Transactions on Machine Learning and Artificial Intelligence - Vol. 10, No. 6
Publication Date: December, 25, 2022
DOI:10.14738/tmlai.106.13645.
Nansamba, S., & Harb, H. (2022). Developing a Neural Network Based Fault Prediction Tool for a Solar Power Plant in Uganda. Transactions on Machine Learning and Artificial Intelligence, 10(6). 52-70.

Developing a Neural Network Based Fault Prediction Tool for a Solar Power Plant in Uganda

Salmah Nansamba
Engineering Institute of Technology, Perth, WA 6005, Australia

Hadi Harb
Engineering Institute of Technology, Perth, WA 6005, Australia

ABSTRACT

Solar photovoltaic (PV) systems are one of the fastest growing renewable energy technologies and plenty of research has been and continues to be carried out in this domain. Maximization of solar PV power plant production, efficiency and return on investment can only be achieved by having adequate and effective maintenance systems in place. Of the various maintenance schemes, predictive maintenance is popular for its effectiveness and minimization of resource wastage. Maintenance activities are scheduled based on the real time condition of the system with priority being given to the system components with the highest likelihood of failure. A good predictive maintenance system is based on the premise of being able to anticipate faults before they occur. In this study therefore, a fault prediction tool for a solar plant in Uganda is proposed. The hybrid tool is developed using both feed forward and long short term memory neural networks for power prediction, in conjunction with a mean chart statistical process control tool for final fault prediction. Results from the study demonstrate that the feed forward and long short term memory neural network modules of the proposed tool attain mean absolute errors of 4.2% and 6.9% respectively for power production predictions. The fault prediction capability of the tool is tested under both normal and abnormal operating conditions. Results show that the tool satisfactorily discriminates against the fault and non-fault conditions thereby achieving successful solar PV system fault prediction.

Keywords: Machine learning; Feedforward Neural Networks; Long Short Term Memory; Solar fault prediction.

INTRODUCTION

There is an ever growing shift from the use of conventional power sources to renewables such as wind and solar energy sources. Photovoltaic has an ever replenishing, vast and clean energy supply and is therefore the most promising renewable energy source [1]. Solar photovoltaic systems are subject to various environmental constraints that result in faults. These faults can potentially lower the annual power generated by a PV system by about 18.9% [1]. It is therefore important that adequate maintenance is done in solar PV systems so as to maximize energy production and the use of these systems till their end of life thereby ensuring the highest return on investment. Maintenance strategies are crucial towards maintaining the technical and financial performance of solar photovoltaic systems. The early detection and prediction of

faults in photovoltaic systems is crucial towards increasing the reliability, efficiency and lifetime of solar photo voltaic systems [1]. The lack of a predictive component in some maintenance systems however is a hindrance to the minimization of downtime costs. In Uganda particularly, grid connected solar PV generation is on the rise, with four solar plants contributing up to 50MW annually, equivalent to 21.277 GWh [2]. There is still a lot of potential for growth, it is therefore prudent that the existing solar PV systems are adequately maintained to ensure efficient performance. The existing solar power plants have Supervisory Control and Data Acquisition (SCADA) systems that monitor crucial plant parameters such as AC power output, string current and voltage and module temperature.

Without automatic analysis of these huge volumes of important collected data in order to predict faults, there remains a limitation in the development of the most effective predictive maintenance work plans for the solar plants. This is a gap that needs to be filled and as such forms the basis of this research project. In this paper, a fault prediction system for a solar power plant is proposed. The system is based on machine learning techniques, namely Feedforward Neural Networks and Long Short Term Memory.

FAULT DETECTION IN SOLAR PV SYSTEMS

Faults in Solar PV Systems

Faults in solar photovoltaic systems are failures that happen in the solar modules, electrical conductors and other system components such as the inverters. Generally, the operating lifetime of photovoltaic panels is set to 25 years by manufacturers. However, decline in power output performance is usually observed before this time elapses and is attributed to the faults that occur in the system [3]. During the course of operation, the photovoltaic system is exposed to unfavourable conditions; solar cells may experience overheating owing to the higher ambient temperatures, dirt and droppings on solar modules can cause shading that results into hotspot problems in solar modules. Inverters on the other hand, owing to their major electrical component composition are prone to sudden total failure [4]. These resulting solar system faults not only lower the PV system efficiency but also its useful life. A brief discussion of PV system faults as presented in [5] follows.

Arc faults

Arc faults typically occur when fault current is discharged into air and other dielectrics. This can be as a result of a break in the electrical conductors which causes discontinuity in the electrical circuit. It can also be as a result of insulation damage which leads to conductors of different voltages being in close proximity with each other. Arc faults can potentially destroy entire PV strings and also cause electrical fires [5].

Line to line faults

Line to line faults occur when there is an accidental connection between two points at different potentials in a PV string. They can also arise as a result of the creation of an undesired path of low impedance between two points. This can be due to various causes for example conductor aging and DC connector damage. Line to line faults are dangerous because they not only destroy PV strings and the system conductors but are also potential sources of electrical fires [4] [5] [6] [7].

URL: http://dx.doi.org/10.14738/tmlai.106.13645

Ground faults

Also referred to as zero efficiency faults, ground faults are the commonest fault type in electrical systems. They occur as a result of breakdown in insulation of the equipment, which causes leakage currents to flow to the ground [7].

Hot spots

Photovoltaic hotspots occur when there is reverse bias operation of one or more solar cells as a consequence of solar cell degradation and cell mismatch. The cells then dissipate the power produced instead of transmitting it and this eventually causes abnormally high cell temperatures. Consequently, the output power of the entire PV module gradually declines. Hotspots often result in open circuits and damage of the affected solar cells [5] [8].

Diode faults

Bypass diodes serve to facilitate the flow of current around cracked or shaded solar cells thereby maintaining optimal power production by the cell. When these diodes are not exposed to sufficient air flow to enable cooling, they undergo thermal stress which makes them defective [9].

PV array faults

The faults in the PV panels themselves can be temporary or permanent. Temporary faults occur as a result of the accumulation of dirt and dust on the panels while some of the permanent defects' causes include burnt solar cells, cracking and yellowing of the cells, bubbles and delamination. Normal production in the event of permanent faults can be restored by replacing the faulty modules. Temporal faults on the other hand do not always necessitate replacement of the affected components; for example, cleaning the dirty solar panels can be effective in restoring optimum power production [5].

Machine Learning

Machine learning is a subset of artificial intelligence approaches that is focused on training computer systems to gain the ability to learn from data and make meaningful predictions and classifications. No explicit programming is required for machine learning. The resulting algorithms have the ability to adapt to new situations in form of new data that is presented to them and this facilitates the making of more accurate classifications and predictions. Machine learning is classified into supervised and unsupervised approaches. Under supervised learning, labelled datasets are used to train the algorithms while unlabelled data sets are used in training for unsupervised learning. Both traditional machine learning and deep learning techniques find a lot of application in the photovoltaic arena such as in maximum power point tracking, photovoltaic design parameter extraction, modelling of photovoltaic modules, solar irradiance, solar power production forecasting, photovoltaic storage system energy management and anomaly detection [10]. This study is focused on the application of machine learning to solar power production and fault prediction.

Predictive Maintenance in Solar PV Systems

Preventive maintenance is about mitigating system failure by intervening to equipment before they fail. It is hinged on the premise that before a system component actually fails, it bears signs of deterioration; whether visible or not. Focus is therefore on the ability to perceive these signs and this is accomplished by monitoring and analysing defined performance parameters in

respect to system degradation [7]. Machine learning is one of the approaches used to evaluate the performance of photovoltaic systems for possible degradation. Techniques like artificial neural network and long short term memory (LSTM) models are used to study the relationship and effect of weather parameters like temperature, irradiance and wind on PV power production [11]. Based on this, power production predictions and forecasts are made. These models are also used in making predictions of the weather parameters. Accurate weather parameter predictions improve the accuracy of the power predictions. Machine learning fault prediction models have the capability to pick up on anomalies in the solar PV system. In the event that the model output is less than the expected, this can be an indication of an issue such as dirt on the solar panels that should be removed during maintenance works so as to restore optimum power production.

LITERATURE REVIEW

Photovoltaic panels are subject to changes due to a number of varying ambient weather conditions such as humidity, solar irradiance, dust, wind speed, temperature, rain et al. as well as the inevitable aging and breakdown of individual solar cells. All these factors in turn influence the performance of the photovoltaic system. The fact that there is a lot of variation makes it challenging to set boundaries for fault detection in PV systems. Machine learning is about learning to recognize patterns in data with little to no human interaction [12]'. Machine learning techniques therefore, by their inherent ability to automatically learn and modify given conditions based on availed data sets can be used for self-learning in photovoltaic systems and thus overcome the limitation posed by the immense variability [1].

In [13], an algorithm for detecting anomalies in photovoltaic systems and generating predictive maintenance alerts to inform operators in the maintenance planning process was proposed. Active power produced at the plant, solar irradiance and module temperature parameters were used to train the Artificial Neural Network (ANN) based algorithm. Autocorrelation and Triangular Moving Average (TMA) were used to process the resulting residual vector in order to pick up anomalies and establish long term trends from which future anomalies were identified. Results showed that the model had a validation error of only 2.3% and high prediction accuracy with a positive prediction rate above 90%.

An inverter fault prediction tool with specific class fault and generic fault prediction levels was presented in [14]. Results demonstrated that the Self-Organizing Map (SOM) based supervision diagnosis model component of the tool was effective in predicting high frequency inverter failures up to about 7 days prior, with up to 95% sensitivity and 80% specificity while the pattern recognition neural network based short term fault prediction model component showed effective prediction capabilities for high frequency fault classes to about 7 days prior with its sensitivity was subject to degradation depending on the available fault statistics. Together both models were showed to be effective at predicting faults in PV solar plants, thereby allowing for proactive maintenance.

The authors of [15] proposed a data driven approach to early fault detection and prediction in large PV power plants hinged on a self-organizing map neural network. Supervisory Control and Data Acquisition (SCADA) historical data was used for model development and an original key performance indicator formulated based on the frequency of cells occupancy. Results

URL: http://dx.doi.org/10.14738/tmlai.106.13645

demonstrated the efficiency of the proposed approach in incipient fault prediction of up to an average of seven days prior with true positive rates of up to 95%.

Aziz et al [16] presented a novel approach to detecting and classifying six PV array faults under severe conditions. It was based on deep two dimensional scalogram generation using PV system data and two dimensional convolutional neural networks that applied the AlexNet algorithm for feature extraction and classification. Results from the study showed that the approach achieved a high fault detection accuracy of 73.53% and a best performance of 70.45% in fault classification using representative and discriminative features.

The authors of [17] proposed a method of automatic fault detection in PV arrays based on artificial deep neural networks. The fault types considered in the study included the bypass diode short circuit fault, blocking diode short circuit fault and blocking diode open circuit fault. PV simulation data was used to develop a binary classifier model to detect the presence of a fault in the system and a multiclass classifier to identify the fault type. Results from the study showed that the binary classifier attained an accuracy of 95.55% while the multiclass classifier attained an accuracy of 97.28% resulting in a total system accuracy of 92.95%.

The authors of [18] applied statistical and deep learning methods in the forecasting of solar PV power. Seasonal Autoregressive Integrated Moving Average (SARIMA) and LSTM network models were developed and applied to data for varying weather conditions i.e. sunny, cloudy, partially cloudy, dusty, rainy and foggy days to make hourly forecasts of the PV power. Results from the study indicated that the recurrent neural network models performed better than the statistical models. The bidirectional LSTM model had a better performance than the standard unidirectional LSTM model with a root mean square error (RMSE) value of 0.18 against 0.08 for the bidirectional LSTM under dusty weather conditions.

The reviewed related works applied statistical processing, data analytics and machine learning techniques such as artificial neural networks in the development of models for fault detection, diagnosis, classification and prediction in solar PV and wind energy systems. The reviewed works indicated that neural networks are popular and well suited for building complex systems with nonlinear data. In addition, it was noted that weather conditions, particularly solar irradiance and temperature are the most crucial parameters in power production prediction which is the first step in most of the fault detection approaches. However, a research gap exists as most of the studies were based on data from solar plants in Europe. There is need to consider variation in climate and technology in PV fault prediction. In addition, huge data volumes are required to build fault prediction models of good accuracy and this is a research challenge in some scenarios. Therefore, the proposed research is focused on data specific to a solar PV plant in Uganda which is characterized by tropical climate. In addition, it is intended to generalize the proposed model so that it forms a basis for future model development in related research studies.

METHODOLOGY

Figure 1 summarizes the research methodology.

Data Collection and Preprocessing

The overall performance of solar PV systems is hinged on the quality and quantity of solar irradiation received by the solar modules. This makes sunlight the most important parameter in solar power prediction as without it, no power production can be achieved. Ambient temperature and solar module temperature are the other crucial factors that are crucial towards the performance of solar energy technologies [11]. Historical SCADA data for June, 2020 to June, 2022 was used in the study. Solar irradiance and ambient temperature were used to predict the active power produced by the inverters at the plant. As part of data preprocessing, the data was aggregated by sum from second to hourly intervals to minimize the variations in readings. The data was also inspected for outliers and missing values. Observations at zero irradiance and zero production were removed from the data. Ten sun hours were considered i.e. 09:00 to 18:00 hrs. Only data typical of normal conditions was used for training the models. The data was also normalized by fitting it within the range from zero to one as this is known to enhance machine learning model training. It was also split into 80% training and 20% testing datasets. 10% of the training data was used for model validation.

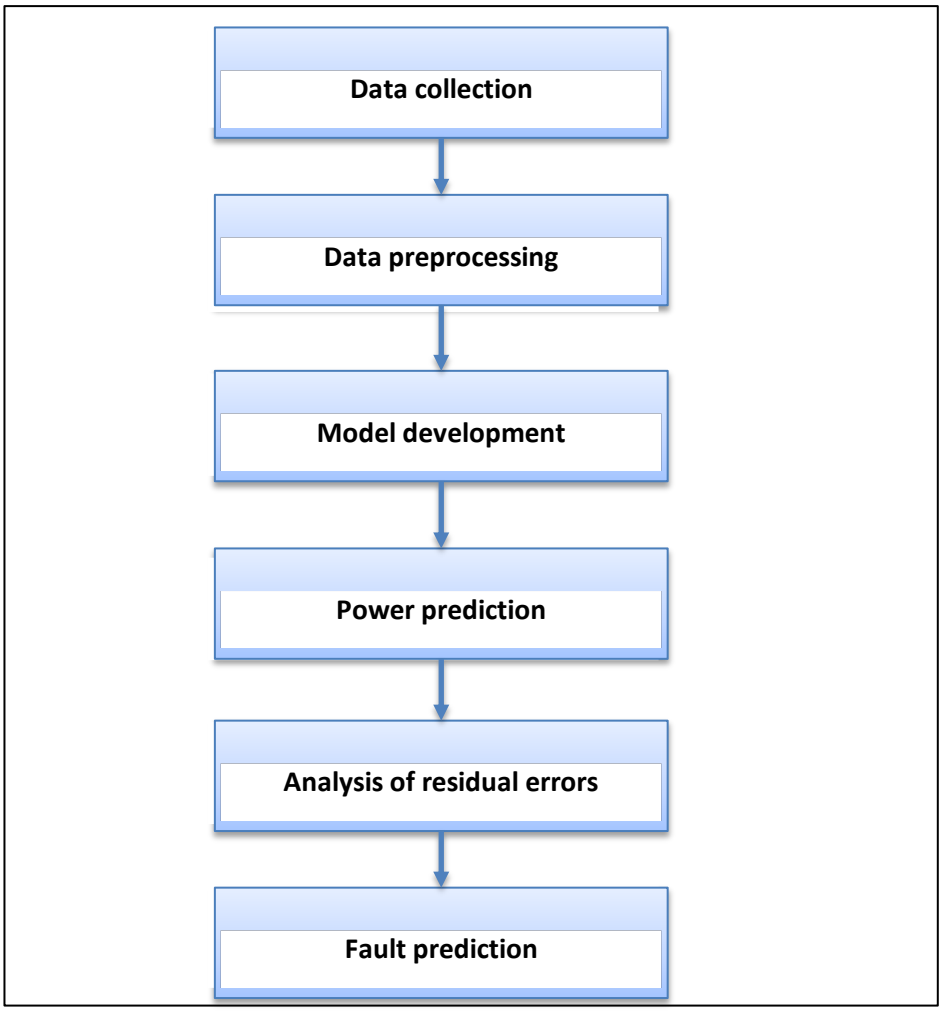


Figure 1. Methodology flow chart.

URL: http://dx.doi.org/10.14738/tmlai.106.13645

Model development

Linear Regression Model

Linear regression is a linear mathematical method of establishing the relationship between variables through finding the line that best fits the data point. The equation of the line of best fit is then used to make future predictions for the variables under study [19].

Feed Forward Neural Network Model

Artificial neural networks, are computational networks whose structure and functionality are inspired by the biological neurons in the human brain. They are popular and robust machine learning algorithms widely used in prediction and classification problems. A feed forward multi-layer perceptron artificial neural network was used in this study. It is referred to as a feed forward network because the data only propagates forward in the network, the connections between the neurons do not form a loop as in the case of recurrent neural networks. It is a Multi-Layer Perceptron network because it has an input, output and at least one hidden layer of neurons. The neurons connect to each other and possess associated weights and biases.

Figure 2 shows a feed forward Multi-Layer Perceptron (MLP) neural network with one hidden layer of four neurons. Information enters the network via the input layer whose number of neurons correspond to the number of feature variables. The training process involves transformation of the input data in the hidden layers. The transformation entails multiplying each of the inputs to the hidden layer with a suitable weight, summation of the weighted inputs and application of biases as required. Activation functions are applied to the modified inputs before being passed on to the next layer, which in Figure 2 is the output layer [20]. The output layer has the target variable and contains one neuron for a regression problem and two neurons for a binary classification problem.

Long Short Term Memory Neural Network Models

LSTMs are an advanced type of recurrent neural networks (RNNs). Recurrent Neural Networks are a class of artificial neural networks with loops that facilitate the movement of information from one step to another. As they have memory, they use the output from the previous step as an input to the next step [18]. They have the ability to capture short term time dependencies in data and this makes them well suited for applications involving ordered time series data. However, they are also faced with the challenge of vanishing gradient and are therefore unable to effectively capture inconstant long term dependencies in the data [21].

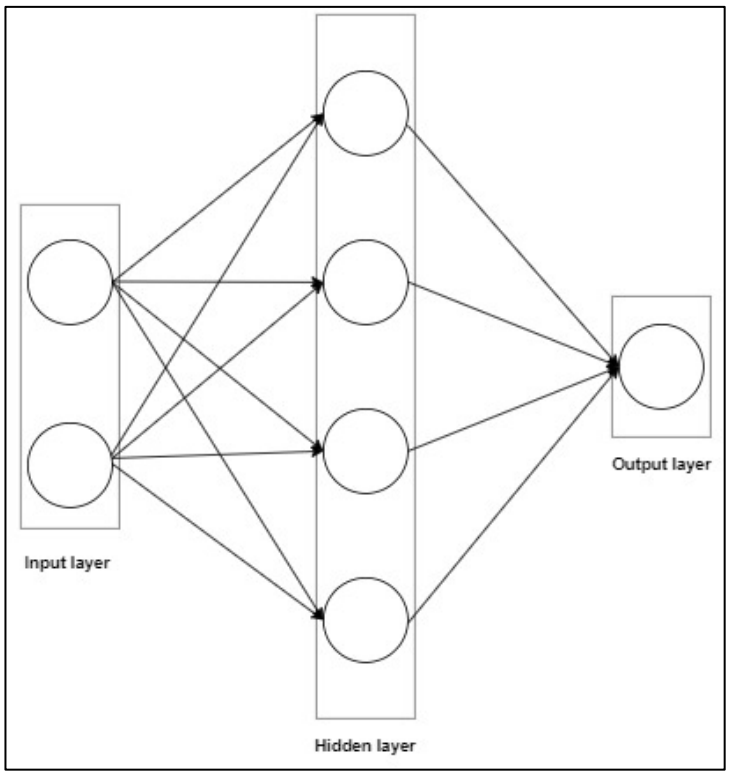


Figure 2. Feed forward multi-layer perceptron neural network.

LSTM neural networks overcome the long term dependency issue typical of general recurrent neural networks. Each LSTM unit contains a memory cell or cell state, an input gate, output gate and forget gate. The gates have weights and bias parameters that they use to ensure controlled flow of information within the cell state, thereby facilitating the memorization of multiple time dependencies of various characteristics by the LSTM network [21]. The previous output, current input as well as status of the cell state are used in the LSTM unit's computation. The outcome of the computation serves not only as the output from one unit to the next but is also used to update the cell state. The input gate controls the updating of the cell state with input values while the forget gate determines what data to retain or discard from the cell state. The output gate determines what data is passed as the output from the cell state [18] [22].

Figure 3 is an illustration of the structure of an LSTM unit. Each of the gates has a sigmoid activation function which fits input data into the range of 0 to 1. Values closer to 1 are considered to be more important while values closer to 0 are considered to be less relevant. The LSTM unit also has two tanh activation functions that fit their input data into the range of -1 to 1 as a means of regularization of the network. x_t and h_{t-1} are the current time input and the previous hidden state respectively. c_{t-1} is the previous cell state while c_t is the new modified cell state. h_t is the new hidden state. f_t, i_t and o_t are outputs from the forget, input and output gates respectively while t_t is the output of the tanh activation function associated with the input gate. Information flow and transformation within the LSTM unit is described by the equations (1) to (6) [23] [24].

URL: http://dx.doi.org/10.14738/tmlai.106.13645

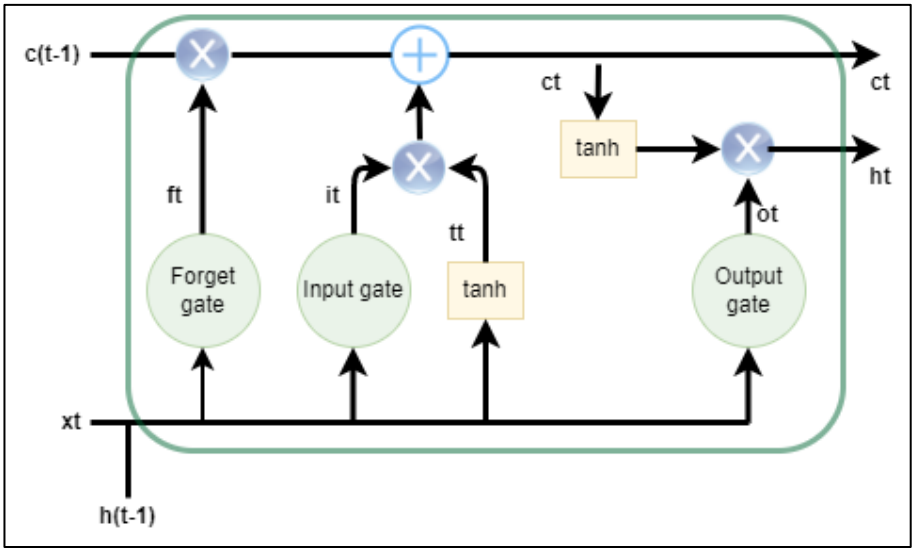


Figure 3. Structure of an LSTM unit

$$f_t = \sigma\left(w_{xf}x_t + w_{hf}h_{t-1} + b_f\right) \quad (1)$$
$$i_t = \sigma(w_{xi}x_t + w_{hi}h_{t-1} + b_i) \quad (2)$$
$$t_t = tanh(w_{xt}x_t + w_{ht}h_{t-1} + b_t) \quad (3)$$
$$c_t = (f_t * c_{t-1}) + (i_t * t_t) \quad (4)$$
$$o_t = \sigma(w_{xo}x_t + w_{ho}h_{t-1} + b_o) \quad (5)$$
$$h_t = c_t(\tanh) * o_t \quad (6)$$

Bidirectional LSTMs are a type of LSTM network architecture. They have both forward and backward recurrent components and thus transmit the cell state vector in both forward and backward directions. The forward component extracts time dependencies from the data based on previous information while the backward component does the extraction based on future information. This additional capability enhances the performance of the LSTM model [18] [21].

Model Evaluation Criteria

The performance of regression models is assessed by comparing the measured values and predicted values. The mean absolute error (MAE), root mean squared error (RMSE), mean squared error (MSE) and coefficient of determination (R^2) parameters are used as the evaluation metrics that define the precision of the models' estimation of the solar PV power produced. They are defined in Equations (7) to (10) where n is the number of observations, y_m are the actual measured values for the power produced, $\widehat{y_m}$ the average of the measured values and y_p are the model predicted power values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_m - y_p\right| \quad (7)$$
$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_m - y_p\right)^2 \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_m - y_p)^2} \quad (9)$$

$$R^2 = \frac{\sum_{i=1}^{n}(y_m - y_p)^2}{\sum_{i=1}^{n}(y_m - \widehat{y_m})^2} \quad (10)$$

Lower values of the mean absolute error and root mean square error indicate higher accuracy model accuracy, Similarly, coefficient of determination values closer to one indicate more accurate predictions [22].

Fault Prediction

The test data set comprised both fault and normal system conditions. The residual error, which is the difference between the actual and predicted power was assessed against statistical control chart thresholds computed as a factor of deviation from the normal distribution so as to predict faults in the system.

RESULTS

Linear regression, feed forward MLP, stacked and bidirectional LSTM solar power prediction models were developed. Irradiance and ambient temperature were selected as the feature variables while the active power produced was the target variable. The LSTM model training data was input in sequences of two days.

Linear Regression Model

The linear regression model achieved a MAE of 4.3% and an R^2 of 90.2%. Although a high positive correlation is observed which is indicative of good performance, the non-linearity in the relationship between the feature and target variables cannot be ignored.

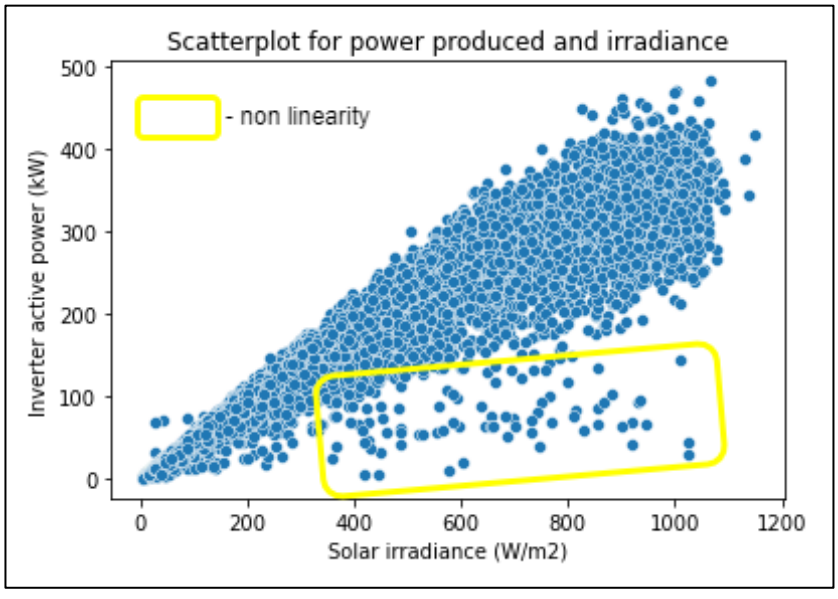


Figure 4. Non linearity between the power produced and irradiance.

URL: http://dx.doi.org/10.14738/tmlai.106.13645

Figure 4 is a scatter plot for the active power produced and the solar irradiance highlighting an element of nonlinearity which is attributed to environmental factors such as dirt, dust, as well as cloud and vegetation cover. Therefore, the linear regression model is not sufficient for the goal of developing a power prediction model that is robust with good generalization. As machine learning accommodates this non linearity in the data during model training, results for the neural network models follow.

Machine Learning Models

The model hyper parameters were selected by trial and error to maximize performance. **Error! Reference source not found.** and **Error! Reference source not found.** summarize the hyper parameters for the feed forward MLP and LSTM neural network models respectively.

Table 1. Hyper parameters for the feed forward neural network model.

Characteristic	Feed Forward MLP
Hidden layer	1
Neurons	5
Epochs	30
Optimizer	Stochastic gradient descent

Table 2. Hyper parameters for stacked and bidirectional LSTM models.

Characteristic	Value
LSTM layers	2 with 64 units each
Dense layer	1 with 1 unit
Dropout rate	0.2
Optimizer	Adam
Activation function	Tanh
Time steps	2 days
Batch size	32
Epochs	70
Bidirectional layers	2 (for bidirectional LSTM model)

Given its backward recurrent component, the bidirectional LSTM model was expected to outperform the stacked LSTM model but the results depicted otherwise. The stacked LSTM attained a MAE of 6.9% while the bidirectional LSTM attained 7.2% for the MAE. Table 3 summarizes the evaluation metrics for all the models.

Table 3. Model evaluation metrics.

Metric	Linear Regression	Feed Forward	Stacked LSTM	Bidirectional LSTM
MSE	1326.6	1369.1	3096.6	3273.5
MAE	4.3%	4.2%	6.9%	7.2%
RMSE	6.1%	6.2%	9.3%	9.5%
R^2	90.2%	89.9%	76.9%	75.6%

The table shows that besides the linear regression model, the feedforward neural network had the highest accuracy while the bidirectional LSTM model had the least accuracy. Further analysis of the feedforward and LSTM neural network model outputs was done before final model selection by studying the plots for their predicted power against the measured power. The results are presented in Figure 5. Figure 6 provides a closer look at the predicted power from both models for the dates 24th to 26th March, 2022. Generally, some deviations were observed in the model outputs between the predicted and real power produced; these are the residual errors that were used to predict faults in the system. The feed forward model consistently predicted higher values during peak power production times and closely followed the pattern for the real power data outside of the peak times. The stacked LSTM model on the other hand made power estimations closest to the measured power during the peak production hours and consistently predicts higher power values during non-peak production hours. As most power production occurs during peak hours, this is a discrepancy that can be managed. Given more training data, it is believed that the stacked LSTM model would have an even better performance. Despite the deviations, it was observed that both models managed to predict the output power satisfactorily.

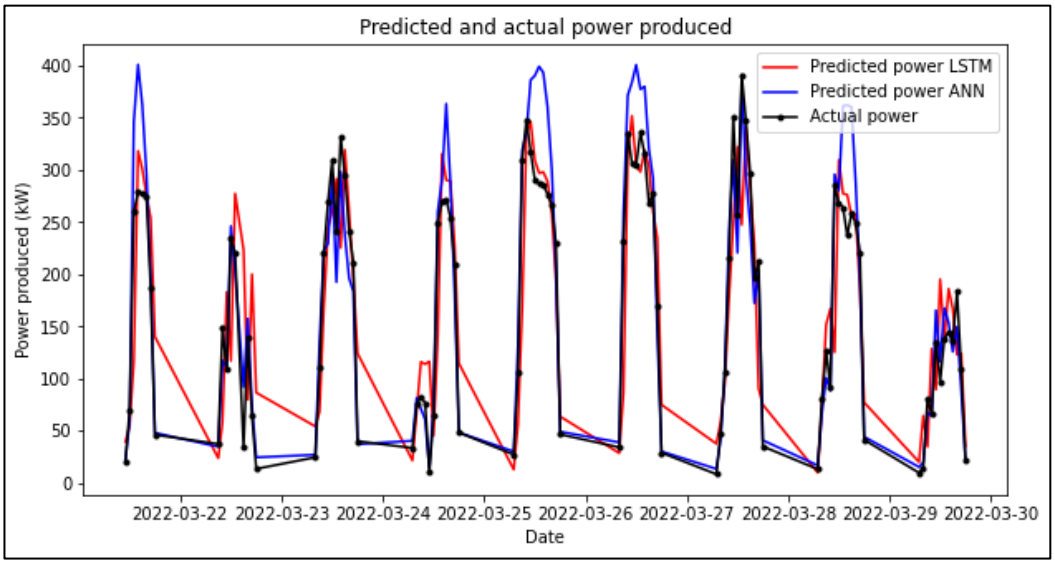


Figure 5. Model output prediction comparisons.

URL: http://dx.doi.org/10.14738/tmlai.106.13645

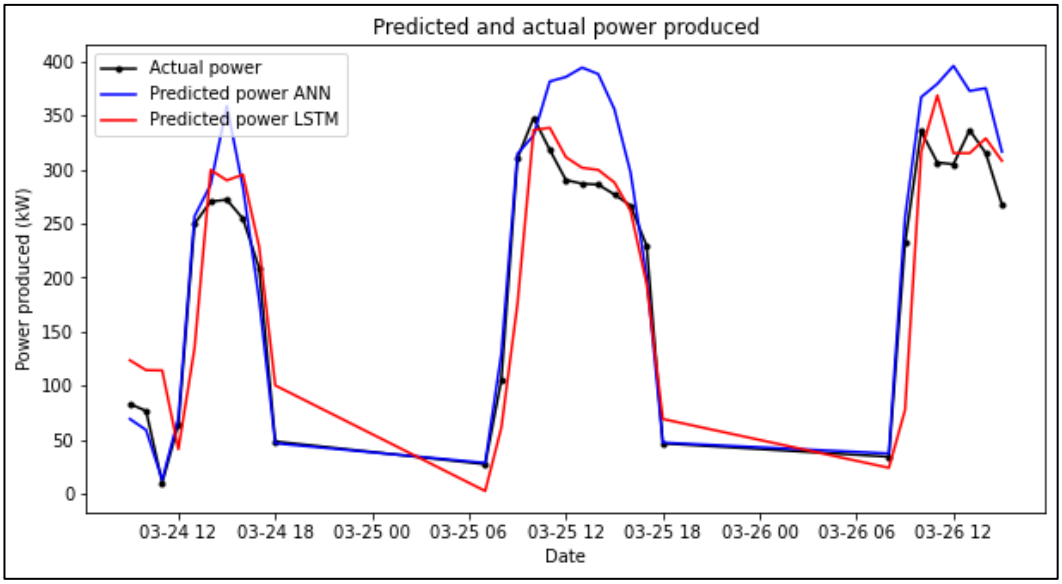


Figure 6. Power predictions for the period 24th to 26th March, 2022.

FAULT PREDICTION SYSTEM

Basing on the results, a hybrid fault prediction system that combines the feed forward and stacked LSTM model is selected in order to harness the strengths of both architectures.

A system is deemed to be in statistical control when it is not changing its range and mean attributes. Mean control charts are used to monitor the mean of a process variable as a form of statistical process control. They are used to determine whether the variation is part of the normal operating conditions or whether it signifies an abnormal situation. For data that follows a normal distribution, the warning and action limits on a mean chart are defined as below [25]:

- Action limits: Mean ± 3*Standard deviation
- Warning limits: Mean ± 2*Standard deviation

Thresholds for the proposed system are derived from these limits and values outside the range of the action limits are interpreted as representative of fault conditions.

Threshold Determination

The residual errors are used to predict faults. First, density and histogram plots for the residual error are obtained to confirm that the data follows a normal distribution. They are presented in Figure 7. It is observed that the data indeed follows a normal distribution hence qualifies for the application of the mean control chart. The warning level and threshold limits for the fault prediction tool are calculated and presented in Table 4.

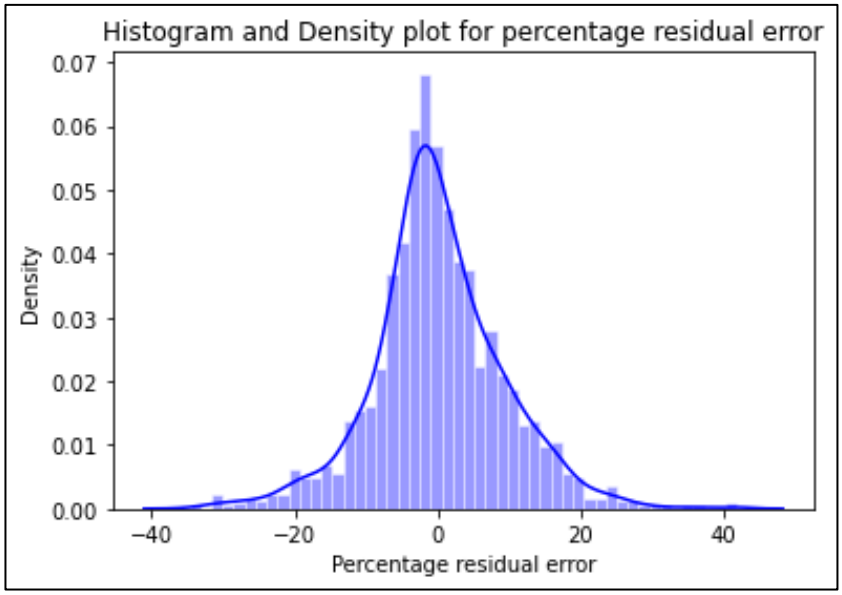


Figure 7. Density plot for residual error percentage

Table 4. Limits for the hybrid fault prediction tool

Limit	Stacked LSTM module	Feed forward module
Upper threshold	25.47%	19.91%
Upper warning limit	19.25%	14.83%
Lower threshold	-11.84%	-10.57%
Lower warning limit	-5.63%	-5.49%

Fault Prediction

The test dataset is inspected and fault conditions in the system identified basing on personal judgement. The proposed tool is then tested on both normal and abnormal system conditions. The results are presented in the figures that follow.

LSTM Model Fault Prediction Tool Component

Figure 8 shows the mean chart for the residual error of the LSTM fault prediction module for the duration of 13th to 19th April, 2022. It shows that the tool is able to predict the faults experienced on 14th, 17th and 18th April, 2022. On these days, the percentage residual error is observed to exceed both the warning limit and threshold. On 18th April for example, the percentage residual error is observed to be 32%, exceeding the threshold by 6.53% and the tool successfully predicts this anomaly.

The performance of the LSTM module is also tested under normal system conditions and the results presented in Figure 9. The figure shows data for the duration of 10th to 16th May, 2022 during which the fault prediction tool does not predict any anomalies on these days as the residual error does not exceed the threshold on any of them. The ability of a fault prediction

URL: http://dx.doi.org/10.14738/tmlai.106.13645

system to distinguish between normal and abnormal operating conditions is at the core of its functionality. This feature is successfully demonstrated by the LSTM component of the proposed tool in this scenario. The residual error on some days like the 11th is observed to be close to the warning limit. If such a trend were observed over a week, this would be interpreted as an impending fault in the system.

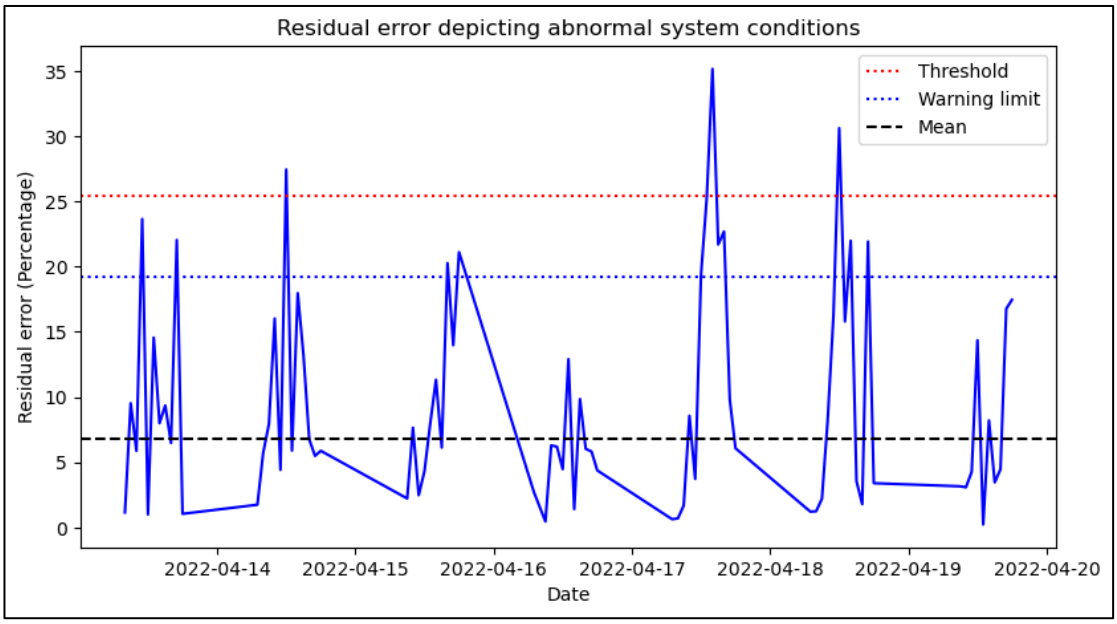


Figure 8. LSTM module functionality under fault conditions.

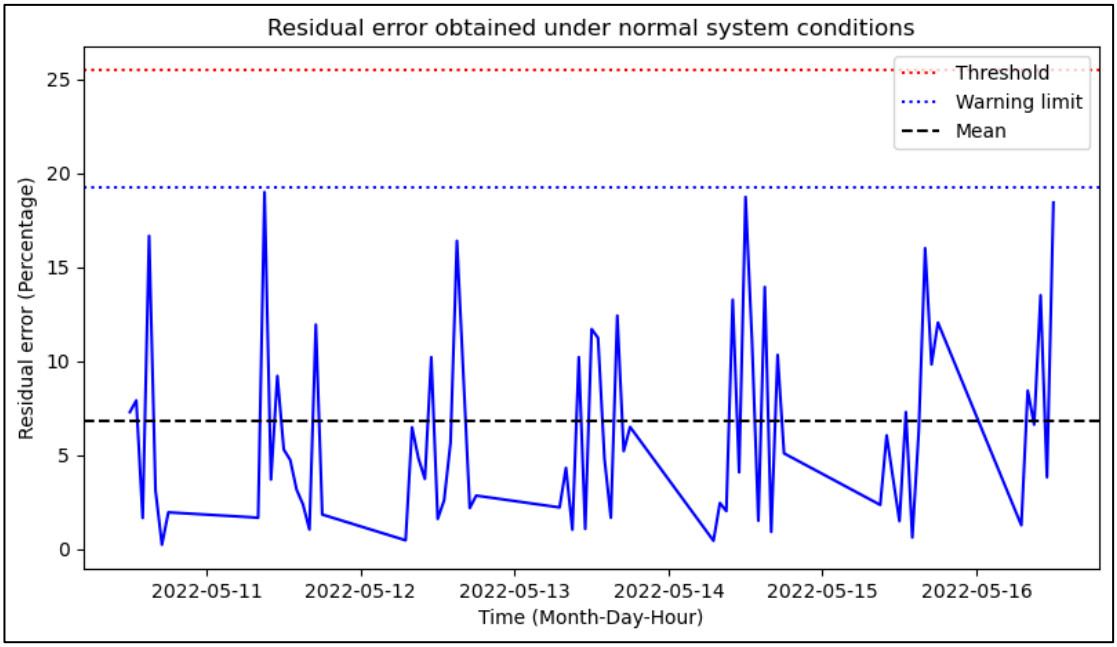


Figure 9. LSTM module functionality under normal conditions.

Feed Forward Neural Network Model Component

The performance of the feed forward neural network model module is also verified. Figure 10 shows the mean chart for the residual error on randomly selected days for this component of

the fault prediction tool. The tool predicts the faults experienced on 13th, 17th and 18th April, 2022 with the residual error percentage being recorded as above the threshold. This verifies for the anomaly detection capability of the feed forward model component of the fault prediction tool.

Figure 11 shows the functionality of the feed forward module under normal system conditions from 11th to 17th May, 2022. As expected, no faults are detected during this period as the residual error does not exceed the threshold at any point in time. This confirms the capability of the feed forward model component of the fault prediction tool to successfully discriminate between fault and non-fault conditions.

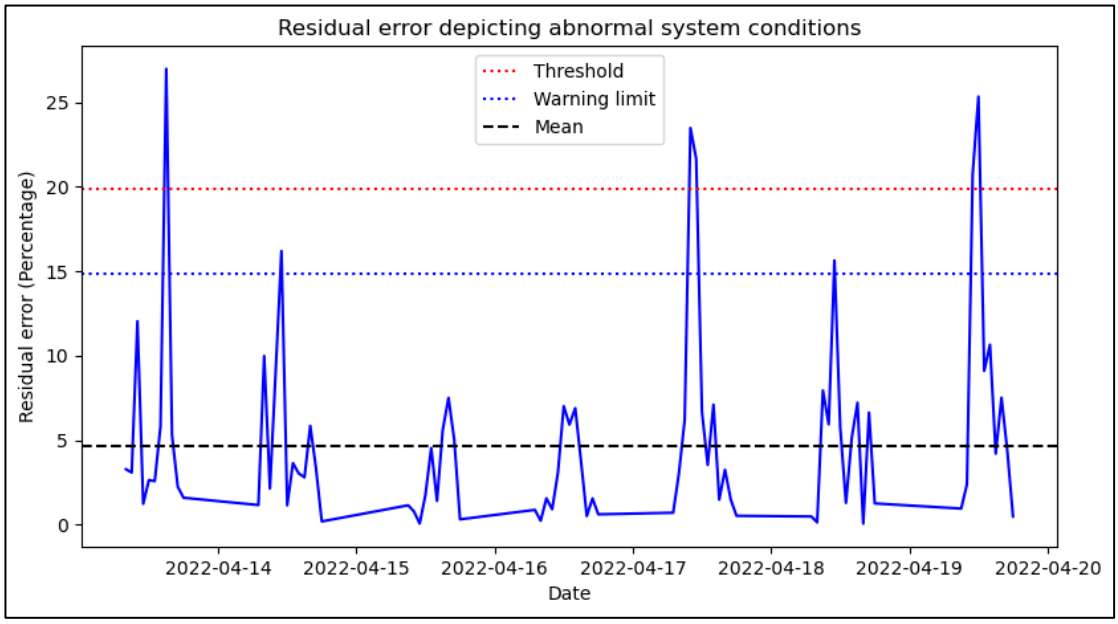


Figure 10. Feedforward module functionality under fault conditions.

Hybrid Fault Prediction System

Having demonstrated the ability of both the feed forward and LSTM modules to predicts faults in the photovoltaic system, the proposed tool combines the functionality of both modules to capitalize on each of their strengths. The fault prediction tool is a hybrid system with a 2oo2 voting system architecture. Under the 2oo2 voting arrangement, fault prediction is dependent on both modules detecting the same anomaly. This lowers the probability of false alarm generation by the proposed system [26].

URL: http://dx.doi.org/10.14738/tmlai.106.13645

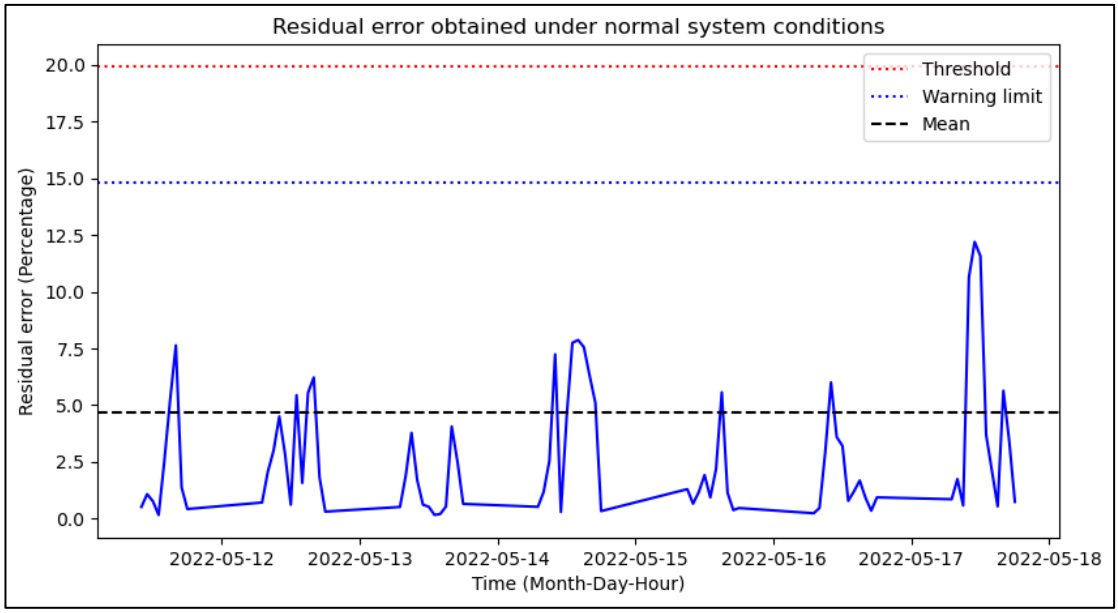


Figure 11. Feed forward module functionality under normal conditions.

CONCLUSION

In this study, a neural network based fault prediction tool for a solar plant in Uganda was developed. Two years' SCADA historical data for a 1.6MW solar hybrid plant in Uganda was used to develop a linear regression model as well as feedforward MLP, stacked LSTM neural network and bidirectional LSTM neural network models. The linear regression model depicted good performance with a MAE of 4.3%. However, it was not considered suitable for solar system fault prediction as it did not account for non-linearity in the plant data that is due to environmental factors like dirt. The feedforward neural network model also demonstrated a good performance with a MAE of 4.2%. In order to harness the ability of LSTM neural networks to learn data dependencies in the order of the time series data, stacked and bidirectional LSTM models were built and their performances compared. Although the bidirectional LSTM model was expected to offer better results given its additional capability of learning data dependencies in the backward direction, results showed otherwise with the stacked LSTM model doing better and attaining a MAE of 6.9%.

The proposed solar fault prediction tool is a hybrid system of both the stacked LTSM and feedforward model components combined with a mean chart statistical process control tool. The system utilizes a 2oo2 voting arrangement to decide on the presence of anomalies in the PV system. The warning and threshold limits of the fault prediction tool were set at 19.25% and 25.47% of the residual error respectively for the LSTM component and at 14.83% and 19.91% of the residual error respectively for the feedforward model component. The test data set was inspected for abnormal conditions and these were assumed based on personal judgement. The tool was applied to randomly selected normal and abnormal conditions in the solar plant system and its performance studied in both scenarios. Results demonstrated that both modules of the fault prediction tool successfully predicted faults conditions in the solar PV system.

In utilization of the proposed tool, we recommend that the user only considers data captured during the peak production hours as input to the power prediction model so as to overcome the observed model behaviour of consistently predicting higher power values than the actual

during non-peak hours. In addition, the user should inspect the plant data and use their personal judgement to distinguish between abnormal and normal system conditions.

With more training and failure data available, the LSTM model performance will be enhanced further and the predicted faults classified during future works. Long term fault predictions will also be achieved by determining the moving averages for the residuals. With enhanced accuracy, the model could also be applied in transfer learning.

The major limitation faced in conducting this research project was the lack of access to detailed failure data in form of alarm logs, incidence reports and maintenance logs to facilitate classification of predicted faults.

To conclude, the research findings met the research objectives with the solar fault prediction tool depicting satisfactory performance. It is anticipated that if implemented at the solar power plant after further improvement, it will be useful in the prediction of system faults to inform predictive maintenance scheduling so as to maximize plant production efficiency.

References

[1] S. S. R. a. W. G. M. K. AbdulMawjood, "Detection and prediction of faults in photovoltaic arrays: A review," *IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering,* pp. 1-8, 2018.

[2] C. Kavuma, D. Sandoval and H. K. Jean de Dieu, "Analysis of solar photo- voltaic for grid integration viability in Uganda," *Energy Science and Engineering,* vol. X, no. 3, pp. 694-706, 2022.

[3] Syafaruddin and D. S. Zinger, "Review on Methods of Fault Diagnosis in Photovoltaic System Applications," *Journal of Engineering Science and Technology Review,* vol. 12, no. 5, pp. 53-66, 2019.

[4] T. Berghout, M. Benbouzid, T. Bentrcia, X. Ma, S. Djurovic and L.-H. Mouss, "Machine Learning-Based Condition Monitoring for PV Systems: State of the Art and Future Prospects," *Energies,* vol. 14, no. 6316, pp. 1-24, 2021.

[5] B. Long-Dong, Y.-K. Wu and M.-H. Pham, "Fault identification and diagnosis methods for photovoltaic system: A review," in *2021 7th International Conference on Applied System Innovation*, Chiayi, 2021.

[6] M. S. Arani and M. A. Hejazi, "The Comprehensive Study of Electrical Faults in PV Arrays," *Journal of Electrical and Computer Engineering,* pp. 1-10, 2016.

[7] Y. Hammoudi, I. Idrissi, M. Boukabous, Y. Zerguit and H. B. , "Review on maintenance of photovoltaic systems based on deep learning and internet of things," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 26, no. 2, pp. 1060-1072, 2022.

[8] M. Dhimish, V. Holmes, P. Mather and M. Sibley, "Novel hot spot mitigation technique to enhance photovoltaic solar panels output power performance," *Solar Energy Materials and Solar Cells,* vol. 179, pp. 72-79, 2018.

[9] L. Cristaldi, M. Khalil and P. Soulatiantork, "A root cause analysis and a risk evaluation of PV balance of systems failures," *Acta IMEKO,* vol. VI, no. 4, pp. 113-120, 2017.

[10] G. M. Tina, C. Ventura, S. Ferlito and S. D. Vito, "A State-of-Art-Review on Machine-Learning Based Methods for PV," *Applied Sciences,* vol. 11, no. 16, pp. 1-34, 2021.

[11] L. B. Bosman, W. D. Leon-Salas, W. Hutzel and E. A. Soto, "PV System Predictive Maintenance: Challenges, Current Approaches, and Opportunities," *Energies,* vol. 13, no. 1398, pp. 1-16, 2020.

URL: http://dx.doi.org/10.14738/tmlai.106.13645

[12] J. A. Dhanraj, M. Ali, V. Karthikeyan, T. Kuaanan and C. Prem Kumar, "An effective evaluation on fault detection in solar panels," *Energies,* vol. XIV, no. 22, pp. 1-14, 2021.

[13] M. De Benedettia, F. Leonardi, M. Fabrizio and S. Corrado, "Anomaly detection and predictive maintenance for photovoltaic systems," *Neurocomputing,* vol. 310, p. 59–68, 2018.

[14] A. Betti, M. L. L. Trovato, F. S. Leonardi, G. Leotta, F. Ruffini and C. Lanzetta, "Predictive Maintenance in Photovoltaic Plants with a Big Data Approach," *33rd European Photovoltaic Solar Energy Conference and Exhibition,* pp. 1895-1900, 2019.

[15] A. Betti, M. Tucci, E. Crisostomi, S. Barmada, A. Piazzi and D. Thomopulos, "Fault Prediction and Early-Detection in Large PV Power Plants Based on Self-Organizing Maps," *Sensors,* vol. XXI, no. 5, pp. 1-16, 2021.

[16] F. Aziz, S. Ahmad, A. Ul Haq, Y. Mahmoud and M. Jalal, "A Novel Convolutional Neural Network Based Approach for Fault Classification in Photovoltaic Arrays," *IEEE Access,* vol. 8, pp. 41889-41904, 2021.

[17] R. F. Colmenares-Quintero, E. R. Rojas-Martinez, F. Macho-Hernantes, K. E. Stansfield and J. C. Colmenares-Quintero, "Methodology for automatic fault detection in photovoltaic arrays from artificial neural networks," *Cogent Engineering,* vol. 8, no. 1, pp. 1-21, 2021.

[18] S. Khan, F. Shaikh, M. M. Siddiqui, T. Hussain, L. Kumar and A. Nahar, "Hourly Forecasting of Solar Photovoltaic Power in Pakistan Using Recurrent Neural Networks," *International Journal of Photoenergy,* vol. 2022, pp. 1-11, 2022.

[19] A. Patel, O. V. G. Swathika, U. Subramaniam, T. S. Babu, A. Tripathi, S. Nag, A. Karthick and M. Muhibbullah, "A Practical Approach for Predicting Power in a Small-Scale OffGrid Photovoltaic System using Machine Learning Algorithms," *International Journal of Photoenergy,* vol. 2022, pp. 1-21, 2022.

[20] D. Amaratunga, "Optimization of Coagulant (ALUM) Dosage using an Advanced Process Control Strategy based on Artificial Neural Network (ANN)," Engineering Institute of Technology, 2018.

[21] B. Lindemann, T. Müller, H. Vietz, N. Jazdi and M. Weyrich, "A survey on long short-term memory networks for time series prediction," *Procedia CIRP,* vol. 99, pp. 650-655, 2021.

[22] G. Li, S. Xie, B. Wang, J. Xin, Y. Li and S. Du, "LSTM neural networks Photovoltaic Power Forecasting With a Hybrid Deep Learning Approach," *IEEE Access,* vol. 8, pp. 175871-175880, 2020.

[23] N. E. Michael, M. Mishra, S. Hasan and A. Al-Durra, "Short-Term Solar Power Predicting Model Based on Multi-Step CNN Stacked LSTM Technique," *Energies,* vol. 15, no. 2150, pp. 1-20, 2022.

[24] F. Harrou, F. Kadri and Y. Sun, "Forecasting of Photovoltaic Solar Power Production Using LSTM Approach," in *Advanced Statistical Modeling, Forecasting, and Fault Detection in Renewable Energy Systems*, IntechOpen, 2020.

[25] J. S. Oakland, "Process Control," in *Statistical Process Control*, Oxford, Elsevier, 2008, pp. 106-113.

[26] K. W. See, G. Wang, Y. Zhang, Y. Wang and L. Meng, "Critical review and functional safety of a battery management system," *International Journal of Coal Science & Technology,* vol. 9, no. 36, pp. 1-17, 2022.