degradation assessment and long-term lifetime predictions for power devices based on failure precursor data. With an online data acquisition system, prognostic models can be employed in the condition monitoring of power devices.

**Author Contributions:** Conceptualization and methodology, M.S.I., K.-H.L., H.H.L. and J.F.; data curation, C.L., M.W. and W.A.; formal analysis and writing—original draft preparation, M.S.I.; visualization, C.L., W.A. and M.W.; writing—review and editing, W.A., M.W., H.H.L., C.L. and J.F.; funding acquisition, supervision, and project administration, H.H.L. and K.-H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work described in this paper is supported by the Centre for Advances in Reliability and Safety (CAiRS), admitted under the AIR@InnoHK Research Cluster and partially supported by the National Natural Science Foundation of China (52275559), and the Shanghai Pujiang Program (2021PJD002).

**Data Availability Statement:** The dataset presented in this study is available upon request from the corresponding author.

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

## References

1. Hanif, A.; Yu, Y.; DeVoto, D.; Khan, F. A Comprehensive Review Toward the State-of-the-Art in Failure and Lifetime Predictions of Power Electronic Devices. *IEEE Trans. Power Electron.* **2019**, *34*, 4729–4746. [[CrossRef](#)]
2. Yang, S.; Xiang, D.; Bryant, A.; Mawby, P.; Ran, L.; Tavner, P. Condition Monitoring for Device Reliability in Power Electronic Converters: A Review. *IEEE Trans. Power Electron.* **2010**, *25*, 2734–2752. [[CrossRef](#)]
3. Goudarzi, A.; Ghayoor, F.; Waseem, M.; Fahad, S.; Traore, I. A Survey on IoT-Enabled Smart Grids: Emerging, Applications, Challenges, and Outlook. *Energies* **2022**, *15*, 6984. [[CrossRef](#)]
4. Fahad, S.; Goudarzi, A.; Li, Y.; Xiang, J. A coordination control strategy for power quality enhancement of an active distribution network. *Energy Rep.* **2022**, *8*, 5455–5471. [[CrossRef](#)]
5. Ni, Z.; Lyu, X.; Yadav, O.P.; Singh, B.N.; Zheng, S.; Cao, D. Overview of Real-Time Lifetime Prediction and Extension for SiC Power Converters. *IEEE Trans. Power Electron.* **2020**, *35*, 7765–7794. [[CrossRef](#)]
6. Goudarzi, A.; Davidson, I.E.; Ahmadi, A.; Venayagamoorthy, G.K. Intelligent analysis of wind turbine power curve models. In Proceedings of the 2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG), Orlando, FL, USA, 9–12 December 2014; pp. 1–7.
7. Pu, S.; Yang, F.; Vankayalapati, B.T.; Akin, B. Aging Mechanisms and Accelerated Lifetime Tests for SiC MOSFETs: An Overview. *IEEE J. Emerg. Sel. Top. Power Electron.* **2022**, *10*, 1232–1254. [[CrossRef](#)]
8. Dusmez, S.; Ali, S.H.; Heydarzadeh, M.; Kamath, A.S.; Duran, H.; Akin, B. Aging Precursor Identification and Lifetime Estimation for Thermally Aged Discrete Package Silicon Power Switches. *IEEE Trans. Ind. Appl.* **2017**, *53*, 251–260. [[CrossRef](#)]

9. Patil, N.; Das, D.; Goebel, K.; Pecht, M. Failure Precursors for Insulated Gate Bipolar Transistors (IGBTs). In Proceedings of the 9th International Seminar on Power Semiconductors (ISPS 2008), Prague, Czech Republic, 27–29 August 2008.
10. Song, S.; Munk-Nielsen, S.; Uhrenfeldt, C.; Trintis, I. Failure mechanism analysis of a discrete 650V enhancement mode GaN-on-Si power device with reverse conduction accelerated power cycling test. In Proceedings of the 2017 IEEE Applied Power Electronics Conference and Exposition (APEC), Tampa, FL, USA, 26–30 March 2017; pp. 756–760.
11. Guran, I.-C.; Florescu, A.; Perișoară, L.A. A Novel ON-State Resistance Modeling Technique for MOSFET Power Switches. *Mathematics* **2023**, *11*, 72. [[CrossRef](#)]
12. Ibrahim, M.S.; Jing, Z.; Yung, W.K.; Fan, J. Bayesian based lifetime prediction for high-power white LEDs. *Expert Syst. Appl.* **2021**, *185*, 115627. [[CrossRef](#)]
13. Edrisian, A.; Goudarzi, A.; Ebadian, M.; Swanson, A.G.; Mahdian, D. Assessing the effective parameters on operation improvement of SCIG based wind farms connected to network. *Int. J. Renew. Energy Res.* **2016**, *6*, 585–592.
14. Nguyen, M.H.; Kwak, S. Enhance reliability of semiconductor devices in power converters. *Electronics* **2020**, *9*, 2068. [[CrossRef](#)]
15. Dusmez, S.; Duran, H.; Akin, B. Remaining Useful Lifetime Estimation for Thermally Stressed Power MOSFETs Based on on-State Resistance Variation. *IEEE Trans. Ind. Appl.* **2016**, *52*, 2554–2563. [[CrossRef](#)]
16. Wu, L.; Yang, J.; Peng, Z.; Wang, H. Remaining useful life prognostic of power metal oxide semiconductor field effect transistor based on improved particle filter algorithm. *Adv. Mech. Eng.* **2017**, *9*, 1687814017749324. [[CrossRef](#)]
17. Zhao, S.; Wang, H. Enabling Data-Driven Condition Monitoring of Power Electronic Systems With Artificial Intelligence: Concepts, Tools, and Developments. *IEEE Power Electron. Mag.* **2021**, *8*, 18–27. [[CrossRef](#)]
18. Pugalenth, K.; Park, H.; Raghavan, N. Prognosis of power MOSFET resistance degradation trend using artificial neural network approach. *Microelectron. Reliab.* **2019**, *100*, 113467. [[CrossRef](#)]
19. Zhao, S.; Chen, S.; Yang, F.; Ugur, E.; Akin, B.; Wang, H. A Composite Failure Precursor for Condition Monitoring and Remaining Useful Life Prediction of Discrete Power Devices. *IEEE Trans. Ind. Inform.* **2021**, *17*, 688–698. [[CrossRef](#)]
20. Lipton, Z.C.; Berkowitz, J.; Elkan, C. A critical review of recurrent neural networks for sequence learning. *arXiv* **2015**, arXiv:1506.00019.
21. Staudemeyer, R.C.; Morris, E.R. Understanding LSTM—a tutorial into long short-term memory recurrent neural networks. *arXiv* **2019**, arXiv:1909.09586.
22. Rezaeianjouybari, B.; Shang, Y. Deep learning for prognostics and health management: State of the art, challenges, and opportunities. *Measurement* **2020**, *163*, 107929. [[CrossRef](#)]
23. Jing, Z.; Liu, J.; Ibrahim, M.S.; Fan, J.; Fan, X.; Zhang, G. Lifetime prediction of ultraviolet light-emitting diodes using a long short-term memory recurrent neural network. *IEEE Electron. Device Lett.* **2020**, *41*, 1817–1820. [[CrossRef](#)]
24. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
25. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
26. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
27. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
28. Zhang, Y.; Xiong, R.; He, H.; Pecht, M.G. Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries. *IEEE Trans. Veh. Technol.* **2018**, *67*, 5695–5705. [[CrossRef](#)]
29. Baharani, M.; Biglarbegian, M.; Parkhideh, B.; Tabkhi, H. Real-Time Deep Learning at the Edge for Scalable Reliability Modeling of Si-MOSFET Power Electronics Converters. *IEEE Internet Things J.* **2019**, *6*, 7375–7385. [[CrossRef](#)]
30. Celaya, J.R.; Saxena, A.; Saha, S.; Goebel, K.F. Prognostics of power MOSFETs under thermal stress accelerated aging using data-driven and model-based methodologies. In Proceedings of the Annual Conference of the PHM Society, Montreal, QC, Canada, 25–29 September 2011.
31. Ochella, S.; Shafiee, M. Performance metrics for artificial intelligence (AI) algorithms adopted in prognostics and health management (PHM) of mechanical systems. In Proceedings of the 2020 International Symposium on Automation, Information and Computing (ISAIC 2020), Beijing, China, 2–4 December 2020; Volume 1828, p. 012005.
32. Chen, Z.; Xia, T.; Li, Y.; Pan, E. A hybrid prognostic method based on gated recurrent unit network and an adaptive Wiener process model considering measurement errors. *Mech. Syst. Signal Process.* **2021**, *158*, 107785. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.