



## OPEN Prediction of power conversion efficiency parameter of inverted organic solar cells using artificial intelligence techniques

Mustapha Marzouglal<sup>1</sup>, Abdelkerim Souahlia<sup>2</sup>, Lakhdar Bessissa<sup>3</sup>, Djillali Mahi<sup>1</sup>, Abdelaziz Rabehi<sup>2</sup>✉, Yahya Z. Alharthi<sup>4</sup>, Amanuel Kumsa Bojer<sup>5</sup>✉, Aymen Flah<sup>6,7,8,9</sup>, Mosleh M. Alharthi<sup>10</sup> & Sherif S. M. Ghoneim<sup>10</sup>

Organic photovoltaic (OPV) cells are at the forefront of sustainable energy generation due to their lightness, flexibility, and low production costs. These characteristics make OPVs a promising solution for achieving sustainable development goals. However, predicting their lifetime remains a challenging task due to complex interactions between internal factors such as material degradation, interface stability, and morphological changes, and external factors like environmental conditions, mechanical stress, and encapsulation quality. In this study, we propose a machine learning-based technique to predict the degradation over time of OPVs. Specifically, we employ multi-layer perceptron (MLP) and long short-term memory (LSTM) neural networks to predict the power conversion efficiency (PCE) of inverted organic solar cells (iOSCs) made from the blend PTB7-Th:PC<sub>70</sub>BM, with PFN as the electron transport layer (ETL), fabricated under an N<sub>2</sub> environment. We evaluate the performance of the proposed technique using several statistical metrics, including mean squared error (MSE), root mean squared error (rMSE), relative squared error (RSE), relative absolute error (RAE), and the correlation coefficient (R). The results demonstrate the high accuracy of our proposed technique, evidenced by the minimal error between predicted and experimentally measured PCE values: 0.0325 for RSE, 0.0729 for RAE, 0.2223 for rMSE, and 0.0541 for MSE using the LSTM model. These findings highlight the potential of proposed models in accurately predicting the performance of OPVs, thus contributing to the advancement of sustainable energy technologies.

**Keywords** Power conversion efficiency, Inverted organic solar cells, Prediction, Multi-layer perceptron, Long short-term memory, Machine learning, Degradation

The overconsumption of traditional energy resources is leading to their depletion, which in turn contributes to increased pollution and global warming. Consequently, the depletion of these resources and the resulting pollution have created a growing demand for renewable energy sources<sup>1,2</sup>. Organic photovoltaic (OPV) cells have gained significant attention in recent years due to their potential for producing cost-effective, lightweight, and flexible solar energy solutions. Despite achieving power conversion efficiencies (PCE) exceeding 18% in laboratory settings<sup>3,4</sup>, challenges such as stability and long-term degradation continue to hinder their widespread adoption. Among the various alternatives, organic solar cells stand out as a promising option due to their affordability and flexibility. In particular, inverted organic solar cells (iOSCs) are of special interest because of

<sup>1</sup>Laboratory of Studies and Development of Semiconductor and Dielectric Materials, LeDMaScD, University Amar Telidji of Laghouat, BP 37G Route of Ghardaïa, Laghouat 03000, Algeria. <sup>2</sup>Telecommunication and Smart Systems Laboratory, Faculty of Sciences and Technology, Ziane Achour University, Djelfa, Algeria. <sup>3</sup>Materials Science and Informatics Laboratory, MSIL, Ziane Achour University of Djelfa, Road Moudjbara, PO Box 3117, Djelfa 17000, Algeria. <sup>4</sup>Department of Electrical Engineering, College of Engineering, University of Hafr Albatin, Hafr Al Batin 39524, Saudi Arabia. <sup>5</sup>Ethiopian Artificial Intelligence Institute, PO Box 40782, Addis Ababa, Ethiopia. <sup>6</sup>Processes, Energy, Environment, and Electrical Systems, National Engineering School of Gabès, University of Gabès, Gabes, Tunisia. <sup>7</sup>College of Engineering, University of Business and Technology (UBT), 21448 Jeddah, Saudi Arabia. <sup>8</sup>Applied Science Research Center, Applied Science Private University, Amman 11931, Jordan. <sup>9</sup>ENET Centre, VSB-Technical University of Ostrava, Ostrava, Czech Republic. <sup>10</sup>Department of Electrical Engineering, College of Engineering, Taif University, Taif 21944, Saudi Arabia. ✉email: abdelaziz.rabehi@univ-djelfa.dz; amanuel.kumsa@aau.edu.et

their enhanced stability, especially when fabricated using materials like PTB7-Th as the active layer. However, the PCE of iOSCs tends to degrade over time, underscoring the need for accurate predictive models to forecast their performance over extended periods<sup>5,6</sup>.

As a third-generation photovoltaic technology, organic solar cells utilize an active layer composed of organic molecules with a  $\pi$  conjugated structure. When exposed to sunlight, this layer generates excitons (electron–hole pairs), which are subsequently separated into free electrons and holes, creating an electrical current<sup>7,8</sup>. Although the PCE — defined as the ratio of electrical power delivered by a solar cell to the incident solar energy — of organic solar cells currently lags behind that of inorganic cells, researchers have made significant progress since the discovery of plastic solar cells in 1986<sup>8</sup> and<sup>6</sup>. Various structural designs and materials are being explored to enhance their efficiency.

The problem of degradation in inverted solar cells is a major challenge that affects their long-term performance<sup>10</sup>. Exposure to environmental factors such as humidity and high temperatures leads to the deterioration of materials used in these cells, resulting in a decline in their efficiency<sup>11</sup>. Therefore, predicting the degradation of these cells is crucial, as it helps identify weaknesses and improve cell design, contributing to enhanced performance and sustainability<sup>12</sup>. When discussing the PCE of solar cells, it is considered one of the most important characteristics in evaluating the performance of these cells<sup>13</sup>. Predicting PCE is vital for the development and optimization of this technology<sup>14</sup>. As organic solar cells develop, predicting their PCE is essential for comparing them with inorganic cells and tracking their progress. Unlike parameters such as short-circuit current ( $J_{sc}$ ) and open-circuit voltage ( $V_{oc}$ ), which require extensive experimental data for accurate modeling, PCE can guide improvement efforts more directly, saving time and resources. PCE is calculated using a formula that includes  $J_{sc}$ ,  $V_{oc}$ , fill factor ( $FF$ ), and incident light power<sup>8</sup>. Traditional methods for predicting the performance of organic solar cells involve evaluating the electrical and structural characteristics of the devices and simulating their current density–voltage ( $J$ - $V$ ) curves<sup>15</sup>. While simple equivalent circuit models can aid in comparative analysis, they do not offer detailed insights into the nuanced performance characteristics of the devices<sup>16</sup>. These traditional methods are also often time-consuming and costly, as they require specific parameters to achieve accurate predictions<sup>5</sup>. Although analytical models and more complex simulations provide a deeper understanding of charge transport mechanisms, they are challenging to implement due to the difficulties in obtaining the necessary parameters<sup>17</sup>.

Given the remarkable success of Machine Learning (ML) techniques in various engineering fields— such as medical imaging<sup>18</sup>, solar energy prediction<sup>19</sup>, traffic flow prediction<sup>20</sup>, financial market predictions<sup>21</sup>, energy demand forecasting<sup>22</sup>, and solar cell efficiency prediction<sup>23</sup>—we aimed to address these challenges by initially leveraging the conventional double exponential model, commonly used to simulate the degradation of organic solar cells<sup>24,25</sup>. Specifically, we proposed optimizing the double exponential model using the Levenberg–Marquardt algorithm, rather than the Nonlinear Least Squares optimization typically used in the literature. Additionally, we introduced four ML-based techniques, namely LSTM, MLP, Support Vector Machine (SVM) and Random Forest (RF), for monitoring the degradation of iOSCs through the prediction of the PCE. Our goal is to predict the quality of iOSCs by selecting certain experimental points and attempting to predict the remaining points, thus enhancing the efficiency and effectiveness of the organic solar cell technology. The results show the high accuracy of our proposed technique, as evidenced by the calculation of the error between the predicted and experimentally measured PCE values.

## Methodology

Many researchers focus on studying, developing, and optimizing the structure of organic solar cells before they are manufactured. They then subject these cells to various influencing factors to evaluate their performance characteristics, efficiency, and lifespan. In addition, they compile comprehensive databases for use in various scientific studies and research.

In this work, we have used a database obtained from an experimental work by Sacramento et al.<sup>10</sup> studying the stability and the degradation analysis of iOSCs formed by PTB7-Th: PC70BM Blend with Different Stacks of Electron Transport Layers (*ETLs*). These iOSCs were fabricated in a nitrogen (N<sub>2</sub>) environment.

## Device structure and fabrication

Figure 1 shows the proposed device structure for the iOSCs studied by A. Sacramento et al.<sup>10</sup>, which is composed of the following layers: (ITO)/(ETL)/(PTB7-Th:PC<sub>70</sub>BM)/(V<sub>2</sub>O<sub>5</sub>)/(Ag). Five types of configurations for the *ETLs* have been studied: PFN/LiF (type I), LiF/PFN (type II), PFN (type III), LiF (type IV), and without an *ETL* (type V). The iOSCs were fabricated, and their  $J$ - $V$  characteristics were measured as reported elsewhere<sup>10,12,26,27</sup>.

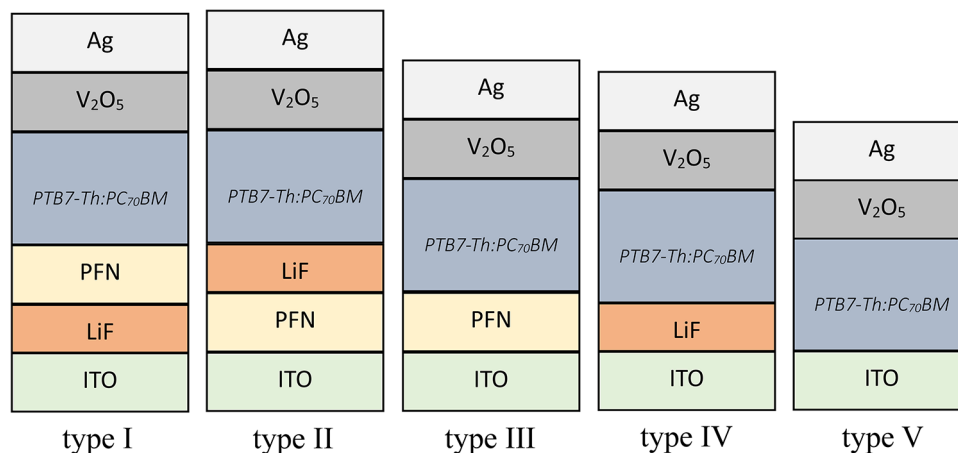
Five groups of iOSCs were formed to study the behavior of their parameters over time, which were:

- Group I: ITO/PFN/LiF/PTB7-Th: PC<sub>70</sub>BM / V<sub>2</sub>O<sub>5</sub>/Ag,
- Group II: ITO/LiF/PFN/PTB7-Th: PC<sub>70</sub>BM / V<sub>2</sub>O<sub>5</sub>/Ag,
- Group III: ITO/PFN/PTB7-Th: PC<sub>70</sub>BM / V<sub>2</sub>O<sub>5</sub>/Ag,
- Group IV: ITO/LiF/PTB7-Th: PC<sub>70</sub>BM / V<sub>2</sub>O<sub>5</sub>/Ag,
- Group V: ITO/PTB7-Th: PC<sub>70</sub>BM/V<sub>2</sub>O<sub>5</sub>/Ag.

### Layers description

The different layers involved in the iOSC structure are described as follows:

- ITO (indium tin oxide): Typically used as a transparent conductive electrode due to its high transparency and conductivity. It's often used in displays, solar cells, and other optoelectronic devices.



**Fig. 1.** Device structure of iOSCs.

- PFN (Poly[(9,9-bis(3'-(N,N-dimethylamino)propyl)-2,7-fluorene)-alt-2,7-(9,9-dioctylfluorene)]): A polymer used as an electron-blocking layer or hole-transporting layer in organic electronic devices. It helps improve the device's efficiency and performance.
- LiF (lithium fluoride): is another material used as an ETL in organic solar cells, often in combination with other layers. The search results show that the performance of OPV cells with LiF as the sole ETL is not as good as when LiF is used in an ETL stack, such as (LiF/PFN) or (PFN/LiF)<sup>28</sup>.
- PTB7-TH: PC70BM: A blend of organic semiconductors commonly used in organic solar cells. PTB7-Th is a polymer donor material, while PC70BM is a fullerene acceptor material. Together, they form the active layer responsible for converting light into electricity.
- HTL (hole-transport layer): This layer facilitates the transport of positive charges (holes) from the active layer to the anode. This configuration uses V<sub>2</sub>O<sub>5</sub> (Vanadium Pentoxide) as the HTL, serving as a control/reference structure.

Ag (silver): Often used as a reflective electrode or cathode in organic electronic devices. In this configuration, it's likely serving as the cathode, collecting electrons generated in the active layer. Optimization of layers structure in OSCs The optimization of the ETL, Hole Transport Layer (HTL), and the active layer is crucial for enhancing the performance and longevity of organic solar cells (OSCs). ETL materials' choice and stacking order significantly influence the stability and efficiency of OSCs. Studies show that using Poly[(9,9-bis(3'-(N,N-dimethylamino)propyl)-2,7-fluorene)-alt-2,7-(9,9-dioctylfluorene)] (PFN) in combination with a thin layer of Lithium Fluoride (LiF) improves stability compared to other configurations, with the PFN/LiF stack providing the best long-term performance<sup>29</sup>. For instance, cells with this stack exhibited a TS80 of 201 h, demonstrating the importance of optimizing ETL structures for durability and efficiency<sup>30</sup>.

The HTL, such as Vanadium Pentoxide (V<sub>2</sub>O<sub>5</sub>), plays a vital role in facilitating hole transport from the active layer to the anode<sup>31</sup>. By establishing a reference structure with V<sub>2</sub>O<sub>5</sub>, researchers can evaluate the impact of various ETL configurations on the performance of iOSCs<sup>32</sup>. Optimizing the HTL alongside the ETL is essential for achieving long-term stability and performance<sup>33</sup>.

The active layer's morphology, thickness, and material composition are critical for maximizing charge separation and transport<sup>34</sup>. Fine-tuning these parameters can lead to significant improvements in PCE<sup>35</sup>. For example, a well-defined fibril network morphology can enhance charge transport and reduce recombination losses, achieving PCEs as high as 16.5%<sup>36</sup>. Additionally, controlling the active layer thickness is vital, as it affects key electrical parameters like short-circuit current density ( $J_{sc}$ ) and open-circuit voltage ( $V_{oc}$ )<sup>12</sup>.

Overall, careful optimization of the ETL, HTL, and active layers is essential for enhancing the efficiency, stability, and overall performance of iOSCs, ultimately contributing to their practical application and longevity.

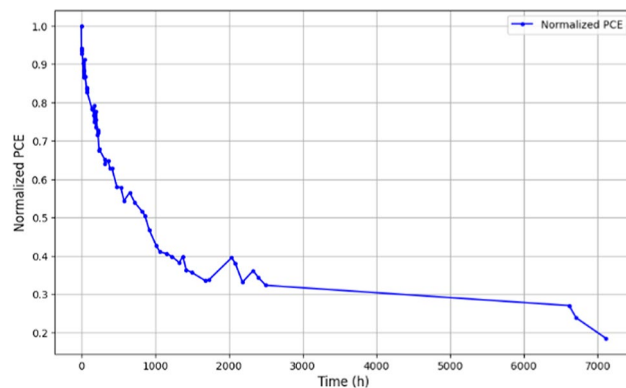
### Power conversion efficiency

The PCE is one of the most important parameters for evaluating the performance of solar cells<sup>37</sup>. It provides a measure of the electrical power that can be generated by a solar cell compared to the incident light power<sup>3</sup>. The PCE is generally expressed as a percentage and is calculated using the formula<sup>38</sup>:

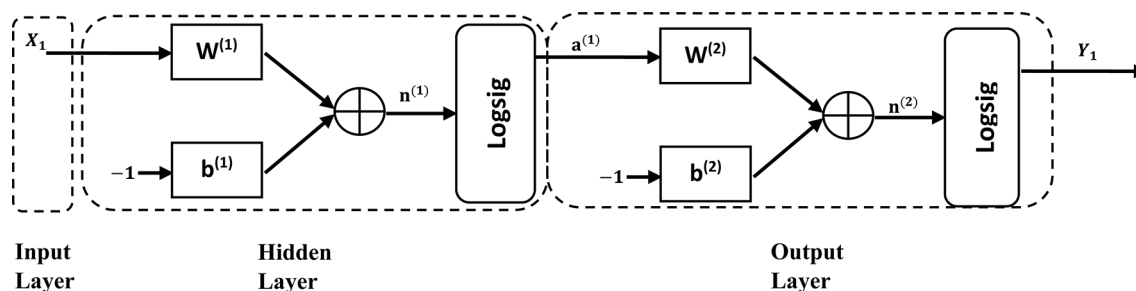
$$PCE = \frac{J_{sc} * V_{oc} * FF}{P_{in}} \quad (1)$$

where  $P_{in}$  is the power of the incident light<sup>39</sup>.

Alternatively, the PCE can be determined directly from the current density–voltage (J–V) characteristics of the solar cell, taking into account the measured  $V_{oc}$ ,  $J_{sc}$  and  $FF$ <sup>40</sup>. This equation highlights that all three parameters



**Fig. 2.** Degradation of the PCE over time for group I.



**Fig. 3.** Generic architecture of MLP.

( $V_{oc}$ ,  $J_{sc}$ , and  $FF$ ) can limit the overall device performance<sup>38</sup>. Optimizing these key parameters through careful design and engineering of the solar cell structure and materials is crucial for achieving high-efficiency photovoltaic devices<sup>17</sup>.

Figure 2 illustrates the degradation of the (PCE) of (iOSCs), Group 1: (ITO/PFN/LiF/PTB7-Th: PC<sub>70</sub>BM/V<sub>2</sub>O<sub>5</sub>/Ag) over time. Initially, the PCE exhibits a rapid decrease during the first few hundred hours of operation, a phenomenon known as "burn-in"<sup>41</sup>. This initial decay is attributed to photochemical reactions occurring between the various layers of the cell<sup>42</sup>. Following this "burn-in" phase, the PCE experiences a slower exponential decay, which aligns with observations reported in the literature for organic solar cells based on the PTB7-Th:PC<sub>70</sub>BM blend<sup>11</sup>. Over time, PCE degradation poses a significant challenge to the commercialization of iOSCs. Understanding and minimizing degradation mechanisms, particularly during the "burn-in" phase, is essential for enhancing the long-term stability of these photovoltaic devices.

### Degradation mechanisms in inverted organic solar cells

Degradation refers to the decline in the performance of organic solar cells beyond the initial "burn-in" phase<sup>41</sup>. This decline can be caused by environmental factors such as oxidation and humidity, morphological changes in the active and transport layers, and interactions between the different layers of the device<sup>43</sup>. In summary, while "burn-in" represents an initial degradation phenomenon, overall degradation denotes the long-term performance decline of iOSCs<sup>42</sup>. Optimizing the structure of the device layers is crucial for mitigating both types of degradation<sup>44</sup>.

### Model development

The present work is divided into two main contributions. First, we propose a double exponential model based on Levenberg–Marquardt algorithm rather than the simple Nonlinear Least Squares model used by Sacramento et al.<sup>10</sup> to predict the PCE of iOSC. Second, we introduce four different ML techniques namely MLP, LSTM, SVM and RF to predict the PCE of iOSC. In the following subsections, we provide a theoretical background of these proposed methods.

### Multi-layer perceptron

An MLP is a non-dynamic kind of artificial neural networks that is commonly used in ML for supervised learning problems, including fitting, prediction and classification problems. As can be seen from Fig. 3, the MLP consists of three main layers: an input layer, a hidden layer and an output layer. The input layer receives the 0external excitation signals and fore propagates them to the hidden layer neurons. The hidden layer consists of a given number of hidden layers, with given number of neurons in each, that act as the MLP's actual processing

engine. The output layer generates a decision or prediction corresponding to the input signals. The output of each neuron is updated according to Eq. (3):

$$y_j = f \left( \left( \sum_i w_{ij} * X_{ij} \right) - b_j \right) \quad (2)$$

where  $y_j$  is the  $j$ th neuron output.  $X_{ij}$  is the excitation coming from the neurons of the previous layer  $i$ .  $w_{ij}$  is the weights linked the neurons of the previous layer  $i$  to the neuron  $j$ .  $b_j$  is the bias.  $f$  is the activation function. It can be a sigmoid, a hyperbolic tangent or a rectified linear unit function. Scientists often apply MLP to supervised learning tasks using training on a set of input–output pairs helping to extract the relation between those inputs and outputs. The training process involves adjusting the weights and biases of the model to minimize the error between computed and desired outputs. Different optimization algorithms like gradient descent or its variants along with an activation function are used to adjust weights and bias relative to the error.

In this study, we have chosen to work on the prediction of the PCE of iOSCs based on MLP as it has shown good prediction capabilities for complex systems and patterns due to its ability to learn and store knowledge (in the form of weights) for future predictions<sup>45</sup>. MLPs are powerful models that have been and still are used for a wide range of engineering tasks<sup>21</sup> including solar energy applications<sup>22,46,47</sup>.

### Long short-term memory

LSTM is a kind of dynamic ANNs. It belongs to the family of recurrent neural networks (RNN) initially designed to deal with time-dependent tasks<sup>48</sup>. LSTM was proposed by Hochreiter and Schmidhuber in 1997 essentially to overcome the limitations of learning long-term dependencies based on traditional RNNs due to the vanishing and exploding gradient problems usually accompany the well-known Back Propagation Through Time (BPTT) learning algorithm<sup>49</sup>. LSTMs can learn and remember over extended sequences through their sophisticated memory cells. As shown by Fig. 4, the memory cell is the core of the LSTM. It maintains a cell state  $c_t$  updatable via a series of operations that selectively add or remove information. LSTMs has also special neural units that regulate the information flow within the cell called gates. LSTMs employ several kinds of gates. Forget Gates  $f_t$  select the information to be discarded from the cell state. Input Gates  $i_t$  updates the cell state by injecting new information. Output Gates  $o_t$  select the information from the cell state to be used for the output production. Activation functions like the sigmoid and hyperbolic tangent are typically used by LSTM to control the information flow and to update the state of the memory cell.

The LSTM architecture can capture long-term dependencies in input data sequences via the ability to keep relevant information for long periods of time. This property renders LSTMs a popular solution in many engineering tasks where learning long-term dependencies is essential such as natural language processing<sup>51</sup>, time series prediction<sup>52</sup>, solar energy<sup>53,54</sup>, to name but a few.

### Support vector machine

Support Vector Machine (SVM) is a robust algorithm frequently used to tackle non-linear problems, even when the dataset size is small. It is applicable to both classification and regression tasks, with the regression variant referred to as Support Vector Regression (SVR).

For a dataset containing  $N$  samples of input–output pairs:

$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\} \in R^k \times R$ . The regression function can be described as:  $f(X_i) = W^T \cdot X_i + b, w \in R^K$  (3).

- o  $W$  is the normal vector to the hyperplane,
- o  $b$  is the bias term (the distance of the hyperplane from the origin),
- o  $x_i$  is the input vector.

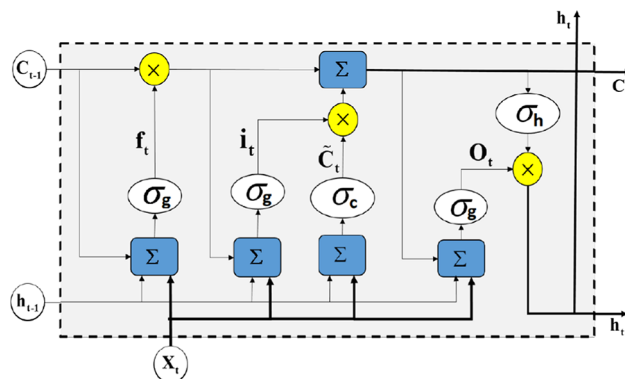


Fig. 4. Generic architecture of LSTM<sup>50</sup>.

SVM for non-linear problems uses a non-linear mapping function that transforms the input data into a higher-dimensional feature space, making it easier to separate the data linearly in that space. The optimization problem for SVM is:

$$\min \left( \frac{1}{2} \|w\|^2 + C \sum_{i=1}^M \tau_i \right) \quad (4)$$

subjects to boundary conditions

$$y_i (w^T \cdot \varphi(x_i) + b) \geq 1 - \tau_i, \tau_i \geq 0 \quad (5)$$

where:

- $\tau_i$  is a slack variable allowing some misclassification,
- $C$  is the regularization parameter that controls the trade-off between fitting the training data and maintaining model simplicity. A larger  $C$  tends to increase model accuracy on training data but may cause overfitting, while a smaller  $C$  results in a simpler model that might misclassify some points.

The non-linear transformation  $\varphi(x_i)$  can be replaced with kernel functions, such as  $K(x_i, x_j)$ , which compute the necessary values in the input space, making SVM computationally efficient for non-linear problems.

### Random forests

The RF is a highly adaptable supervised ML algorithm, introduced by Leo Breiman and Adele Cutler in 2001<sup>55</sup>. Its popularity stems from its ease of use and flexibility, which have led to widespread adoption in both classification and regression tasks. At its core, RF leverages ensemble learning, a powerful technique where multiple models (in this case, decision trees (DT)) are combined to solve complex problems and enhance model performance.

While DTs are a fundamental component of many supervised learning algorithms, they are often susceptible to issues such as bias and overfitting, where the model becomes too tailored to the training data and struggles with generalization. However, when DTs are aggregated into an ensemble using the Random Forest method, the algorithm mitigates these issues. By utilizing a large number of decision trees that are uncorrelated with one another, the Random Forest algorithm achieves significantly more accurate and reliable predictions.

A flow chart of the RF is shown in Fig. 5. Each DT in the forest is built using a random subset of the dataset and a random selection of features at each split. This randomization injects variability into the individual trees, making the model less prone to overfitting and improving its predictive power. During the prediction phase, RF combines the results of all trees through voting (in classification tasks) or averaging (in regression tasks). This

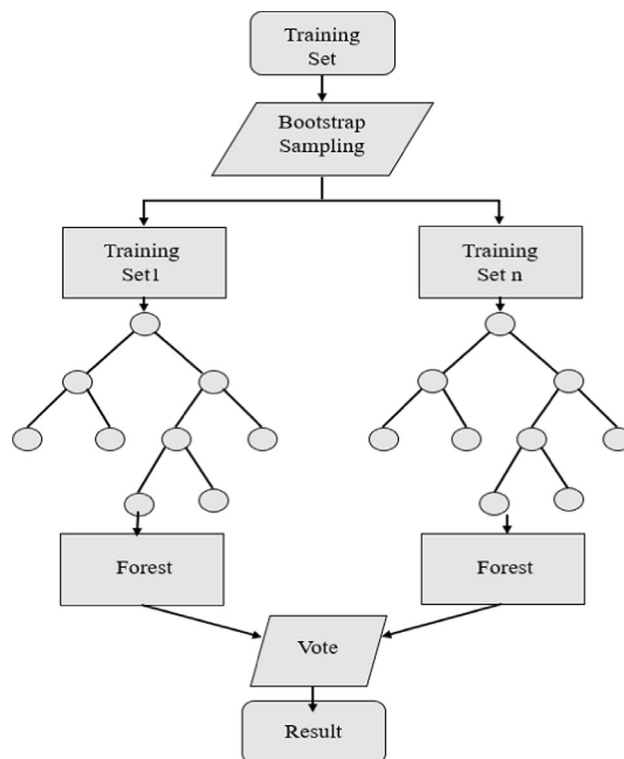


Fig. 5. The flowchart of random forest.

collaborative approach allows the algorithm to arrive at stable, precise predictions by drawing from the collective wisdom of multiple trees.

To effectively train a RF model, there are three key hyperparameters that need to be set: node size, which controls the depth of the trees; the number of trees, which determines how many trees will be included in the forest; and the number of features sampled, which dictates how many features are considered at each split. Properly tuning these parameters is critical for optimizing the model's performance.

Random Forests are highly regarded for their ability to handle complex datasets, effectively reduce overfitting, and provide robust predictions across various environments. They are also capable of dealing with missing data and scaling to large datasets. Another key advantage is their ability to determine feature importance, offering insights into which variables contribute most to the model's predictions.

These strengths make Random Forests applicable across a wide range of fields. In finance, they are used for tasks like customer evaluation and fraud detection<sup>56</sup>. In healthcare, they are employed in areas such as computational biology, biomarker discovery, and sequence annotation<sup>57</sup>, demonstrating their versatility across diverse engineering and scientific applications.

### Double exponential model

The double exponential model, often used in curve fitting and data modeling, can be expressed as:

$$f(x) = Ae^{-Bx} + Ce^{-Dx} \quad (6)$$

where:  $A$ ,  $B$ ,  $C$ , and  $D$  are parameters to be determined through fitting,  $x$  represents the independent variable and  $e$  denotes the base of the natural logarithm. Figure 6 shows an example of this model for different values of  $A$ ,  $B$ ,  $C$ , and  $D$  parameters.

Fitting the double exponential model to a given data involves finding the best values for these parameters that minimize the difference between the model's predictions and the actual data points.

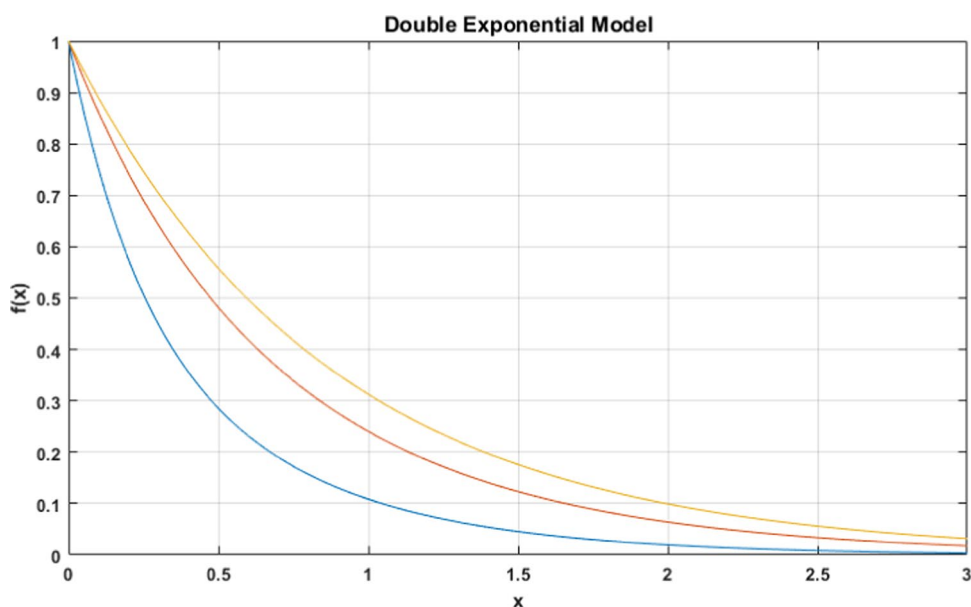
A general approach to fit the double exponential model to a given data consists of:

- Initial Parameter Estimation: One might start with random initialization of the parameters  $A$ ,  $B$ ,  $C$ , and  $D$  to begin the fitting process. The initialization can be also based on the data or any prior knowledge of the system.
- Curve Fitting Algorithm: A curve fitting algorithm or optimization technique is used to adjust the parameters of the double exponential model. Popular optimization methods include: Nonlinear Least Squares, Gradient Descent or Levenberg–Marquardt Algorithm<sup>24</sup>.

### Experimental setup

#### Used dataset

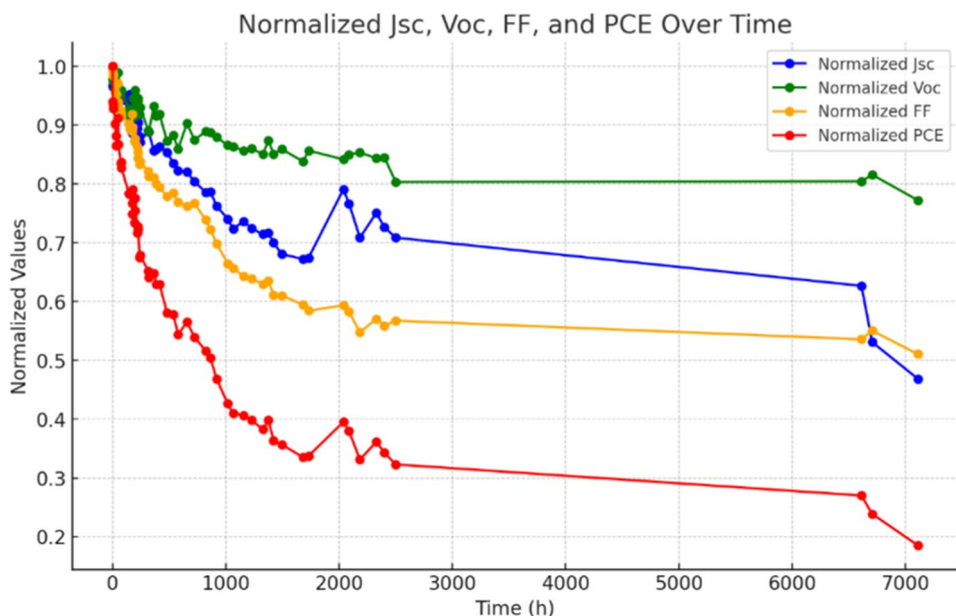
Our study focused on leveraging the experimental data obtained by Sacramento et al.<sup>10</sup> by examining the degradation of inverted organic solar cells (iOSCs) using a PFN/LiF stacking arrangement that showed improved cell stability. Key performance parameters were evaluated, including open circuit voltage (Voc), short circuit current density (Jsc), fill factor (FF), and power conversion efficiency (PCE). These parameters were extracted from current density–voltage (J-V) measurements conducted under both dark and light conditions over a period of 7,110 h<sup>10</sup>. The measurements were performed at room temperature (23°C) with an average relative humidity (RH) of 54%, using a Keithley 2450 source measurement unit. The samples were tested using a solar



**Fig. 6.** The double exponential model for different values of  $A$ ,  $B$ ,  $C$  and  $D$  parameters.

Time (h)	Jsc (mA/cm <sup>2</sup> )	Voc (V)	FF (%)	PCE (%)
0	18.06692	0.7033	56.42521	7.16964
2	17.47278	0.68772	56.10268	6.74151
...	...	...	...	...
6707	9.59309	0.57366	31.06226	1.7094
7110	8.46811	0.54319	28.81635	1.32549

**Table 1.** Performance parameters of (i)OSCs over time during degradation testing.



**Fig. 7.** Normalized performance parameters for (i)OSCs, regarding their initial values as a function of time.

simulator under the AM 1.5G spectrum, with a light intensity of 100 mW/cm<sup>2</sup> (1 sun), calibrated with a Solarex Corporation-certified monocrystalline silicon photodiode<sup>58</sup>.

As shown in Table 1, the dataset is organized in a tabular format with columns representing the performance parameters of iOSCs over time during degradation testing. This comprehensive dataset enables a thorough analysis of the degradation mechanisms affecting the solar cells' long-term performance. The data spans a wide timeframe, allowing for detailed tracking of how each parameter evolves during the degradation process.

Figure 7 shows the normalized performance parameters for iOSCs with respect to their initial values as a function of time. These parameters include Jsc, Voc, FF, and PCE. The normalization allows for a clearer comparison of the degradation rates of each parameter over time, highlighting the stability and long-term performance of the iOSCs. The data demonstrate the varying rates of degradation, crucial for understanding and improving the longevity and efficiency of iOSCs. In our study, we focus particularly on the degradation of the time-dependent power conversion efficiency (PCE), which is critical for evaluating the performance of organic solar cells over time.

It turns out that in our study, we are interested in the degradation of the PCE due to its significance in determining the efficiency of organic solar cells. It is important to note that the dataset is complete and representative, as it was collected under controlled climatic conditions, including constant temperature, humidity, and radiation. This ensures that the degradation observed in the data is purely due to time and internal factors rather than external environmental fluctuations, allowing for more accurate modeling and analysis of iOSCs' long-term performance.

### Data pre-processing

In order to improve the proposed model's learning process and performance, we use two main preprocessing operations in this study. First, we handled missing values by fitting the original PCE using a double exponential model, and then imputing the missing points, allowing the model to work with a complete dataset and avoid biases caused by missing data points. Second, we applied data scaling to ensure that all values are on a comparable scale, which is highly beneficial for ML training. Details of these preprocessing operations are presented in this section.



### Data imputation

In our experiments, the original dataset consisted of 58 measured PCE values over 7100 h, but there were several significant gaps in the measurements. One of the most critical gaps occurred between the 2,500-h mark and the 6500-h mark, leaving a missing interval of approximately 4,000 h. This absence of data could have introduced uncertainty into the analysis and compromised the accuracy of any conclusions drawn regarding the long-term degradation of the iOSC devices.

To address the issue of missing data, several techniques are typically used, such as linear interpolation, extrapolation, mean or median imputation, and forward/backward fill, among others. However, in our work, we employed an imputation technique based on fitting the original data to a double exponential model, which closely matched the observed degradation trend, as will be discussed in Section "Data fitting". By applying this model, we were able to generate 102 additional data points within the 4,000-h gap and other smaller gaps in the dataset. This allowed us to increase our data to 160 values and to maintain a continuous dataset while ensuring the imputed points followed the same degradation behavior as the original observations.

Figure 8 illustrates the original and imputed data, along with the double exponential fit. The imputation method allowed us to bridge the gaps in the dataset, particularly in the critical range between 2,500 and 6,500 h, ensuring the continuity of the data for subsequent analysis and modeling. This approach enhanced the robustness of our analysis, making it possible to better evaluate the long-term stability and performance of the iOSCs.

### Data normalisation

To prepare the data for analysis and model training, the PCE values are scaled between 0 and 1 using the following formula:

$$PCE_{Normalized} = \frac{PCE - \min(PCE)}{\max(PCE) - \min(PCE)} \quad (7)$$

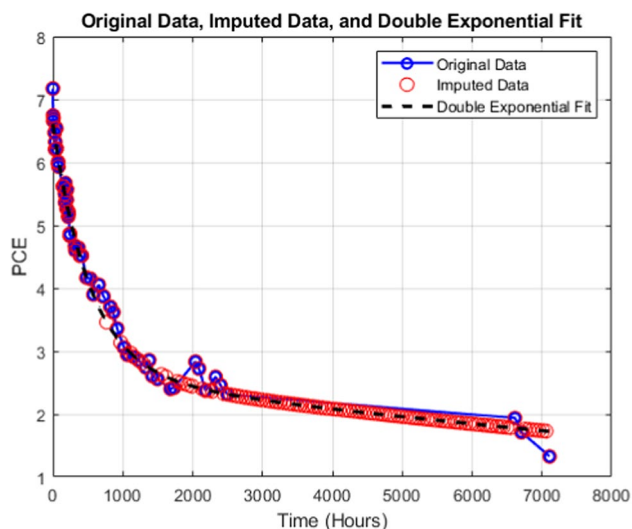
where PCE and  $PCE_{Normalized}$  represent the original and the normalized values of the PCE, respectively. The  $\max$  and  $\min$  operators are the maximum and minimum PCE values in the within the dataset.

Scaling the PCE values ensures that all values are on a comparable scale, which is important for machine learning-based methods. Additionally, scaling can improve the convergence speed of optimization algorithms by reducing the likelihood of skewed gradients, leading to more efficient ML training processes.

### Training process

Our leveraged dataset contains 160 PCE values recorded over approximately 7,100 h, with irregular time steps. To ensure a fair evaluation of the ML techniques used for predicting future PCE values based on the learned history of PCE degradation, we applied a cross-validation approach based on the rolling block technique. The rolling technique, also known as a rolling or sliding window technique, is commonly used in time-series analysis to train models. It involves creating a dynamic or shifting window of data, where each window contains a subset of sequential observations. The model is trained on the data within this window, and predictions are made for subsequent points. After making predictions, the window shifts forward by one or more-time steps, and the process repeats.

This method offers several advantages, particularly for time-dependent data like PCE values. First, it helps maintain the temporal structure of the dataset by ensuring that training is done on sequential data, which is essential for capturing degradation patterns over time. Additionally, the rolling technique allows the model to



**Fig. 8.** Imputation of the missing PCE data.

learn from different portions of the dataset while avoiding data leakage between training and testing sets. It also ensures that the model is tested on unseen data, which is crucial for evaluating the model's generalization ability.

In our experiments, we used training blocks of 80 values and test blocks of 20 values. Table 2 illustrates the various windows used for training and testing across the four machine learning techniques applied in this study.

### Performance evaluation

To evaluate the performance of our proposed models for the PCE prediction of the iOSCs degradation, we have used an objective evaluation based on various statistical metrics to measure the accuracy and reliability of the model's predictions. These metrics include Mean Squared Error (MSE), Root Mean Squared Error (rMSE), Relative Root Mean Squared Error (rrMSE), Relative Absolute Error (RAE), Residual Standard Error (RSE), Residual Sum of Squares (RSS), and the Coefficient of Determination ( $R^2$ ). Below is an explanation of each metric<sup>59–63</sup>:

**Mean squared error (MSE):** The average of the squared differences between the actual and predicted values.

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

Lower values of MSE indicate better model performance. This metric is sensitive to outliers due to the squaring of errors.

**Root mean squared error (rMSE):** The square root of the mean squared error:

$$rMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (9)$$

It represents the standard deviation of the residuals, making it easier to interpret as it is in the same units as the target variable.

**Relative root mean squared error (rrMSE):** rMSE normalized by the mean of the actual values:

$$rrMSE = \frac{rMSE}{\bar{y}} \quad (10)$$

It provides insight into how large the residuals are relative to the magnitude of the actual values, offering a normalized view of prediction error.

**Relative absolute error (RAE):** Measures the sum of absolute errors relative to the sum of absolute errors of a baseline model (such as the mean of actual values):

$$RAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{\sum_{i=1}^N |y_i - \bar{y}|} \quad (11)$$

It indicates the proportion of error in the model compared to a baseline model. Lower values of RAE suggest better model performance.

**Residual standard error (RSE):** Estimates the standard deviation of the residuals from the model, It is defined by the following formula:

$$RSE = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (12)$$

It Reflects the model's goodness of fit; lower values indicate a better fit of the model to the data.

**Residual sum of squares (RSS):** The sum of the squared residuals, representing the total deviation of the predictions from the actual values:

$$RSS = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (13)$$

It provides a measure of the total error in the model's predictions. Lower values indicate a better fit.

	Training set	Test set
Block 1	From 1st to 80th samples	From 81st to 100th samples
Block 2	From 21st to 100th samples	From 101st to 120th samples
Block 3	From 41st to 120th samples	From 121st to 140th samples
Block 4	From 61st to 140th samples	From 141st to 160th samples

**Table 2.** The division of the data into training and test sets.

**Coefficient of determination ( $R^2$ ):** The proportion of variance in the dependent variable that is predictable from the independent variables:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (14)$$

$R^2$  Ranges from 0 to 1, with higher values indicating a better fit of the model to the data. An  $R^2$  of 1 signifies a perfect fit.

## Results

In this section, we present the results obtained from this study. First, we begin by discussing the results of data fitting achieved using our leveraged double exponential model. Following that, we will present the findings from applying four different machine learning techniques—namely, MLP, LSTM, SVM, and RF to predict the PCE of the studied iOSC structure.

### Data fitting

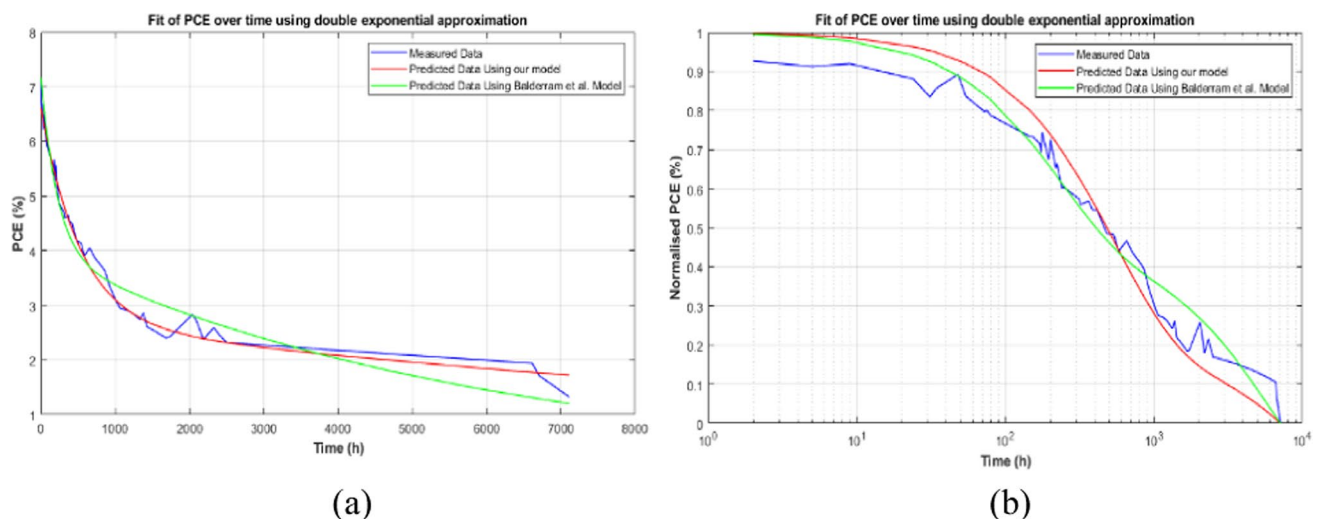
According to Sacramento et al.<sup>10</sup>, the degradation of the PCE over the time follows a mathematical model based on two-term exponential function:

$$\frac{PCE(t)}{PCE(0)} = A_1 e^{-\frac{t}{T_1}} + A_2 e^{-\frac{t}{T_2}} \quad (15)$$

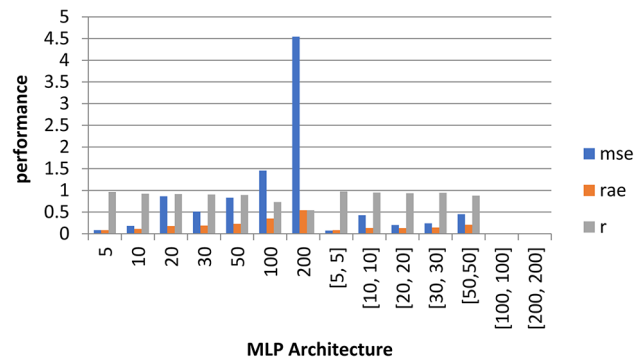
where  $PCE(0)$  is the PCE at time  $t=0$  h,  $t$  is the time,  $(T_1, T_2)$  are time constants and  $(A_1, A_2)$  are weights of the degradation. This model for PCE degradation in solar systems is widely accepted in the literature<sup>24,25</sup>. The parameters  $(A_1, A_2)$  and  $(T_1, T_2)$  are typically determined by minimizing the least squares error between the measured and modeled PCE values based on this double exponential model.

For instance, Sacramento et al.<sup>10</sup> optimized the time constants and degradation weights for the data presented in Section "Used dataset", yielding the following values:  $A_1=0.45$ ,  $A_2=1-A_1=0.55$ ,  $T_1=220$  h and  $T_2=6000$  h. In this work, we utilized the same model but employed the Levenberg–Marquardt algorithm to optimize the double exponential parameters, as opposed to the Nonlinear Least Squares method used by Sacramento et al.<sup>10</sup>. Our optimization resulted in the following parameter values:  $A_1=3.9510$ ,  $A_2=2.6580$ ,  $T_1=529.3806$  and  $T_2=1637.2$ .

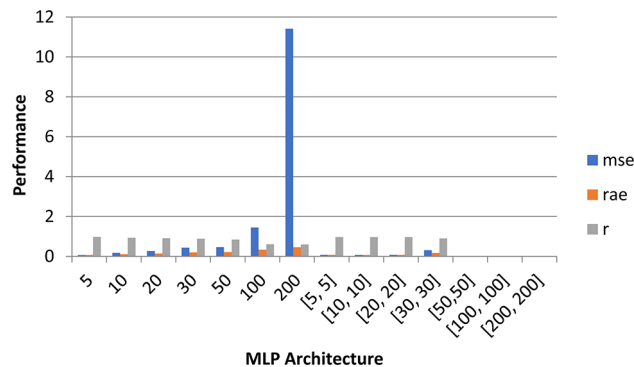
Figure 9 compares the measured PCE data with the predictions made using the double exponential models proposed by Sacramento et al. and in this study. Overall, both models fit the measured PCE data well. However, it is evident from the figure that our proposed model, based on the Levenberg–Marquardt optimization, more closely aligns with the measured PCE data compared to the model by Sacramento et al., which used the Nonlinear Least Squares method. Another observation is that in both cases, the surface of the double exponential model is smooth and cannot capture the fluctuations in the measured PCE data.



**Fig. 9.** Fitting of the PCE over the time based on the double exponential model (a) PCE in function of time (b) Normalized PCE in function of logarithmic scaled time.



**Fig. 10.** MLP performance in function of different architectures for Levenberg Marquardt training algorithm.



**Fig. 11.** MLP performance in function of different architectures for Bayesian regularization backpropagation training algorithm.

## Data prediction

### Hyper parameters setting

In this section, we analyze the impact of hyperparameter configurations on the performance of various ML architectures, highlighting the importance of hyper parameters like architecture size and training algorithms in achieving optimal predictive accuracy.

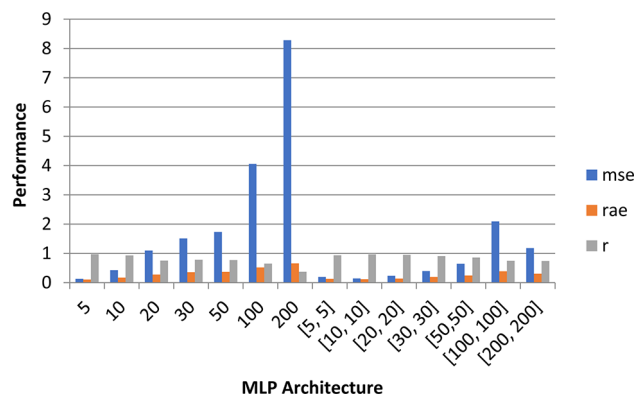
**MLP** Figure 10 shows the performance of MLP using the Levenberg–Marquardt (LM) training algorithm across different architectures. For single-layer architectures, the best performance occurs with 5 neurons, achieving an MSE of 0.082 and a high correlation coefficient ( $r$ ) of 0.967. As the number of neurons increases, performance declines, with 50 neurons showing a significant drop (MSE = 0.831,  $r$  = 0.894) and the largest architectures (100 and 200 neurons) performing poorly (MSE > 1.4,  $r$  < 0.73).

For multi-layer architectures, the architecture with two hidden layers, five neurons each (<sup>5,5</sup>) performs the best (MSE = 0.072,  $r$  = 0.978), while larger architectures like <sup>50,50</sup> result in reduced accuracy (MSE = 0.449,  $r$  = 0.878). This indicates that smaller, well-balanced architectures outperform larger, more complex ones due to the risk of overfitting.

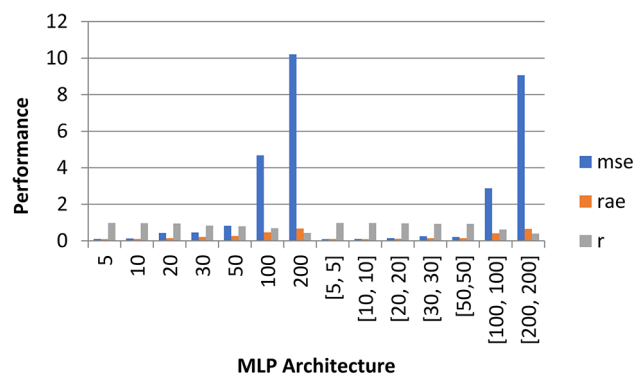
Figure 11 illustrates the performance of MLP using the Bayesian Regularization Backpropagation (BR) training algorithm across various architectures. For single-layer architectures, the best performance is achieved with 5 neurons, yielding an MSE of 0.068 and a high correlation coefficient ( $r$  = 0.973). As the number of neurons increases, the performance worsens, with architectures of 100 and 200 neurons showing significant increases in MSE (1.445 and 11.416, respectively) and lower  $r$  values (< 0.61), suggesting overfitting or diminishing returns with larger architectures.

In multi-layer architectures, the <sup>10,10</sup> architecture performs the best (MSE = 0.074,  $r$  = 0.970), closely followed by <sup>5,5</sup> neurons. However, larger architectures like <sup>30,30</sup> show a considerable drop in performance (MSE = 0.310,  $r$  = 0.901). This pattern indicates that smaller and moderate-sized architectures with the BR algorithm provide more robust generalization and better accuracy, while larger models may struggle with overfitting or poor convergence.

Figure 12 illustrates the performance of the MLP using the Gradient Descent (GD) training algorithm for various architectures. For single-layer architectures, performance generally declines as the number of neurons increases. The best performance is observed with 5 neurons, yielding a relatively low MSE of 0.129 and a high correlation coefficient ( $r$  = 0.973). However, as the number of neurons increases, the MSE significantly worsens,



**Fig. 12.** MLP performance in function of different architectures for Gradient descent.



**Fig. 13.** MLP performance in function of different architectures for resilient backpropagation training algorithm.

with 100 and 200 neurons showing MSE values of 4.06 and 8.28, respectively, accompanied by a substantial drop in correlation ( $r < 0.65$ ), indicating poor model generalization and potential overfitting.

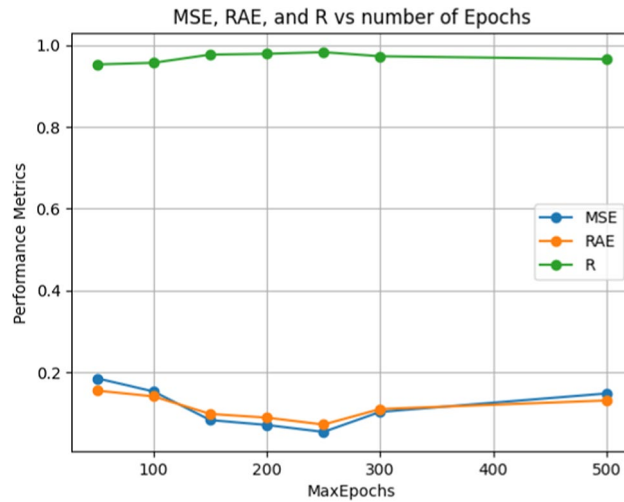
For multi-layer architectures, a similar trend is observed. Architectures with  $5,5$  and  $10,10$  neurons perform relatively well, with low MSE (0.192 and 0.144) and high correlation ( $r = 0.934$  and  $0.965$ , respectively). However, larger architectures, such as  $50,50$  and  $[100, 100]$ , show significantly higher MSE and a sharp decline in correlation ( $r = 0.857$  and  $0.748$ ), confirming the diminishing returns with increasing model complexity. Overall, this highlights that smaller architectures are more effective with Gradient Descent, while larger models struggle with convergence and generalization.

Figure 13 illustrates the performance of the MLP utilizing the Resilient Backpropagation (Rprop) training algorithm across various architectures. For single-layer configurations, the optimal performance is achieved with 5 neurons, yielding a low Mean Squared Error (MSE) of 0.095 and a high correlation coefficient ( $r = 0.977$ ). As the number of neurons increases, performance deteriorates, with 50 neurons showing a significant increase in MSE (0.824) and a drop in correlation ( $r = 0.794$ ). The largest architectures, specifically those with 100 and 200 neurons, demonstrate poor performance with MSE values exceeding 4.67 and 10.21, and correlation coefficients below 0.69.

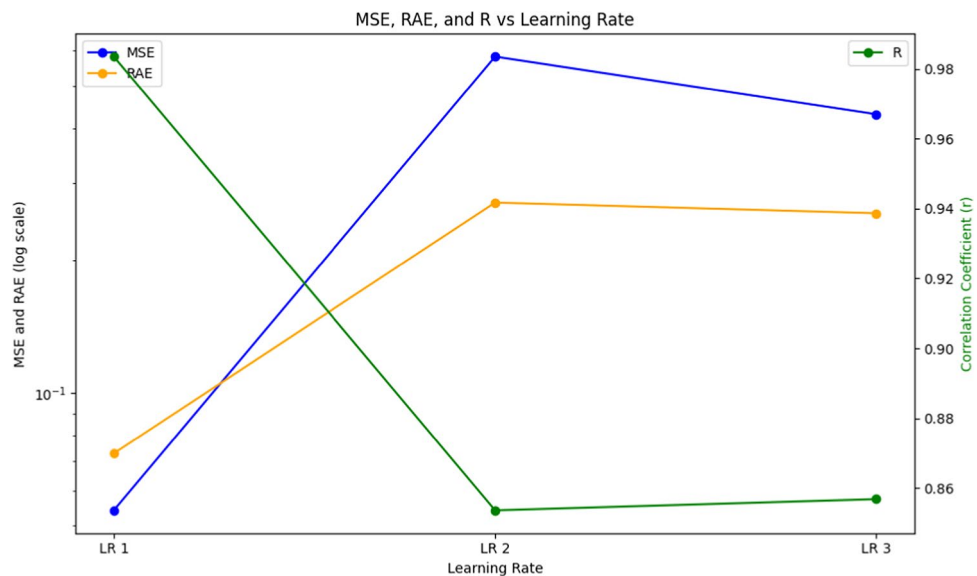
In the case of multi-layer architectures, the configuration of  $5,5$  neurons yields the best results (MSE = 0.095,  $r = 0.976$ ), while larger setups, such as  $50,50$ , show reduced accuracy (MSE = 0.209,  $r = 0.928$ ). These findings emphasize that smaller, balanced architectures are more effective with the Rprop algorithm, as they minimize the risk of overfitting compared to larger, more complex structures.

**LSTM** Figure 14 presents the relationship between the number of MaxEpochs and three key performance metrics: MSE, RAE, and  $r$ . As the MaxEpochs increases, MSE and RAE decrease steadily until reaching their lowest values at 250 epochs, where the model achieves optimal performance with an MSE of 0.054 and an RAE of 0.072. The correlation coefficient ( $r$ ) peaks at 250 epochs with a value of 0.983, reflecting the highest predictive accuracy. Beyond 250 epochs, MSE and RAE begin to rise, while  $r$  slightly decreases, indicating overfitting or diminishing returns in model training. This analysis demonstrates that 250 MaxEpochs offer the best balance between error reduction and model accuracy.

Figure 15 illustrates the relationship between MSE, RAE, and correlation coefficient ( $r$ ) across three data rows. MSE and RAE are plotted on a logarithmic scale, revealing significant differences in error across the rows. Row 1 achieves the lowest MSE (0.0541) and RAE (0.0729), coupled with the highest correlation ( $r = 0.9835$ ),



**Fig. 14.** LSTM prediction performance in function of the number of training epochs.



**Fig. 15.** LSTM prediction performance in function of the learning rate.

indicating optimal model performance. In contrast, Rows 2 and 3 show larger MSE and RAE values, with reduced correlation coefficients, reflecting a decline in model accuracy and efficiency. The semi-log scale highlights the disparities in error magnitude, demonstrating the clear superiority of Row 1 in terms of error minimization and correlation strength. This analysis justifies Row 1 as the most accurate model configuration.

Figure 16 showcases the performance of the LSTM model as the number of hidden units increases. The MSE and RAE curves reflect a consistent decrease in error until 200 hidden units, where the model performs optimally (MSE=0.0541, RAE=0.0729). Beyond 200 hidden units, both metrics rise again, signaling over fitting. The correlation coefficient (r) reaches its maximum at 200 hidden units (r=0.9835), further confirming this as the most efficient configuration. Thus, 200 hidden units offer the best balance between accuracy and model complexity.

Figure 17 compares the performance of three optimizers—SGDM, RMSprop, and Adam—based on MSE, RAE, and R. Adam emerges as the top performer, achieving the lowest MSE (0.0541) and RAE (0.0729), and the highest correlation (R=0.9835), indicating accurate predictions and minimal error. SGDM follows with a moderate MSE of 0.1727 and R of 0.9500, while RMSprop performs the worst, with an MSE of 0.3056 and R of 0.9386. These results highlight Adam's superior convergence and efficiency in this context. The clear performance gap between optimizers emphasizes the importance of selecting the right training algorithm to maximize model accuracy and minimize errors.

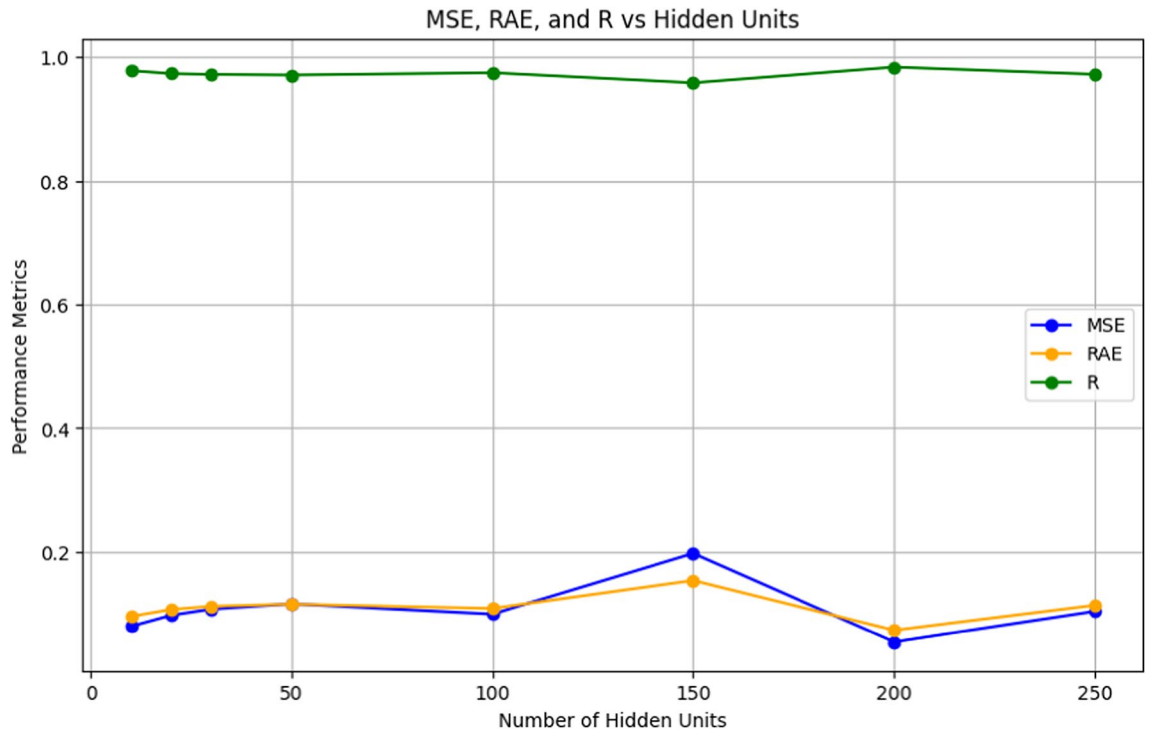


Fig. 16. LSTM prediction performance in function of the number of hidden units.

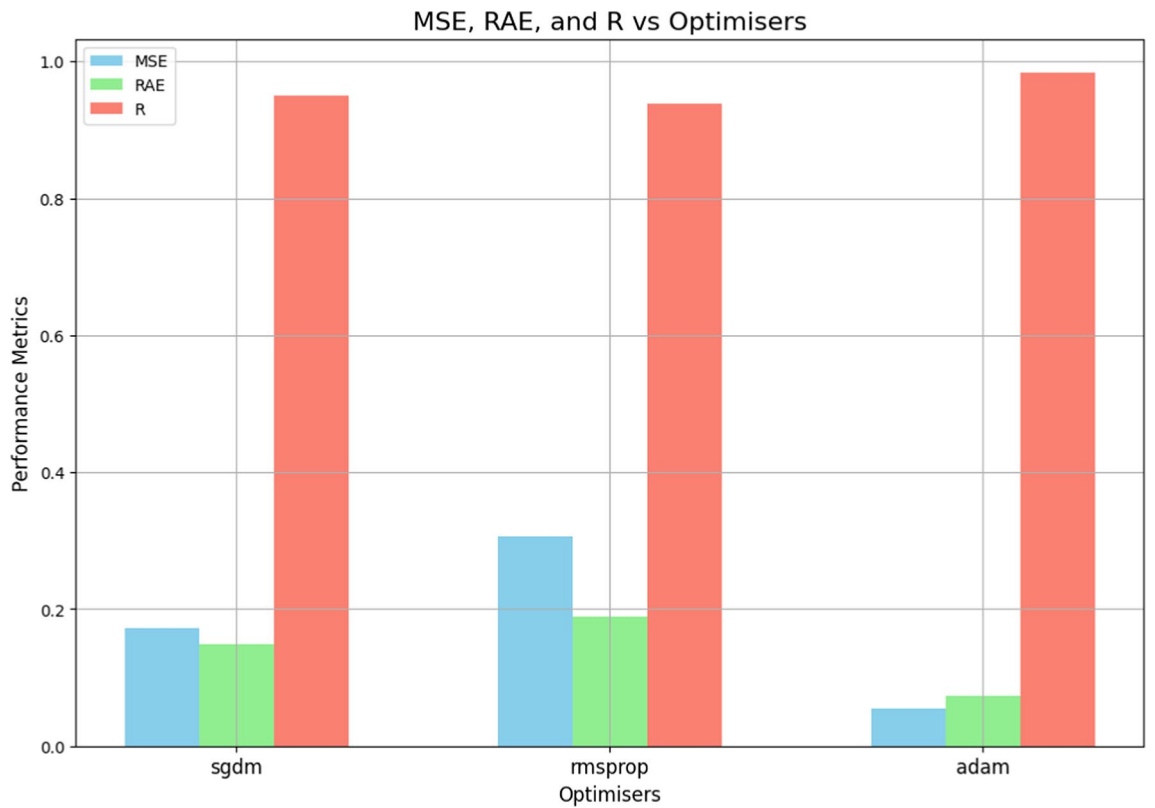


Fig. 17. LSTM prediction performance in function of the optimizer.

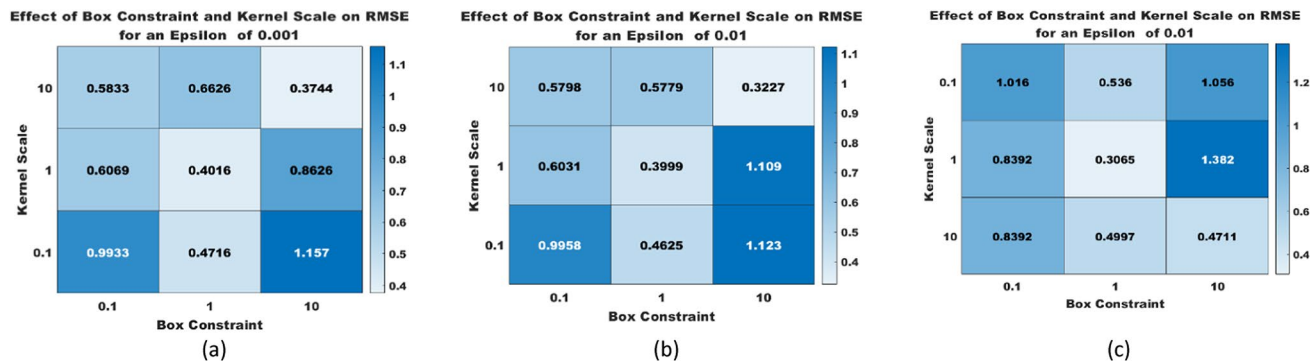


Fig. 18. The impact of SVM hyperparameters tuning on the performance of PCE prediction.

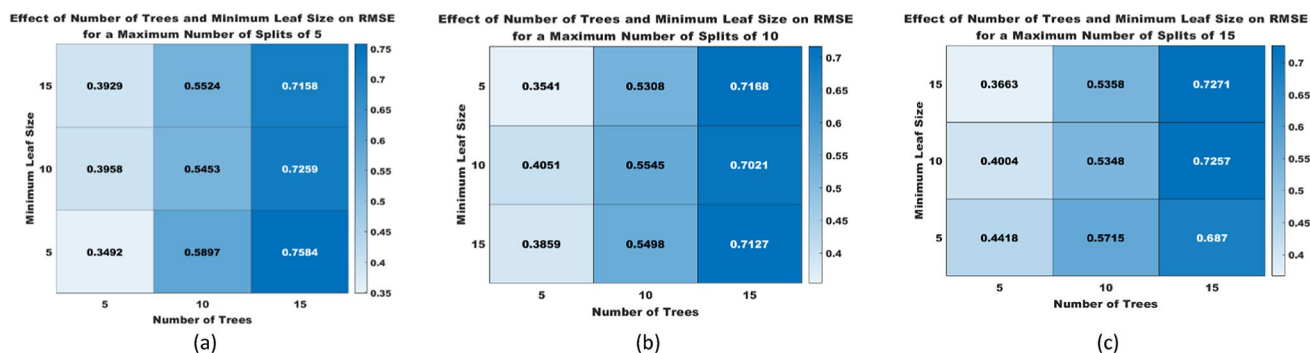


Fig. 19. The impact of RF hyper parameters tuning on the performance of PCE prediction.

**SVM** This sub-section studies the impact of SVM hyperparameters tuning on the performance of PCE prediction. As shown in Fig. 18, we examined the effect of Box Constraint and Kernel Scale on RMSE for different values of Epsilon. The results show that the performance of the model, in terms of prediction accuracy, is highly dependent on the interaction of these parameters. For an Epsilon of 0.001 and 0.01, the study finds that a higher Box Constraint of 10 and a Kernel Scale of 10 consistently yield the best performance, with the lowest RMSE values. Specifically, for Epsilon = 0.01, this combination produces an RMSE of 0.3227.

In contrast, when Epsilon increases to 0.1, a moderate Box Constraint of 1 performs best, yielding the lowest RMSE of 0.3065. This result suggests that for larger Epsilon values, which introduce more tolerance for errors, a smaller Box Constraint is necessary to avoid overfitting. Overall, the results indicate that Box Constraint of 10 and Kernel Scale of 10 form the best architecture for smaller Epsilon values, while for larger Epsilon, adjusting the Box Constraint downwards improves the model's generalisation and reduces errors.

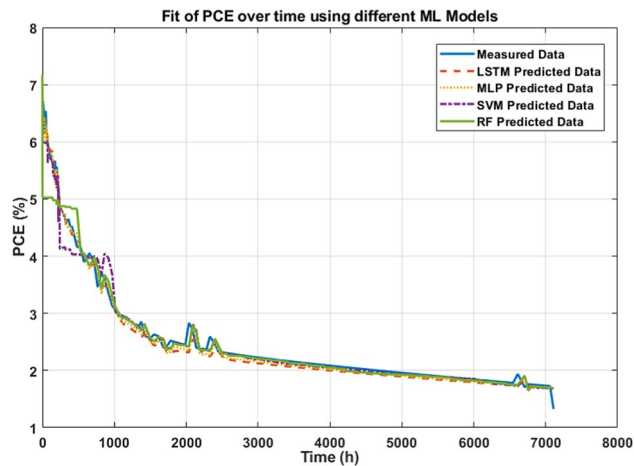
The most optimal SVM architecture among all experiments corresponds to an RMSE of 0.3065, achieved with a Box Constraint and Kernel Scale of 1 for an Epsilon of 0.1. This configuration strikes an ideal balance between fitting the data closely and maintaining sufficient flexibility in the decision boundary, effectively preventing overfitting while ensuring high prediction accuracy.

**RF** This sub-section investigates the impact of RF hyperparameter tuning on the performance of PCE prediction. As shown in Fig. 19, we examined the effect of the number of trees and minimum leaf size on RMSE for different values of the maximum number of splits.

Each panel of the figure corresponds to a value of maximum number of splits. It appears clearly that the three panels are similar regardless the value of this latter. Therefore, the maximum number of splits seems to do not have a big influence on the performance of the prediction. So, a small value of maximum number of splits are preferable as they decrease the complexity of the model. Increasing the number of trees leads to higher RMSE values, indicating that overfitting may occur with too many trees. For instance, for a maximum number of splits of 10, the lowest RMSE value of 0.3541 is observed with 5 trees and a minimum leaf size of 5. This suggests that small tree numbers combined with a lower leaf size yield better results, preventing the model from becoming too complex and improving generalization.

Across all experiments, the most optimal RF architecture corresponds to a configuration of 5 trees and a minimum leaf size of 5 for a maximum split of 5, yielding the lowest overall RMSE of 0.3492. This setup combines between low model complexity and high prediction accuracy, preventing overfitting while maintaining a high level of generalisation.





**Fig. 20.** Fitting of the PCE over time based on the proposed ML models. (a) PCE in function of time (b) Normalized PCE in function of logarithmic scaled time.

Model	MSE	rMSE	rrMSE	RAE	RSE	RS	R-squared
LSTM	0.0541	0.2223	8.0691	0.0729	0.0325	0.9675	0.9835
MLP	0.0716	0.2556	9.2620	0.0838	0.0437	0.9562	0.9778
SVM	0.1098	0.3065	10.9581	0.1001	0.0725	0.9275	0.9625
RF	0.1610	0.3723	13.4778	0.1239	0.1336	0.8664	0.9274

**Table 3.** A comparison between the measured and the predicted PCE data of the studied iOSC cell using different ML-based prediction techniques over all the lifetime of the cell.

#### Comparative study

Figure 20 shows the fit of PCE values over time, comparing measured data against predictions from our four proposed ML models: LSTM, MLP, SVM, and RF. Each model's predicted performance is compared against the actual measured PCE data, highlighting how well each ML technique can capture the degradation trends over time.

The LSTM model follows the general trend of the measured PCE values closely throughout most of the time series. It captures both the initial sharp decline and the steady long-term degradation. While LSTM maintains a reasonable fit across the entire timeline, it slightly underestimates the values toward the later stages of degradation, particularly between 5,000 and 6,000 h.

The MLP model shows a similar fit to LSTM in the early stages (up to around 2,000 h). However, it slightly overestimates the PCE values in the mid-range (2,000 to 2,500 h). MLP's performance declines more noticeably in the long-term prediction (after 6,000 h), where it fails to fully capture the rate of degradation.

The SVM model performs well during the early stage (up to 200 h), tracking the measured data relatively closely. However, it shows a notable overestimation in the early to mid-range (between 200 and 1,500 h) and then diverges slightly after 6,000 h, where it underestimates the PCE values.

The RF model diverges significantly from the measured data, particularly in the early stages of degradation (up to around 500 h), where it shows an oversimplified, flat prediction. RF is unable to capture the fluctuations in the data and performs the worst among the models, particularly in the early and mid-range, although it improves slightly in the later stages.

LSTM continues to be one of the best-performing models for time-series prediction in this study, particularly because it can handle sequential data and retain information over long periods. Its ability to model long-term dependencies makes it effective in capturing the overall degradation trend, although some fine-tuning might be necessary for more accurate long-term predictions.

The MLP model captures the general trend of the PCE degradation but struggles with accuracy over longer time intervals. MLP, being a simpler feedforward neural network, lacks the temporal dynamics handling capabilities of LSTM, which explains its weaker performance in the later stages.

SVM performs reasonably well during the initial stage but struggles with capturing the full complexity of the degradation in the later stages. This model, while effective in capturing some general patterns, may not be flexible enough for long-term time-series forecasting with fluctuating patterns.

RF, as expected, performs the worst in this context, particularly due to its limitations in capturing the sequential and temporal dynamics inherent in time-series data. It oversimplifies the degradation pattern and fails to adapt to the fluctuations in the measured data, especially in the early and mid-range.

Furthermore, Table 3 compares the measured and predicted PCE data for an iOSC cell using different ML models, including LSTM, MLP, SVM, and RF, across MSE, rMSE, rrMSE, RAE, RSE, RSS and R-squared

evaluation metrics. The MSE and rMSE values indicate that LSTM provides the closest fit to the measured data, with the lowest error (MSE = 0.0541, rMSE = 0.2223) and the highest R-squared value of 0.9835. This suggests that LSTM effectively captures the long-term degradation pattern of the PCE values, making it the most accurate model. The RF, on the other hand, shows the highest error values (MSE = 0.1610, rMSE = 0.3723) and the lowest R-value of 0.9274, indicating that it struggles to generalize over the lifetime of the iOSC cell, likely due to its inability to capture the sequential nature of the data effectively.

From a broader perspective, the results highlight the strengths of LSTM and MLP models in time-series predictions, where long-term dependencies and temporal trends are crucial, such as in the degradation behavior of PCE over time. The SVM and RF, although capable of capturing general trends, perform less favorably because they lack mechanisms to retain memory over long sequences. This suggests that for predicting the performance and degradation of OPV cells over their lifetime, models like LSTM that can model sequential dependencies are more effective compared to non-sequential techniques like SVM and RF, which are better suited for tasks with less temporal complexity.

Furthermore, Table 4 summarizes the 95% confidence intervals (CIs) and coverage probabilities of four LSTM, MLP, SVM, and RF models for predicting PCE in iOSCs. Each model's performance is evaluated based on the mean lower bound, mean upper bound, mean CI width, and coverage probability.

The LSTM model demonstrates the narrowest CI with a mean CI width of 0.5035. This indicates that LSTM's predictions are relatively more confident compared to the other models. Despite this narrow width, LSTM maintains a high coverage probability of 96.27%, which means the model's CIs capture almost all actual PCE values. This balance of narrow CIs and high coverage suggests that LSTM is well-suited to making precise PCE predictions while maintaining high reliability. The LSTM model's ability to capture temporal dependencies in the data likely contributes to this strong performance, particularly in modeling sequential data.

The MLP and SVM models show a slightly wider mean CI width of 0.5919 and 0.5916, respectively, compared to LSTM. Their coverage probabilities of 95.65% and 95.03% are also slightly lower but still respectable, indicating that MLP and SVM remain reliable models for PCE prediction. Their wider CIs suggest that they are slightly less confident in their predictions, which may be due to the models' simpler architectures compared to LSTM. They lack to capture time dependencies, which could explain the minor drop in certainty. However, they still manage to perform reasonably well in terms of coverage.

The RF model presents the widest mean CI width, at 1.8229, indicating the highest level of uncertainty in its predictions. Its coverage probability of 94.41% is still respectable, but the much wider CIs suggest that RF is less certain about its predictions compared to the other models. The broader intervals are likely a result of the model's ensemble nature, which combines multiple decision trees to make predictions. This ensemble method introduces more variability in the predictions, leading to higher uncertainty. However, RF's ability to maintain reasonable coverage suggests that while it is cautious in its predictions (as shown by the wide CIs), it still captures a significant portion of the actual PCE values.

Figure 21 shows the predicted vs. actual PCE values with the 95% bootstrapped CIs provide a visual representation of each model's uncertainty and performance. These graphs further support the insights gained from the table, offering a clear view of how well each model captures uncertainty.

The LSTM graph visually confirms the model's high confidence in its predictions, with narrow CIs closely surrounding the predicted values. Most of the actual PCE values fall well within the CIs, supporting the high coverage probability of 96.27%. The narrow CIs reflect LSTM's ability to make precise predictions, likely due to its capacity to learn complex temporal patterns in the data. This makes LSTM a highly reliable model for PCE prediction, especially when prediction accuracy is critical.

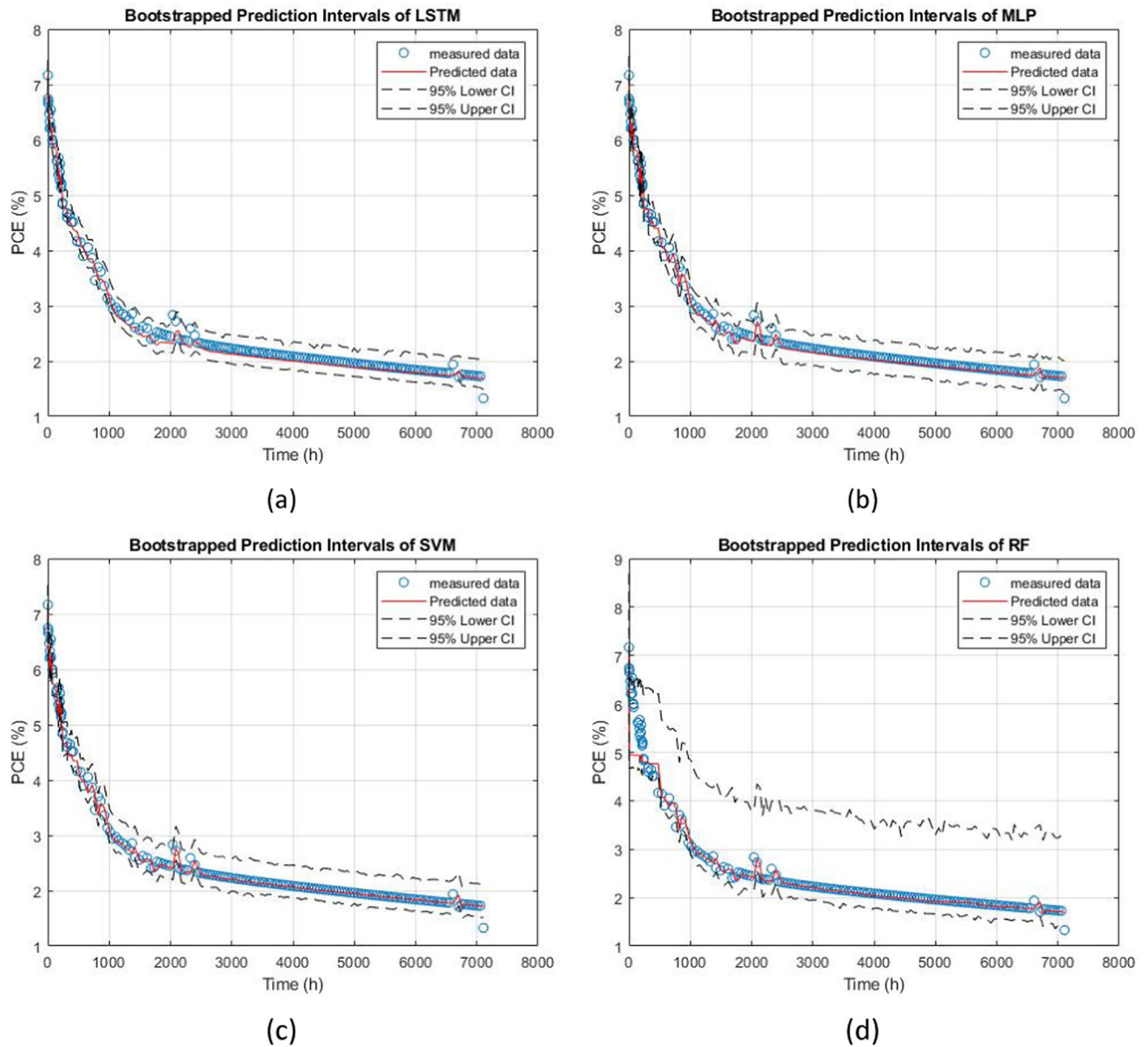
The MLP and SVM graphs show slightly wider confidence intervals compared to LSTM, which is consistent with the table results. While the CIs are somewhat broader, the graphs still indicate that most actual PCE values are captured within the prediction intervals, reflecting the models' coverage of 95.65% and 95.03 for MLP and SVM, respectively. The slightly lower confidences compared to LSTM are evident in the slightly wider bands, which may be attributed to MLP's and SVM simpler architecture, making them less adept at capturing subtle patterns in the data. Nonetheless, these models still perform reasonably well, providing reliable PCE predictions.

The RF graph displays the widest CIs among the four models, reflecting the high level of uncertainty in RF's predictions. The wide intervals suggest that RF is more conservative in its predictions, providing a broader range to account for potential variability in the data. This is consistent with the table's result of the largest mean CI width (1.8229) for RF. Despite the broader intervals, the RF model still captures a reasonable percentage of actual PCE values (94.41% coverage), indicating that while the model is cautious, it is still able to provide relatively reliable predictions.

The results of our study indicate a clear advantage of using LSTM and MLP neural networks over SVM and RF for predicting the PCE of iOSCs. The LSTM model outperformed all the other reported models in this study

Model	Mean lower bound	Mean upper bound	Mean CI width	Coverage probability (%)
LSTM	2.6781	3.1816	0.5035	96.2733
MLP	2.6071	3.1990	0.5919	95.6522
SVM	2.6693	3.2609	0.5916	95.0311
RF	2.4987	4.3216	1.8229	94.4099

**Table 4.** A comparison between LSTM, MLP, SVM, and RF Models for PCE Prediction in terms of Bootstrapped Confidence Intervals Statistics.



**Fig. 21.** Predicted vs. actual PCE with bootstrapped confidence intervals for (a) LSTM, (b) MLP, (c) SVM and (d) RF.

for all the used performance metrics. This significant performance difference can be attributed to its architecture, which is specifically designed to retain information over long sequences. This capability is particularly beneficial for time-series data, where the efficiency of a solar cell can change over time due to various factors such as material degradation, environmental conditions, and operational stresses. By capturing these long-term dependencies, the LSTM model can provide more accurate predictions, which is evident from the lower error metrics compared to other studied prediction models.

The MLP model ranked second with a slight difference after the LSTM and outperformed SVM and RF models in this study for all the used performance metrics. This performance underscores the ability of MLP networks to handle good prediction capabilities for complex systems and patterns due to its ability to learn and store knowledge (in the form of weights) for future predictions. MLPs are highly effective for prediction tasks due to their ability to model complex non-linear relationships in data. In addition, MLP layers and non-linear activation functions enable it to capture intricate dependencies between input features and outputs.

Furthermore, the search for optimal parameters of the double exponential model using the Levenberg–Marquardt optimization algorithm is highly efficient and outperforms the traditional double exponential model that relies on the Nonlinear Least Squares optimization algorithm. The Levenberg–Marquardt algorithm combines the benefits of both gradient descent and the Gauss–Newton method, providing faster convergence and better handling of complex, non-linear relationships. This results in more accurate parameter estimation and improved model performance in fitting data compared to the Nonlinear Least Squares approach.

The application of advanced AI techniques such as LSTM and MLP in predicting the PCE of organic solar cells represents a significant advancement in the field. Traditional methods of optimizing PCE are labour-intensive and time-consuming, often relying on trial-and-error approaches. By leveraging AI, researchers can predict the efficiency of various material and device configurations before physical testing, thus saving considerable time and resources.

The findings of this study highlight the potential of AI to accelerate the development and optimization of organic solar cells. The high accuracy of the LSTM and MLP models, in particular, suggests that AI can provide valuable insights into the factors that influence PCE, guiding researchers towards the most promising materials and configurations. This predictive capability is crucial for advancing the efficiency and commercial viability of organic solar cells, ultimately contributing to the broader goal of sustainable energy generation.

### Future research directions

Despite the promising results, there are several areas for future research that could further improve the prediction accuracy and utility of AI models in this context:

**Integration of environmental data:** Incorporating environmental data such as temperature, humidity, and exposure to pollutants could provide a more comprehensive understanding of the factors affecting PCE and enhance the model's predictive accuracy.

**Long-term stability data:** Including data on the long-term stability and degradation of solar cells can help in predicting the lifespan and performance of solar cells over extended periods, providing more valuable insights for researchers and manufacturers.

**Advanced AI techniques:** Exploring other advanced AI techniques, such as convolutional neural networks (CNNs) or hybrid models that combine multiple AI approaches, could potentially yield even more accurate predictions.

**Material and structural innovations:** Research into new materials and innovative structural designs for organic solar cells can be guided by AI predictions, accelerating the discovery of high-efficiency configurations.

**Real-world testing and validation:** Implementing real-world testing and validation of AI predictions can help refine the models and ensure their practical applicability in diverse environmental conditions and usage scenarios.

### Conclusion

This research demonstrates that both LSTM and MLP neural networks are viable tools for predicting the PCE of iOSCs. The application of these AI techniques can significantly accelerate the optimization of organic solar cells, reducing the reliance on extensive experimental trials. By leveraging the predictive capabilities of AI, researchers can focus their efforts on the most promising materials and device configurations, ultimately advancing the development of high-efficiency organic solar cells.

The findings underscore the potential of artificial intelligence in transforming the research and development landscape of organic photovoltaics. The high accuracy achieved by the LSTM and MLP models, with  $R^2$  values of 0.9675 and 0.9562, respectively, suggests that temporal dependencies and nonlinear relationships in the data can be effectively captured, leading to precise predictions. This is particularly beneficial for long-term performance assessment and optimization of solar cells, where traditional methods fall short due to their labour-intensive and time-consuming nature.

### Data Availability Statement

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 5 August 2024; Accepted: 21 October 2024

Published online: 29 October 2024

### References

- Mohtasham, J. Renewable energies. *Energy Procedia* **74**, 1289–1297. <https://doi.org/10.1016/j.egypro.2015.07.774> (2015).
- Perez, M. & Perez, R. Update 2022 – a fundamental look at supply side energy reserves for the planet. *Sol. Energy Adv.* **2**, 100014. <https://doi.org/10.1016/j.seja.2022.100014> (2022).
- Liu, Q. et al. 18% Efficiency organic solar cells. *Sci. Bull.* **65**(4), 272–275. <https://doi.org/10.1016/j.scib.2020.01.001> (2020).
- National Renewable Energy Laboratory, “Best Research-Cell Efficiencies.” Accessed: Sep. 27, 2024. <https://www.nrel.gov/pv/cell-efficiency.html>
- Guermoui, M. & Rabehi, A. Soft computing for solar radiation potential assessment in Algeria. *Int. J. Ambient Energy* **41**(13), 1524–1533 (2020).
- Guermoui, M., Boland, J. & Rabehi, A. On the use of BRL model for daily and hourly solar radiation components assessment in a semiarid climate. *Eur. Phys. J. Plus* **135**(2), 1–16 (2020).
- Rabehi, A., Amrani, M., Benamara, Z., Akkal, B. & Kacha, A. H. Electrical and photoelectrical characteristics of Au/GaN/GaAs Schottky diode. *Optik* **127**(16), 6412–6418 (2016).
- Baitiche, O., Bendelala, F., Chekane, A., Rabehi, A. & Comini, E. Numerical modeling of hybrid solar/thermal conversion efficiency enhanced by metamaterial light scattering for ultrathin PbS QDs-STPV cell. *Crystals* **14**(7), 668 (2024).
- Cheng, P. et al. Efficient and stable organic solar cells: Via a sequential process. *J. Mater. Chem. C* **4**(34), 8086–8093. <https://doi.org/10.1039/c6tc02338j> (2016).
- Sacramento, A., Balderrama, V. S., Ramírez-Como, M., Marsal, L. F. & Estrada, M. Degradation study under air environment of inverted polymer solar cells using polyfluorene and halide salt as electron transport layers. *Sol. Energy* **198**, 419–426. <https://doi.org/10.1016/j.solener.2020.01.071> (2020).

11. Balderrama, V. S. et al. Degradation of electrical properties of PTB1:PCBM solar cells under different environments. *Sol. Energy Mater. Sol. Cells* **125**, 155–163. <https://doi.org/10.1016/j.solmat.2014.02.035> (2014).
12. Bouabdelli, M. W., Rogti, F., Maache, M. & Rabehi, A. Performance enhancement of CIGS thin-film solar cell. *Optik* **216**, 164948 (2020).
13. Chen, L. X. Organic solar cells: Recent progress and challenges. *ACS Energy Lett.* **4**(10), 2537–2539. <https://doi.org/10.1021/acsenerylett.9b02071> (2019).
14. Eibeck, A. et al. Predicting power conversion efficiency of organic photovoltaics: Models and data analysis. *ACS Omega* **6**(37), 23764–23775. <https://doi.org/10.1021/acsomega.1c02156> (2021).
15. Koster, L. J. A., Mihailetschi, V. D., Ramaker, R. & Blom, P. W. M. Light intensity dependence of open-circuit voltage of polymer:fullerene solar cells. *Appl. Phys. Lett.* **86**(12), 1–3. <https://doi.org/10.1063/1.1889240> (2005).
16. Hossain, N., Das, S. & Alford, T. L. Equivalent circuit modification for organic solar cells. *Circuits Syst.* **06**(06), 153–160. <https://doi.org/10.4236/cs.2015.66016> (2015).
17. Li, Y. et al. Recent progress in organic solar cells: A review on materials from acceptor to donor. *Molecules* **27**(6), 1800. <https://doi.org/10.3390/molecules27061800> (2022).
18. Souahlia, A., Belatreche, A., Benyettou, A. & Curran, K. Blood vessel segmentation in retinal images using echo state networks. *9th Int. Conf. Adv. Comput. Intell. ICACI 2017*, 91–98. <https://doi.org/10.1109/ICACI.2017.7974491> (2017).
19. Souahlia, A., Rabehi, A. & Rabehi, A. Hybrid models for daily global solar radiation assessment. *J. Eng. Exact Sci.* **9**(4), 1–19. <https://doi.org/10.18540/jcecvl9iss4pp15926-01e> (2023).
20. Bouchakour, A. et al. MPPT algorithm based on metaheuristic techniques (PSO & GA) dedicated to improve wind energy water pumping system performance. *Sci. Rep.* **14**(1), 17891 (2024).
21. Lazcano, A., Jaramillo-Morán, M. A. & Sandubete, J. E. Back to basics: The power of the multilayer perceptron in financial time series forecasting. *Mathematics* **12**(12), 1–18. <https://doi.org/10.3390/math12121920> (2024).
22. A. H. Elsheikh, S. W. Sharshir, M. Abd Elaziz, A. E. Kabeel, W. Guilan, and Z. Haiou, “Modeling of solar energy systems using artificial neural network: A comprehensive review,” *Sol. Energy*, vol. 180, no. January, pp. 622–639, 2019, <https://doi.org/10.1016/j.solener.2019.01.037>.
23. Sahu, H., Rao, W., Troisi, A. & Ma, H. Toward predicting efficiency of organic solar cells via machine learning and improved descriptors. *Adv. Energy Mater.* **8**(24), 1–27. <https://doi.org/10.1002/aenm.201801032> (2018).
24. Gottschalg, R., Rommel, M., Infield, D. G., & Ryssel, H. Comparison of different methods for the parameter determination of the solar cells double exponential equation. in *14th Eur. Photovolt. Sol. Energy Conf.*, no. January, pp. 321–324 (1997).
25. Bendaoud, R. et al. Validation of a multi-exponential alternative model of solar cell and comparison to conventional double exponential model. *Proc. Int. Conf. Microelectron. ICM* <https://doi.org/10.1109/ICM.2015.7438053> (2016).
26. Balderrama, V. S. et al. High-efficiency organic solar cells based on a halide salt and polyfluorene polymer with a high alignment-level of the cathode selective contact. *J. Mater. Chem. A* **6**(45), 22534–22544. <https://doi.org/10.1039/c8ta05778h> (2018).
27. Lastra, G. et al. High-performance inverted polymer solar cells: Study and analysis of different cathode buffer layers. *IEEE J. Photovoltaics* **8**(2), 505–511. <https://doi.org/10.1109/JPHOTOV.2017.2782568> (2018).
28. Yanagidate, T. et al. Flexible PTB7:PC71BM bulk heterojunction solar cells with a LiF buffer layer. *Jpn. J. Appl. Phys.* <https://doi.org/10.7567/JJAP.53.02BE05> (2014).
29. Sacramento, A. et al. Inverted polymer solar cells using inkjet printed ZnO as electron transport layer: Characterization and degradation study. *IEEE J. Electron Devices Soc.* **8**, 413–420. <https://doi.org/10.1109/JEDS.2020.2968001> (2020).
30. Mbilo, M. et al. Highly efficient and stable organic solar cells with SnO<sub>2</sub> electron transport layer enabled by UV-curing acrylate oligomers. *J. Energy Chem.* **92**, 124–131. <https://doi.org/10.1016/j.jechem.2024.01.022> (2024).
31. Sacramento, A. et al. Comparative degradation analysis of V<sub>2</sub>O<sub>5</sub>, MoO<sub>3</sub> and their stacks as hole transport layers in high-efficiency inverted polymer solar cells. *J. Mater. Chem. C* **9**(20), 6518–6527. <https://doi.org/10.1039/d1tc00219h> (2021).
32. Krishna, B. G., Ghosh, D. S. & Tiwari, S. Hole and electron transport materials: A review on recent progress in organic charge transport materials for efficient, stable, and scalable perovskite solar cells. *Chem. Inorg. Mater.* **1**, 100026. <https://doi.org/10.1016/j.cinorg.2023.100026> (2023).
33. Sanchez, J. G. et al. Effects of annealing temperature on the performance of organic solar cells based on polymer: Non-fullerene using V<sub>2</sub>O<sub>5</sub> as HTL. *IEEE J. Electron Devices Soc.* **8**, 421–428. <https://doi.org/10.1109/JEDS.2020.2964634> (2020).
34. Hou, J. & Guo, X. Active layer materials for organic solar cells. *Green Energy Technol.* **128**, 17–42. [https://doi.org/10.1007/978-1-4471-4823-4\\_2](https://doi.org/10.1007/978-1-4471-4823-4_2) (2013).
35. Daniel, S. G., Devu, B. & Sreekala, C. O. Active layer thickness optimization for maximum efficiency in bulk heterojunction solar cell. *IOP Conf. Ser. Mater. Sci. Eng.* **1225**(1), 012017. <https://doi.org/10.1088/1757-899x/1225/1/012017> (2022).
36. Ramirez-Como, M., Balderrama, V. S., Estrada, M. Performance parameters degradation of inverted organic solar cells exposed under solar and artificial irradiance, using PTB7:PC70BM as active layer. in *2016 13th International Conference on Electrical Engineering, Computing Science and Automatic Control, CCE 2016*, IEEE, 2016, pp. 1–5. <https://doi.org/10.1109/ICEEE.2016.7751205>.
37. Green, M. A. et al. Solar cell efficiency tables (Version 63). *Prog. Photovoltaics Res. Appl.* **32**(1), 3–13. <https://doi.org/10.1002/pip.3750> (2024).
38. Cui, Y. et al. Over 16% efficiency organic photovoltaic cells enabled by a chlorinated acceptor with increased open-circuit voltages. *Nat. Commun.* <https://doi.org/10.1038/s41467-019-10351-5> (2019).
39. Li, Z., Yang, J. & Dezfuli, P. A. N. Study on the influence of light intensity on the performance of solar cell. *Int. J. Photoenergy* **2021**(1), 1–10. <https://doi.org/10.1155/2021/6648739> (2021).
40. Ghorab, M., Fattah, A. & Joodaki, M. Fundamentals of organic solar cells: A review on mobility issues and measurement methods. *Optik (Stuttg)* **267**(2022), 169730. <https://doi.org/10.1016/j.ijleo.2022.169730> (2022).
41. Peters, C. H. et al. The mechanism of burn-in loss in a high efficiency polymer solar cell. *Adv. Mater.* **24**(5), 663–668. <https://doi.org/10.1002/adma.201103010> (2012).
42. Upama, M. B. et al. Organic solar cells with near 100% efficiency retention after initial burn-in loss and photo-degradation. *Thin Solid Films* **636**, 127–136. <https://doi.org/10.1016/j.tsf.2017.05.031> (2017).
43. Osorio, E. et al. Degradation analysis of encapsulated and nonencapsulated TiO<sub>2</sub>/PTB7:PC70BM/V<sub>2</sub>O<sub>5</sub> solar cells under ambient conditions via impedance spectroscopy. *ACS Omega* **2**(7), 3091–3097. <https://doi.org/10.1021/acsomega.7b00534> (2017).
44. Norrman, K. & Krebs, F. C. Degradation and stability of R2R manufactured polymer solar cells. *Org. Photovoltaics X* **7416**, 49–54. <https://doi.org/10.1117/12.833329> (2009).
45. Kruse, R., Mostaghim, S., Borgelt, C., Braune, C., Steinbrecher, M. Multi-layer perceptrons. in *Computational intelligence: A methodological introduction*, Cham: Springer International Publishing, 2022, pp. 53–124. [https://doi.org/10.1007/978-3-030-42227-1\\_5](https://doi.org/10.1007/978-3-030-42227-1_5).
46. Kaya, M. & Hajimirza, S. Application of artificial neural network for accelerated optimization of ultra thin organic solar cells. *Sol. Energy* **165**, 159–166. <https://doi.org/10.1016/j.solener.2018.02.062> (2018).
47. Seifrid, M. et al. Beyond molecular structure: Critically assessing machine learning for designing organic photovoltaic materials and devices. *J. Mater. Chem. A* **12**(24), 14540–14558. <https://doi.org/10.1039/d4ta01942c> (2024).
48. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D Nonlinear Phenom.* **404**, 132306. <https://doi.org/10.1016/j.physd.2019.132306> (2020).

49. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735> (1997).
50. Elmoaqet, H., Eid, M., Glos, M., Ryalat, M. & Penzel, T. Deep recurrent neural networks for automatic detection of sleep apnea from single channel respiration signals. *Sensors* **20**(18), 1–19. <https://doi.org/10.3390/s20185037> (2020).
51. Kasthuri, E. & Balaji, S. Natural language processing and deep learning chatbot using long short term memory algorithm. *Mater. Today Proc.* **81**(2), 690–693. <https://doi.org/10.1016/j.matpr.2021.04.154> (2021).
52. Lindemann, B., Müller, T., Vietz, H., Jazdi, N. & Weyrich, M. A survey on long short-term memory networks for time series prediction. *Procedia CIRP* **99**, 650–655. <https://doi.org/10.1016/j.procir.2021.03.088> (2021).
53. Zhou, N. R., Zhou, Y., Gong, L. H. & Jiang, M. L. Accurate prediction of photovoltaic power output based on long short-term memory network. *IET Optoelectron.* **14**(6), 399–405. <https://doi.org/10.1049/iet-opt.2020.0021> (2020).
54. Sarmas, E., Spiliotis, E., Stamatopoulos, E., Marinakis, V. & Doukas, H. Short-term photovoltaic power forecasting using meta-learning and numerical weather prediction independent long short-term memory models. *Renew. Energy* **216**, 118997. <https://doi.org/10.1016/j.renene.2023.118997> (2023).
55. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32. <https://doi.org/10.1023/a:1010933404324> (2001).
56. Liu, C., Chan, Y., AlamKazmi, S. H. & Fu, H. Financial fraud detection model: Based on random forest. *Int. J. Econ. Financ.* <https://doi.org/10.5539/ijef.v7n7p178> (2015).
57. Wang, L., Yang, M. Q. & Yang, J. Y. Prediction of DNA-binding residues from protein sequence information using random forests. *BMC Genomics* **10**(SUPPL. 1), 1–9. <https://doi.org/10.1186/1471-2164-10-S1-S1> (2009).
58. Reese, M. O. et al. Consensus stability testing protocols for organic photovoltaic materials and devices. *Sol. Energy Mater. Sol. Cells* **95**(5), 1253–1267. <https://doi.org/10.1016/j.solmat.2011.01.036> (2011).
59. Teta, A. et al. Fault detection and diagnosis of grid-connected photovoltaic systems using energy valley optimizer based lightweight CNN and wavelet transform. *Sci. Rep.* **14**(1), 18907 (2024).
60. Ladjal, B., Tibermacine, I. E., Bechouat, M., Sedraoui, M., Napoli, C., Rabehi, A., & Lalmi, D. Hybrid models for direct normal irradiance forecasting: a case study of Ghardaia zone (Algeria). *Natural Hazards*, 1–23(2024).
61. El-Amarty, N. et al. A new evolutionary forest model via incremental tree selection for short-term global solar irradiance forecasting under six various climatic zones. *Energy Convers. Manage.* **310**, 118471 (2024).
62. Guermoui, M. et al. An analysis of case studies for advancing photovoltaic power forecasting through multi-scale fusion techniques. *Sci. Rep.* **14**(1), 6653 (2024).
63. Khelifi, R. et al. Short-term PV power forecasting using a hybrid TVF-EMD-ELM strategy. *Int. Trans. Electr. Energy Syst.* **2023**(1), 6413716 (2023).
64. Rabehi, A., Rabehi, A. & Guermoui, M. Evaluation of different models for global solar radiation components assessment. *Appl. Solar Energy* **57**, 81–92 (2021).

## Acknowledgements

The authors extend their appreciation to Taif University, Saudi Arabia, for supporting this work through the project number (TU-DSPP-2024-14), also this article has been produced with the financial support of the European Union under the REFRESH – Research Excellence For Region Sustainability and High-tech Industries project number CZ.10.03.01/00/22\_003/0000048 via the Operational Programme Just Transition, project TN02000025 National Centre for Energy II and ExPEDite project a Research and Innovation action to support the implementation of the Climate Neutral and Smart Cities Mission project. ExPEDite receives funding from the European Union's Horizon Mission Programme under grant agreement No. 101139527.

## Author contributions

Mustapha Marzouglal, Abdelkerim Souahlia, Lakhdar Bessissa, Djillali Mahi, Abdelaziz Rabehi, Yahya Z. Alharthi, Amanuel Kumsa Bojer, Ayman Flah: Conceptualization, Methodology, Software, Visualization, Investigation, Writing- Original draft preparation. Mosleh M. Alharthi, Sherif S. M. Ghoneim: Data curation, Validation, Supervision, Resources, Writing—Review & Editing, Project administration.

## Funding

This research was funded by Taif University, Taif, Saudi Arabia, Project No. (TU-DSPP-2024–14).

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to A.R. or A.K.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024