Int J Syst Assur Eng Manag (October 2022) 13(5):2371–2390
https://doi.org/10.1007/s13198-022-01649-7

ORIGINAL ARTICLE

Mining sensor data in a smart environment: a study of control algorithms and microgrid testbed for temporal forecasting and patterns of failure

Akram Qashou[1] · Sufian Yousef[1] · Erika Sanchez-Velazquez[1]

Received: 22 December 2021 / Revised: 22 December 2021 / Accepted: 16 February 2022 / Published online: 24 March 2022

Abstract The generation of active power in renewable energy is dependent on several factors. These variables are related to the areas of weather, physical structure, control, and load behavior. Estimating the future value of the active power to be generated is difficult due to their unpredictable character. However, because of the higher precision required of the estimation, this problem becomes more complex if we examine a short-term temporal prediction. This study presents a method for converting stochastic behavior into a stable pattern, which can subsequently be used in a short-term estimator. For this conversion, K-means clustering is employed, followed by Long-Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) algorithms to perform the Short-term estimate. The environment, the operation, and the generated (normal or faulty) signal are all simulated using mathematical models. Weather parameters and load samples have been collected as part of a dataset. Monte-Carlo simulation using MATLAB programming has been realized to conduct an experiment. In addition, the LSTM and the GRU are compared to see how well they perform in this system. The proposed method's end findings outperform the current state-of-the-art.

Keywords Renewable energy · Smart home · Short-term prediction · Stochastic behavior · Deep learning

✉ Akram Qashou
ammq100@pgr.aru.ac.uk

Sufian Yousef
sufian.yousef@aru.ac.uk

Erika Sanchez-Velazquez
erika.sanchez@anglia.ac.uk

[1] Anglia Ruskin University, Cambridge, UK

1 Introduction

A smart grid is an electricity network enabling a two-way flow of electricity and data with digital communications technology. This gives the ability of monitoring, managing, and automatic decision-making. Besides, smart grid uses a wide range of resources based on information technology techniques to enable new and existing guidelines in minimizing energy costs and reducing electricity wastes. According to Ali et al. (2013), the smart grid is one of the most complicated and largest systems considering the design and building processes, although it is one of the easiest to use. It uses all kinds of power plants (including hydro, solar, coal, nuclear, wind turbine, and natural gas, among others), substations, transformers, and high-voltage transmission lines (Hasan et al. 2019), therefore, there is the need for a demand-responsive electrical grid with high efficiency of energy use. The traditional grid uses a one-way limited interaction, in which power flows to the consumers from the power plant. In contrast, the smart grid introduces a two-way interchange in which involve the exchange both information and electricity, in both directions (between consumer and power utilities). The growing network of computers, automation, control, and communications are instrumental in making the grid "greener", more reliable, more secure and more efficient (Hasan et al. 2019).

These grids could provide a rich dataset that could be used for analyzing and monitoring their activities, but these could also be used to provide an opportunity to use this data in different applications as well, such as using them for sensor data (Jakkula and Cook 2007).

This data could be useful when being set to work with different aspects or dimensions of SmartGrid such as integration with renewable energy sources, management of

intermittent power supplies, real-time data responses as well as the energy pricing strategies among others (Jakkula and Cook 2007). As such, it becomes a necessity that we would develop the right tools and methods which could help in conserving the energy by gathering the data from the smart grid using sensors which could then be used to recognize patterns from previous data and forecast or predict in order to conserve energy in the smart grids.

Some of the algorithm that could be used for prediction which are related to deep learning algorithms like Long-short term memory (LSTM), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU). In this work, the used predictor is the most efficient one of them, in terms of accuracy and delay.

LSTM is an RNN variant that is meant specifically for time series data. The LSTM is used in addressing this problem in addition to empowering RNNs algorithms using internal memory cells (Li et al. 2020).

RNNs are a form of neural networks that adopt the feedback connections in various nodes in remembering the previous time steps values. As such, they can capture the time series data's temporal behaviour. (Tealab 2018).

GRU is a kind of gated RNN which is largely used in mitigating the gradient vanishing problem of RNNs using gating mechanism in addition to making the structure simpler without interfering with the effect of LSTM neural network (Luo et al. 2021).

GRU addresses the vanishing gradient problem, which are the values used in updating network weights. According to (Agarap 2018), GRU can solve this problem using two gates, the rest gate and the update gate. The gates are instrumental in deciding the information that is allowed to pass to the output in addition to having the ability to be trained for retaining information from farther back, thus allowing it to pass relevant information through a variety of events for making better predictions.

However, since these prediction methods are based on regression techniques, which tries to find a common pattern for the historical samples to use it to predict the future values. Considering our application, the historical samples from the energy generators and also the load of the smart city, may not have a constant pattern. This is due to the stochasticity behavior of the environment. Therefore, in order to convert this dynamicity to a static pattern, in this work, K-Means clustering algorithm is used.

K-means clustering algorithm refers to a simple unsupervised learning algorithm used in solving clustering problems which is useful in clustering analysis. According to (Xu et al. 2020), the algorithm is applied using certain procedures that classify a certain set of data into clusters defined by the letter "K."

Although several works have tackled the problem of the prediction in the SmartGrid context, most of these works focus only on the long-term prediction. The advantage of the long-term prediction is in bringing long-term strategy and planning, However, the methods that are currently used for this task provides in accurate predictions. On the other hands, short and very shorts terms predictions are important for the real-time control of the SmartGrid which requires real-time information from the real load and the natural resources in addition to the physical structure of the solar panels. Moreover, providing an accurate short-term prediction for the generated energy is also a challenging problem. This is due to the fact, that the amount of generated energy depends on several parameters, from the physical components of the unit to the weather conditions. All these parameters are totally stochastic and do not follow any accurate pattern.

Our methodology in this work is to convert the stochastic behavior of the attributes into an accurate pattern using a clustering algorithm (i.e., K-Means). This allows us to be able to identify their fitting curve and use a suitable regression-based algorithm (i.e., LSTM and GRU) for an accurate short-term prediction.

The main objective of this work is to propose a method that allows enhancing the accuracy of the short-term generated power prediction for the SmartGrid environment.

The main contributions to the existing body of knowledge that this study will make include:

1. Identify the most useful factors that affects the accuracy of the Smart Grid short term prediction process.
2. Implement a model (or a combination of already existing models) for recognizing patterns of failure in the Smart Grid.
3. Identification of the best deep learning algorithm to mine data from a synthetic testbed.
4. Proposing a solution to provide a comparatively enhanced prediction results using unpattern-able (highly dynamic and stochastic) data.
5. Providing a solution that enhances the protection level for smart grid dynamic environment against failures.
6. Providing an accurate future value for the short-term prediction in a relatively faster processing time.

2 Literature review

Efficient delivery of energy resources to the smart grid requires a balanced energy demand and supply by developing energy resource management strategies. However, the significant fluctuations in energy demand and supply enhance the challenges in the development of these energy resource management schemes. This problem has been tackled using different approaches.

For instance, this work (Yu et al. 2015) developed several approaches to predict energy supply and demand effectively. The study then develops machine learning-based methods for accurate energy consumption and generation forecasts. Lastly, the study used the prediction results to establish energy consumption upper and lower bounds realizing optimal demand and anomaly detections.

The problem is the smart meters acquire large amounts of data through sophisticated signal processing algorithms. The methodology will be applied in the study for it will the first develops a new classification scheme that categorizes users based on their consumption patterns. The study will then test the proposed and benchmarked models. Additionally, this research uses semi-Markov models to generate more extensive and more realistic test data due to insufficient power consumption data. (Tornai et al. 2016).

The problem is electrical load prediction is a fundamental factor in the planning, operations, and resource management within the grid system. The numerous restructurings of the grid and the integration of new devices to the grid heighten the need for forecasting to better plan for energy supply and demand. The methodology will be applied in the study for it will the first proposes a prediction model capable of predicting load data. Additionally, the study assesses the prediction model's performance and effectiveness against several metrics. (Chemetova et al. 2017).

The problem is difficulty in harvesting various renewable energy forms led to the use of the smart grids integrated with photovoltaic (PV) power. However, various atmospheric conditions, for instance, rain, affect solar irradiance occurrence. This solution proposes the adoption of wavelet transform and Elman Neural Network (WT-ENN) for short-term solar energy production and irradiance forecasting. The methodology, the first decomposed the original solar irradiance data into consecutive stable wavelet sub-series. Additionally, the study used ENN to optimize the new wavelet coefficients. Lastly, the study reconstructed solar irradiance using the prediction model and the new coefficients. The prediction model's performance was then assessed using two real-world data solar irradiance datasets. (Huang et al. 2019).

The problem is Smart grid systems allow consumers to use more energy from the grid or vend it back to the grid for other consumers. Smart homes with photovoltaic systems can establish the daily energy yield. This solution recommends the use of multi-layer perceptron based on photovoltaic forecasting on rooftop PV systems. The methodology will be applied, the first suggests the use of multi-layer perceptron-based PV forecasting. The study then trains its historical data, conducts cross-validation, and tests the model using real-world PV data. (Parvez et al. 2020).

The problem is emergence of smart grids offers better integration of power systems between energy producers and consumers. The bidirectional nature of these smart grids calls energy consumption optimization measures to maintain the grid's reliability and supply–demand balance. The solution evaluates the available short-term energy consumption prediction models to determine next-day energy consumption forecasts at one-hour intervals realizing a 24-point forecast. This methodology conducted a thorough assessment of various high-level machine algorithms adopted to forecast and evaluate the various model instances to determine the most appropriate algorithm for energy consumption forecasting. (Petrican et al. 2018).

The problem is Efficient energy delivery in the smart grid requires adopting energy resource management strategies that balance energy supply and demand. This solution proposes several techniques that accurately model and predict energy production and demand over time. Similarly, the study recommends modeling analyses that statistical output models of energy consumption and machine learning approaches improve prediction accuracy. The methodology, the first the statistical distribution of real-world meter reading data of several houses for over 200 days acquired from Stanford University. (Yu et al. 2014).

The problem is several countries continue to record an increase in their solar power capacity connections to the distribution grids. Adopting the smart grid concept has since contributed to this increase. The solution study suggests a new forecasting model that uses autoregressive models and gradient boosting algorithms. Such promote the alleviation of these constraints and result in point and probabilistic solar power forecasts for medium voltage and low voltage distribution stations. The methodology, the first proposes a model that overcomes the information and communication technology (ICT) limitations to promote solar energy forecasts at secondary substation levels. It then combined the values obtained from various distributed sensors. (Bessa 2014).

The problem is various new technology appliances that consumers currently use in their households overwhelm the existing smart grid infrastructure as they were initially not developed to support these devices. This solution proposes the implementation of various methods energy providers can improve their energy consumption forecasts for households despite their variability in electrical appliance usage. The methodology, the first assesses the existing prediction models ad their significance. It then describes various modeling techniques that assess the existing statistical approaches and machine learning algorithms. (Lauer et al. 2019).

3 Analytical analysis

To realizes our proposed solution, we need first to identify the environment model, which in this case includes a photovoltaic grid. Then the process of generating the energy depending on the physical structure and the weather conditions in addition to the generated signal need to be modelled and then simulated. After this, the generated data are clustered using the K-means clustering algorithm. Finally, LSTM and GRU are used to provide the short-term predictions.

3.1 Environment and PV model

The earth rotates around the sun approximately 8766 h about 365.242 days. Earth is closest to the Sun (147 million km) on January 2, and this point is called perihelion.

Specific point on earth aligned with sun position. It's determined by two angels; they are altitude angle (α) and azimuth angle θ_s

The altitude angle is the angular height of the Sun is measured from the horizontal. The altitude angle can be given by:

$$sin\alpha = sinLsin\delta + cosLcos\delta cos\omega \tag{1}$$

where,

L: attitude of the location

δ: Angle of declination

ω: Hour angle.

The declination angle is between Earth Sun vector and equatorial plane its calculated degree, arguments to trig function noted in radian mode.

$$6_s = 23.45^0 sin\left[\frac{2\pi(N-81)}{365}\right] \tag{2}$$

Noted hour angel ω is the angular displacement of the sun local point is given by:

$$\omega = 15^0(\mathrm{AST} - 12\mathrm{h}) \tag{3}$$

AST the true daily motion of solar time is given by a daily apparent solar motion of true observed sun. AST is constructed on the actual solar day. The two interval falls between two consecutive returns of local meridian and the sun. Solar time is illustrious as,

$$AST = LMT + E_0T \pm 4^0/(LSMT - LOD) \tag{4}$$

LMT: Local meridian time

LOD: Longitude

LSMT: Local standard meridian time

E_0T: Equation of time.

AST: Apparent solar time

The *LSTM* is a reference meridian used for a particular time zone, used for Greenwich Mean Time.

LSTM is given by:

$$LSMT = 15^0 T_{GMT} \tag{5}$$

The E_0T is the difference between apparent and mean solar times, both taken at a given longitude at the same real instant of time.

E_0T is given by:

$$EoT = 9.87\sin(2B) - 7.53cosB - 1.5sinB \tag{6}$$

where,

B can be given by;

$$B = \frac{2\pi}{365}(N - 81) \tag{7}$$

where,

N: Day number defined as the number of days elapsed in a given year up to a particular date.

Angular displacement of the Sun reference line from the source axis. The azimuth angle can be given by:

$$sin\theta = \frac{cos\delta sin\omega}{cos\alpha} \tag{8}$$

The solar source model is to estimate the emitted radiation from the Sun. The function of the temperature is described as radiant energy of emitting objects.

We associate radiating energy to the blackbody. A blackbody is defined as a perfect absorber and emitter. A perfect absorber can absorb all of the received energy with any reflections.

Planck's law describes the wavelengths emitted by a blackbody at a specific temperature as follows:

$$E_\lambda = \frac{3.74 \times 10^8}{\lambda^5\left[exp\frac{14{,}40}{\lambda T}\right] - 1} \tag{9}$$

E_λ: Total emissive per unit area of blackbody emission rate (W/m2 μm)

T: Absolute temperature of the blackbody (K)

λ: Wavelength (μm).

Solar radiation value outside the atmosphere varies as the Earth orbits the Sun. Therefore, the distance between the Sun and the Earth must be considered in modeling the extraterrestrial solar radiation. Thus, the (G_{ex}) is given by:

$$G_{ex} = G_0\left[\frac{R_{av}}{R}\right]^2 \tag{10}$$

where.

G_{ex}: Extraterrestrial solar radiation.

G_0: Solar constant

R_{av}: Mean distance between the Sun and the Earth

R: Instantaneous distance between the Sun and the Earth depends on the day of the year or day number.

There are different approximations for the factor (R_{av}/R) in the literature. A recommended approximation can be given by:

$$\left[\frac{R_{av}}{R}\right] = 1 + 0.0333cos\left[\frac{2\pi N}{365}\right] \tag{11}$$

By substituting Eqs. (11, 10)

The extraterrestrial solar radiation unit of time falling at a right on square meter of a surface can be given by:

$$G_{ex} = G_0\left(1 + 0.0333cos\left[\frac{2\pi N}{365}\right]\right) \tag{12}$$

Once the surface faces the Sun (normal to a central ray), the solar irradiance falling on, is G_{ex}, utilizes maximum solar radiation at that distance. If the surface is not normal to the Sun, the solar radiation drops on it will be decrease by cosine of the angle between the surface normal and a central ray from the Sun.

Thus, the extraterrestrial solar radiation on a horizontal surface located in a specific location (G_{exH}) can be calculated by:

$$G_{exH} = G_{ex}cos\varphi \tag{13}$$

where.

φ: Solar zenith angle.

The solar zenith angle value is equal to the altitude value, and thus Eq. (13) can be rewritten as follows:

$$G_{exH} = G_0\left[1 + 0.0333cos\left[\frac{360N}{365}\right]\right]sinLsin\delta + cosLcos\delta cos\omega \tag{14}$$

Finally, the total extraterrestrial solar energy E_{ex} (Wh/m2) is calculated as follows:

$$E_{ex} = \int_{T_{sr}}^{T_{ss}} G_{exH}dt \tag{15}$$

There are several components of a solar radiation on a tilted surface are in addition to the direct $(G_{B.\beta})$ and diffuse $(G_{D.\beta})$ solar radiation, reflected solar radiation (G_R) is added to form the global solar radiation incident on a tilted surface.

$$G_{T,\beta} = G_{B,\beta} + G_{D,\beta} + G_R \tag{16}$$

The solar energy components on a horizontal surface as follows:

$$G_{T,\beta} = G_BR_B + G_DR_D + G_T\rho R_R \tag{17}$$

where,

RB, RD, and R_R:are coefficients.

ρ: Ground Aledo.

R_B: Ratio between global solar energy on a horizontal surface and global solar energy on a tilted surface.

R_D: Ratio between diffuse solar energy on a horizontal surface and diffuse solar energy on a tilted surface,

R_R: Factor of reflected solar energy on a tilted surface.

The finding of solar energy components on a tilted surface is to estimate the coefficients R_B,R_D, and R_R. Used model for calculating R_B is the Liu and Jordan model [Liu, B.Y. and Jordan, R.C., 1963]

$$R_B = \frac{\cos(L-\beta)cos\delta sin\omega_{ss} + \omega_{ss}\sin(L-\beta)sin\delta}{cosLcos\delta sin\omega_{ss} + \omega_{ss}sinLsin\delta} \tag{18}$$

The surfaces in the southern hemisphere, the slope toward the equator R_B is given as:

$$R_B = \frac{\cos(L+\beta)cos\delta sin\omega_{ss} + \omega_{ss}\sin(L+\beta)sin\delta}{cosLcos\delta sin\omega_{ss} + \omega_{ss}sinLsin\delta} \tag{19}$$

The most recommended formula R_R is:

$$R_R = \frac{1-cos\beta}{2} \tag{20}$$

R_D Have been classified into isotropic and anisotropic models.

Isotropic radiation has the same intensity regardless of the direction of measurement. Isotropic radiator is a uniform radiation from given point.

Model used of the Liu and Jordan an isotropic diffuse with R_D been expressed:

$$R_R = \frac{1-cos\beta}{2} \tag{21}$$

$$R_D = \frac{1}{3[2+cos\beta]} \tag{22}$$

$$R_D = \frac{3+\cos(2\beta)}{4} \tag{23}$$

$$R_D = 1 - \frac{\beta}{180} \tag{24}$$

The behavior of isotropy and anisotropy radiation. Isotropy indicates identical properties in all direction moreover measure aligned different axes of materials physical property (absorbance, refractive index, density). On the other hands Anisotropy model is based on anisotropic radiation measurements varies in direction because of radiates nonuniformly in all direction. Anisotropic moles are noted R_D:

$$R_D = \frac{G_B}{G_T}R_D + \left(1 - \frac{G_B}{G_T}\right)\left(\frac{1+cos\beta}{2}\right) \tag{25}$$

$$R_D = 0.51R_B + \frac{1+cosTLT}{2} - \frac{1.74}{1.26\pi}\left[sin\beta = \left(\beta\frac{\pi}{180}\right)cos\beta - \pi sin^2\left(\frac{\beta}{2}\right)\right] \tag{26}$$

$$R_D = \frac{G_B}{G_T} R_B + \left(1 - \frac{G_B}{G_T}\right)\left(\frac{1 + cos\beta}{2}\right)\left(1 + \sqrt{\frac{G_B}{G_T}} sin^3\left(\frac{\beta}{2}\right)\right) \quad (27)$$

The circuit of a solar cell is a current source connected in parallel with a diode as shown in Figure. The output of the current source is directly related to the light emitted on the cell. Nor current or voltages are produced during darkness of solar cell. During light falls over solar cell it generates diode current Fig. 1.

The diode, D, determines the I–V characteristics of the cell. A series resistance, R_s, while the shunt resistance, R_{SH}

Assumed that $R_s = 0$ and R_{SH} = infinity. The net current of the cell is the difference between the photocurrent, I_L, and the normal diode current is:

$$I = I_L - I_0\left(e^{\frac{q(V+IR_s)}{nkT}} - 1\right) - \frac{V + IR_s}{R_p} \quad (28)$$

The photocurrent, I_L, depends on reference first and second temperatures, T_1 and T_2, respectively, illustrated by:

$$I_L = I_L(T_1) + K_0(T - T_1) \quad (29)$$

where,

$$I_L(T_1) = I_{scT1,nom}\left(\frac{G}{G_{nom}}\right) \quad (30)$$

$$K_0 = \frac{I_{scT_2} - I_{scT_1}}{T_2 - T_1} \quad (31)$$

where,

G: Present solar radiation

G_{nom}: Solar radiation at the reference test.

The saturation current of the diode, I_0, is given by:

$$I_0 = I_{0T_1}\left(\frac{T}{T_1}\right)^{\frac{3}{n}} e^{\frac{qV_{qT_1}}{nk\left(\frac{1}{T} - \frac{1}{T_1}\right)}} \quad (32)$$

where,

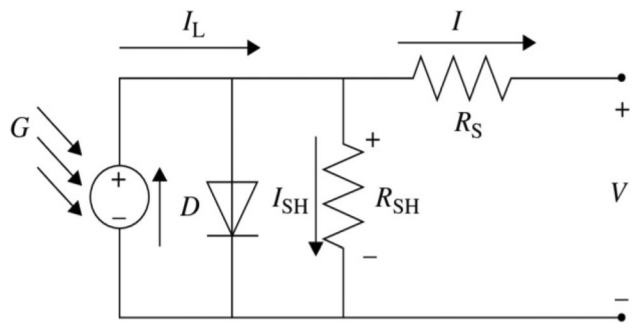

Fig. 1 **a** Solar cell is a current source connected in parallel with a diode

$$I_{0T_1} = \frac{I_{scT_1}}{(e^{\frac{qV_{ocT_1}}{nkT_1}} - 1)} \quad (33)$$

The series resistance of a solar cell is given by:

$$R_s = -\frac{dV}{dI_{V_{oc}}} - \frac{1}{X_V} \quad (34)$$

where,

$$X_V = I_{oT_1}\frac{q}{nkT_1} e^{\frac{qV_{ocT_1}}{nkT_1}} \quad (35)$$

3.2 Operations modelling

A typical V–I characteristic of a solar cell at a certain ambient irradiation, G, and fixed cell temperature.

For a resistive load, the load characteristic is a straight line with slope I/V = 1/R.

If the load is small, the cell operates in the regions M–N of the curve; cell behaves as a constant current source that is almost equal to the short circuit current. When we have high load, the cell function in the P–S zone of the curve and the cell behaves as a constant voltage source that is almost equal to the open circuit voltage.

The short circuit current, I_{sc}, is the greatest value of current generated by a solar cell. When V = 0 the short circuit condition produced. When the photocurrent is zero, the open circuit voltage corresponds to the voltage drop across the diode. It reflects the voltage of the cell at no light conditions, and it can be expressed as:

$$V_{oc} = \frac{nkT}{q} ln\left(\frac{I_L}{I_0}\right) = V_t ln\left(\frac{I_L}{I_0}\right) \quad (36)$$

where ($V_t = mkT_c/q$): is known as the thermal voltage.

T: absolute cell temperature.

Cells are connected together in series to increase the voltage. Several of these series strings of cells may be connected together in parallel to increase the current as well. This package is called as a PV module or PV panel.

The relations between a PV module panel, the cell's voltage (V_c) and current (I_c) and the module's voltage (V_m) and current (I_m) are given by the following equations:

$$I_M = N_{pM} I_c \quad (37)$$

$$V_M = N_{sM} V_c \quad (38)$$

$$R_{sM} R_{sM} = \frac{N_{sM}}{N_{pM}} R_{sc} \quad (39)$$

where,

N_{sM}: Number of series cells.

N_{pM}: Number of parallel cells.

R_{sM}: Equivalent series resistance of the PV module.

Most PV manufacturers provide temperature elements for their crystalline PV modules based on the NOCT as the cell temperature (T_c), which has a standard equation of

$$T_c = T_a + \frac{G}{800}(NOCT - 20^{\circ}C) \tag{40}$$

where,

T_c: Cell temperature.

T_a: Ambient temperature.

G: Instant solar radiation.

NOCT: nominal operating cell temperature is defined as the temperature reached by the open circuit cells in a photovoltaic module under conditions of 800 W/m2 irradiance on the cell surface, 20 °C air temperature, and 1 m/s wind velocity.

As mentioned in Eq. (28), the characteristic of a PV cell is expressed by its relationship between current and voltage (I–V) and power and voltage (P–V) at specific solar radiation and temperature levels. Assuming the R_p in Eq. (27) is too large, and then the I–V a solar cell expressed as:

$$I = I_L - I_0\left[\exp\left(\frac{q(V + IR_s)}{nkT_c}\right) - 1\right] \tag{41}$$

In addition to that, in order to simplify the characterization of a solar cell, assume that

$$\frac{q}{nk} = k_1 \tag{42}$$

$$\frac{T_c}{k_1} = a \tag{43}$$

Based on this, Eq. (27) can be rewritten as follows:

$$I = I_L - I_0\left[exp\left(\frac{q(V + IR_s}{a}\right) - 1\right] \tag{44}$$

Solving Equation for V results:

$$V = a.ln\left(\frac{I_L - I}{I_0} + 1\right) - IR_s \tag{45}$$

The light-generated current, I_{ph}, is linearly proportional to the global solar radiation and is also logarithmically dependent on the operating temperature of cell, T_c. Therefore, I_{ph} can be expressed:

$$I_L = (k_2 + k_3T_c)G_T \tag{46}$$

Finally, the diode saturation current depends on the operating temperature of cell T_c: as follows:

$$I_0 = k_4T_c^3 exp\left(-\frac{k_5}{T_c}\right) \tag{47}$$

Four statistic errors are used, which are.

(1) (*MAPE*): Mean absolute percentage error.
(2) (*MBE*): Mean bias error.
(3) (*MAE*): Mean absolute error.
(4) (*RMSE*): Root mean square error.

The general accuracy of a neural network can be highlighted by *MAPE*. *MAPE* can be defined as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{M - P}{M}\right| \tag{48}$$

where,

M: Measured data.

P: Predicted data.

The information of long-term performance of the neural network model can also be evaluated by *MBE*. *MBE* can be calculated as follows:

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(P_i - M_i) \tag{49}$$

The mean absolute error (*MAE*): is a measure of errors between paired observations expressing the same phenomenon. Examples of Y versus X include comparisons of predicted versus observed, subsequent time versus initial time, and one technique of measurement versus an alternative technique of measurement. *MAE* is calculated as:

$$MAE = \sum_{i=1}^{n}(\frac{|y_i - x_i|}{n}) \tag{50}$$

where,

y_i: Prediction

x_i: True value

n: Total number of data points

The final statistic error is *RMSE*; it represents the measurement of the variation of the predicted data around the measured data.

The short-term performance information of the model can be evaluated by *RMSE*:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - M_i)^2} \tag{51}$$

3.3 Fault modelling

Photovoltaics (PV) component/module, which involves solar cells connected in series. The equation below illustrates the PV voltage as well as current features.

$$I = I_{PH} - I_0\left[\exp\left(\frac{V + R_sI}{\alpha V_t}\right) - 1\right] - \frac{V + R_sI}{R_{sh}} \tag{52}$$

whereby voltage of the PV is denotated by V while the photovoltaics current (output) is denotated by *I*. Therefore, $V = N_{cell}V_{cell}$;N_{cell} as well as$I = I_{cell}$; this refers to the solar cell's quantity connected in series. The produced PV-photocurrent is represented byIPH:$I_{PH} = I_{PH}, cellI_0$;. Also, the thermal voltage is represented by $V_t = N_{cell}\frac{KT}{q}$ whereas

the diode's reverse saturated I is represented by $I_0 = I_0, cell V_t$. The parallel as well as series equal or correspondent resistance(s) is denoted by R_{sh} as well as $R_s, respetcively : R_s = N_{cell} R_{s,cell}$

Whereas $R_{sh} = N_{cell} R_{sh,cell}$ (Villalva et al. 2009).

3.3.1 Strategy used to identify faults

This study focuses on detecting photovoltaics system's faults, which are associated with the degraded photovoltaics array and photovoltaics array's partial shading, short circuits, as well as open circuits. Silvestre et al. (2014) formulate Eqs. (55, 56) for defining the photovoltaics system's indicators of current and voltage, as well as describing it using defect feature quantity. The Eqs. (55, 56) are shown below:

$$R_V = \frac{V}{V_{oc}} \tag{53}$$

$$R_I = \frac{I}{I_{sc}} \tag{54}$$

where $R_V - voltageindcator$, $whereasR_I - indicatorofcurrent/I$ in a photovoltaics module. Similarly, I-is current and V-is voltage of the photovoltaics module in its maximum power point. The photovoltaics array's short-circuit current is-I_{sc} whereas the voltage open-circuit is represented as V_{oc}. According to (Silvestre et al. 2014), Eqs. 57 and 58 below represents theV_{oc},I_{sc}

$$I_{sc} = N_P(\frac{I_{scm\_STC}}{1000} G + K_I(T - T_{STC})) \tag{55}$$

$$V_{oc} = N_s(V_{ocm\_STC} + K_V(T - T_{STC}) + V_t \ln\left(\frac{\frac{I_{sc}}{N_p}}{I_{scm\_STC}}\right)) \tag{56}$$

In which:

N_s-refers to the number of the photovoltaic modules of a photovoltaic string

N_P-refer to the number of the photovoltaic array's photovoltaic string

V_t-implies the photovoltaics's thermal voltage.

$T -$ refers to photovoltaics' temperature

G-refers to photovoltaics component that receives irradiations

K_V-refers to an open-circuit voltage's Temperature coefficient

K_I-refers to short-circuit current's Temperature coefficient

At standard testing settings, which are $T_{STC} = 25^0 C$ [temperature] as well as $G_{STC} = 1000 W/m^2$[irradiation], the V_{ocm_STC} -refers to photovoltaics module's open-circuit voltage whereas I_{scm_STC} refers to the photovoltaic modules short-circuit current.

Silvestre et al. (2014) stated that no fault operations (i.e., no faults in the indicators of current or voltage) in the photovoltaics component as illustrated in Eqs. 55 and 56 is represented in Eqs. 58 and 59 below:

$$R_{VM} = \frac{V_m}{V_{oc}} \tag{57}$$

$$R_{IM} = \frac{I_m}{I_{sc}} \tag{58}$$

whereby:

$I_{m,} V_m$: The photovoltaics module's output current as well as voltage of the PV system in fault-free operation

R_{IM}: Photovoltaics module's current indictors of the PV system in fault-free operation

R_{VM}: Photovoltaics module's voltage indictors of the PV system in fault-free operation

Additionally, Eqs. 61 and 62 below shows the V_m, I_m of the photovoltaics array at a maximum power point (MPP)during a fault-free state. (Silvestre et al. 2014).

$$I_m = N_P(\frac{I_{mm\_STC}}{1000} G + K_I(T - T_{STC})) \tag{59}$$

$$V_m = N_s(V_t \ln\left(1 + \frac{I_{sc} - I_m}{I_{sc}}\left(e^{\frac{V_{oc}}{N_s V_t}} - 1\right)\right) - \frac{I_m}{N_p} R_s \tag{60}$$

whereby at normal testing conditions:

R_s: Photovoltaics component's series equal resistance.

I_{mmSTC}: The current at the maximum power point of the photovoltaics module at Standard Test Conditions.

3.3.2 Defining the verges of detecting faults

A. Open-circuit fault

Yahyaoui and Segatto (2017) stated that the faulty string's I tend to be equivalent with the photovoltaics array's output current's decreased percentage, which usually occurs whenever there is a faulty open circuit. Therefore, the I-indicators in this condition is represented in Eq. 63 below:

$$R_{IO} = \frac{(N_p - 1)}{N_p} \frac{I_m}{I_{sc}} = \alpha R_{IM} \tag{61}$$

where R_{IO} -refers to the current indicator when there's an existence of an open circuit fault in a photovoltaics array where α tend to be:

$$\alpha = 1.0 - \frac{1.0}{N_p} \tag{62}$$

Therefore, the fault detection threshold of open circuit faults is be defined by Eq. 65 (Yahyaoui and Segatto 2017).

$$T_{IO} = \varepsilon R_{IM} \tag{63}$$

whereby:

ε-refers to a fault detection allowable offset coefficient and $\varepsilon = 2.0\%$ (Silvestre et al. 2014).

T_{IO}-refers to an open circuit fault's threshold. Silvestre et al. (2014) argued that T_{IO} Value should be above the $R_I - value$ (see Eq. 56), particularly if one or more the photovoltaics strings given has an open circuit fault.

B) Short-Circuits fault

Likewise, if there is an occurrence of short-circuit photovoltaics module in any photovoltaic strings, the short-circuit voltage of the photovoltaics module is equal to the decreased portion of the output voltage of the photovoltaics array. Therefore, the following equation (Eq. 66) is used to calculate voltage indicator Yahyaoui and Segatto (2017).

$$R_{VS} = \frac{(N_s - 1)}{N_s} \cdot \frac{V_m}{V_{oc}} = \beta R_{VM} \tag{64}$$

where: R_{VS}-refers to an indicator of voltage in presence of a short-circuit in the photovoltaics occurring within photovoltaics string(s). Equation 67 show β, illustrated below:

$$\beta = 1 - \frac{1}{N_s} \tag{65}$$

Therefore, the fault detection threshold of short-circuit faults are be defined by (Yahyaoui and Segatto 2017).

$$T_{VS} = \varepsilon \beta R_{VM} \tag{66}$$

$R_V - value$ (Provided in Eq. 55) should be lower than T_{VS} values.

T_{VS} -refers to short-circuit fault's threshold whenever one/more photovoltaic module string offered with a short circuit fault.

C) Partial shading fault

Partly shading of photovoltaics array causes drastic decrease in the current output since part of the array receives solar irradiation. Therefore, Eqs. 69 and 70 below are used to calculate the indicators of current and voltage in an event of partly or incomplete shading

$$R_{VP} = \frac{V_{mp}}{V_{oc}} \tag{67}$$

$$R_{IP} = \frac{I_{mp}}{I_{sc}} \tag{68}$$

where:

I_{mp} refers to a photovoltaics array's current output at a maximum power point

V_{mp} refers to a photovoltaics array's voltage output maximum power point

R_{IP} refers to fractional shading fault's the indicators of current

R_{VP} refers to partly shading fault's the indicators of voltage

Equation 71 and 72 are used to calculate photovoltaics array's $I_{mp} aswell as V_{mp}$ whenever there is a partly shading, where the photovoltaics array receives maximum irradiance in unshaded side.

$$I_{mp} = N_p \left(\frac{I_{mm\_STC}}{1000} G_P + K_I (T - T_{STC}) \right) \tag{69}$$

$$V_{mp} = N_s \left(V_t \ln \left(1 + \frac{I_{sc} - I_{mp}}{I_{sc}} \left(e^{\frac{V_{oc}}{N_s V_t}} - 1 \right) \right) \right) - \frac{I_{mp}}{N_p} R_s \tag{70}$$

where G_P -gets photovoltaics optimum irradiance in partial-shading states. Thus, Eq. 73 below is instrumental in detecting and identifying the threshold of faults during a partly-shading.

$$T_{IP} = \varepsilon R_{IP} \tag{71}$$

where T_{IP}-refers to the partial-shading fault's thresholds. T_{IP} values should be higher than the R_I Values (provided in Eq. 56) whenever all or partial shaded for the photovoltaics array.

where T_{IP} is the threshold of partial shading faults, and when partial or all shaded for the PV array, the value of R_I (given as Eq. (56) must be below the value of T_{IP} Table 1.

3.4 Problem definition and proposed Model

Having identified the models for the environment, the operation, and the generated signals, now the problem of the accurate generated power prediction can be formulated as below.

$$\mathrm{F(A*)} = \left(\max_{0 \le \eta \le 1} P, \eta, \min_{0 \le Q_S \le 1} Q_S \right)$$

The above definition is a min–max-optimization problem, where,

P: PV power output (LSTM and GRU Final Comparative Results) Table 3.

η: Is conversion efficiency of PV module (Maximum efficiency for GRU and LSTM algorithms) (Durisch et al. 2007) Table 4.

Q_S: Respects the thermal energy losses through radiation and convection heat transfer from modules (Castillejo-Cuberos and Escobar 2020; Lave et al. 2015; Van Haaren et al. 2014) Figs. 16, 17.

While the optimization variables can be defined as:

$A = \{S, \beta, \gamma, h_r, h_c, l, G_T, V_{WS}, \varphi, \theta, \int, \Lambda, £, Ь, \alpha, \mu, \xi, \vartheta, \varpi, \Join, \Pi\cup\}$

However, a problem with all these variables can be defined as an NP-Hard problem. And cannot be solved using tradition optimization techniques. Therefore, to solve this problem we will follow the below methodology Table 2.

Table 1 Nomenclature

Parameter	Description	Value
q	Charge on an electron	$1.602 \times 10 - 19$ [C]
K	Boltzmann's constant	$1.380 \times 10 - 23$ [J/K]
A	Ideality factor of diode	1.3
N_s	Number of series connected cells (diodes)	72
R_s	Series resistance of the PV module	0.221 [Ω]
T	The cell's temperature	25 [°C]
$V = V_{mp}$	The voltage at the maximum power point	26.3 [V]
T_0	Real-time temperature	273.15[K]
K_i	Temperature coefficient of	
I_{sc} cell short circuit current	0.058 [%/ C]	
N_P	Total parallel cells	1
I_{sc}	Short-circuit current at STC	9.06 [A]
V_{oc}	Open circuit voltage	46.22 [V]
G = G_{ref}	Solar Irradiance	1000 at STC
T_{ref}	Reference temperature	25 [°C]
P_m	Maximum power at Standard Test Conditions (STC)	320 [W]
V_m	Maximum power voltage	37.38 [V]
I_m	Maximum power current	8.56 [A]
I_{Ph}	Photocurrent of a solar PV cell generated due to solar irradiation	8.214 [A]
E_g	Forbidden Energy band gap, for silicon	1.12 [ev]
K_v	Temperature coefficient of Open-Circuit Voltage	−80 [mV/°C]
K_1	Temperature coefficient of short circuit Current	0.065 [%/°C]
ε	A fault detection allowable offset coefficient	2.0%
R_p	Parallel resistance of the PV module	415.405 [Ω]

After having modelled the environment, the operation and the signals to generate the patterns that is like the real patterns of the physical instruments, this pattern is clustered using k-mean clustering algorithm.

After that, we implement a monte-carlo simulation with the identification of all the bounds of the remained stochastic variables, and the optimization outputs, as mentioned in the above table. Inside the iterations of this simulation, Particle Swarm Optimization (PSO) algorithm, which is a metaheuristic stochastic-based algorithm is applied in order to identify the optimal values of the selected optimization variables.

After finding the optimal values of the parameters, these parameters as well as the output values, will be used in addition to the current and previous loads in order to produce the short-term prediction, for example, LSTM or GRU algorithms will be used in this phase. The below figure shows the block diagram of our proposed solution. The below figure summarizes this proposed method Fig. 2.

In the prediction phase,initially LSTM, see Fig. 3a, has been selected with training input is the output of the clustering phase with size of 1400 × 34and then this and 300 hidden layers with three output signals, representing the next or the future temporal values of n, p and the Qs in the LSTM we have used look up in order to use only the most useful or the most related samples in building that pattern. Moreover, GRU, see Fig. 3b, followed the same structure to compare both algorithms using the same benchmark in order to be able to figure out which one provides us with the most accurate future temporal value, and which one provides us with a most with the fastest processing time.

3.5 Datasets

Regarding the load, we have acquired it from a with short-term slots of a frequency of 5 min which is very useful for our application in short-term prediction. This dataset (Dataset employed by this research can be retrieved from UK Smart Grid Industry 2021–2024) contains 371 samples each sample is 5 min separated from the other sample from the period of the first of January to the second of January in the year in the previous year 2020.This dataset was

Table 2 While the definitions of the used notations and their ranges, are illustrated in the below table

Notation (Variable)	Definition	Range	Ref
S	is the temperature surface area of PV module	−55 to + 150	Saloux et al. (2011); Sun et al. (2017)
β	is temperature coefficients of PV module	25 to 65	Sun et al. (2017)
γ	is solar irradiance coefficient of PV module	200 to 1000 W/m2	Mills and Schleich (2012)
h_r	Heat transfer coefficient of radiation	0 to 1	Cao (2010)
h_c	Heat transfer coefficient of convection	2.5 to 5 W/(m2 × K)	Patil and Vijay (2012)
l	BOARD length	150 to 165 cm	Sun et al. (2017)
G_T	is surface solar radiance flux on module plane	200 to 1000 W/m2	Mills and Schleich (2012)
V_{WS}	Wind speed	25 to 40 mph	Sun et al. (2017)
φ	Maximum Power Point	0.78 to 0.92	Sarvi et al. (2015)
θ	Life Cycle Impact Assessment	2742 to 2857 kWh/kWp	(Lamnatou et al., 2015)
∫	Life Cycle Inventory	10 to 25	Sun et al. (2017)
ʎ	auxiliary electricity demand	1 to 160 kW	Sun et al. (2017)
£	Cleaning of the panels	0.2 to 0.325	Al-Housani et al. (2019)
Б	Maintenance	$13 to $25/kW/yr	Al-Housani et al. (2019)
α	Decommissioning, dismantling	20 to 30 years	Mahani et al. (2019)
μ	Waste processing	20 to 30 years	Mahani et al. (2019)
ξ	Front electrode deposition	20 to 30 years	Mahani et al. (2019)
ϑ	Electron transport layer deposition	20 to 30 years	Mahani et al. (2019)
ϖ	Active layer deposition	20 to 30 years	Mahani et al. (2019)
Ж	Back electrode deposition	20 to 30 years	Mahani et al. (2019)
Ш	Hole transport layer deposition	20 to 30 years	Mahani et al. (2019)

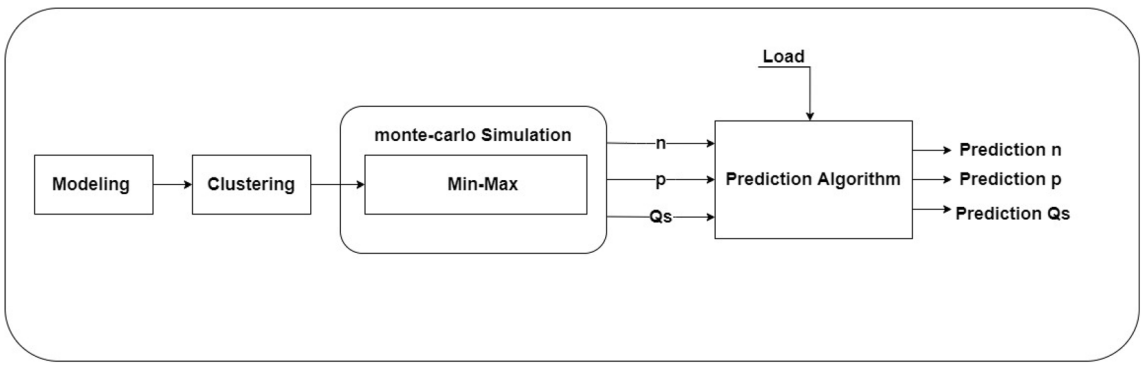

Fig. 2 Methodology Block Diagram

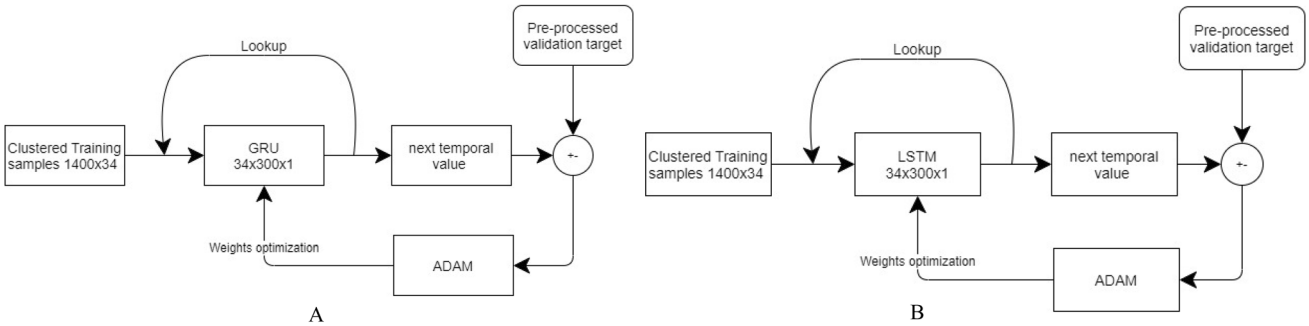

Fig. 3 GRU and LSTM Structures

generated from a real site located in London city in the UK. The exact coordinates are 51.5074° N, 0.1278° W.

4 Implementation and results

The modeling of the sun position for the selected location in London city in the UK, according to the location of the site from which we obtained the datasets which contains the temperature and solar radiation is shown below.

Figure 4 shows, the first plot which is the alpha angle of the sun position from day 0 to day 350. So, for entire year according to the position of the coordinate of London city UK. Where the second plot is the Theta for the period time for the same position for the same location.

Figure 5, below chooses the hourly extraterrestrial solar radiation profile for 16 days of January, each plots of this 16 plots for a specific date is shown that the plots almost similar to each other but the related plot for corresponding days increase with the increase of that day which means the peak value of each day as increasing according to the day number for example day 1 we have the value around 350 for day 2 its around 360 and so on the big value.the x-axis here is (LMT) and y -axis the$(GextH)$.While Fig. 6 shows five minutes step for only one day Fig. 7.

Figure 7a shows the solar radiation for one day in January for 5 min step. And 7b the diffuse solar radiation for the same period of time also the step time 5 min.

Figure 8 shows sample of the global solar radiation and diffuse solar radiation for sample days, the first day, day number 50, 100, 180, 250, and 360, which choose the variation of the solar radiation in this day. In this result we have used 60 min step so hourly based solar radiation for each day this value helps us to predict the next solar radiation for following minutes.

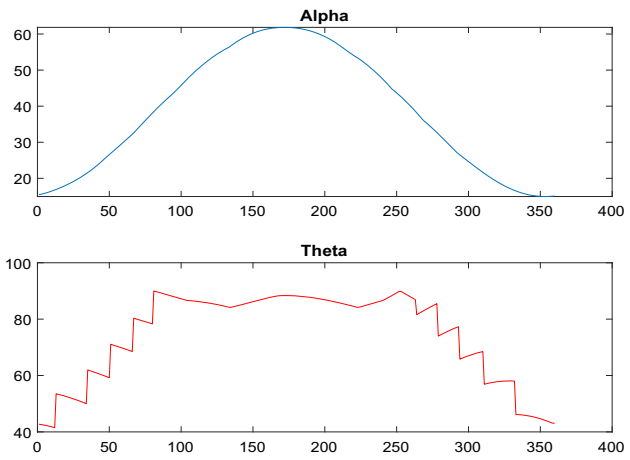

Fig. 4 The alpha angle of the sun position

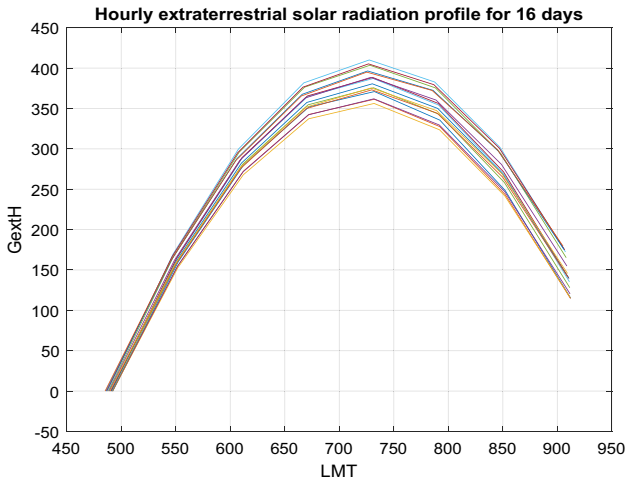

Fig. 5 Five minutes step for only one day

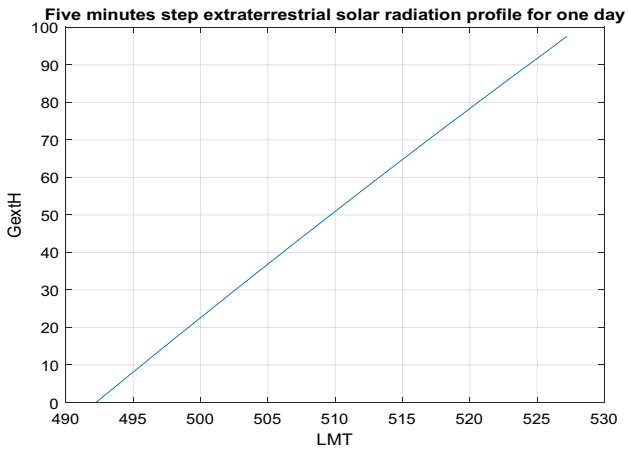

Fig. 6 The hourly extraterrestrial solar radiation

Figure 9 shows the 5 min step solar radiation for the same days which also helps us to predict the short-term solar radiation for each 5 min, as we can see the pattern is almost similar, it looks smoother than the previous figure.

Figure 10a shows the sun position for 5 min of one day in January the sun position alpha angle and theta angle.

it shows corresponding voltage, current and power for each sun position Fig. 10. between the sun position and generated power.

It appears from the above figures, that the sun position for the same day in different years is not the same. This is a physical fact that applies on all the natural resources and factors, such as wind, dust, shading, etc. Consequently, and as clearly shown in the figures above, the behavior of the natural factors is totally stochastic and cannot be predicted. Since the generation of the electricity is based and affected by these resources and factors. This leads to the fact, that the prediction of the generated energy based on renewable resources is a challenging problem. Therefore, this work

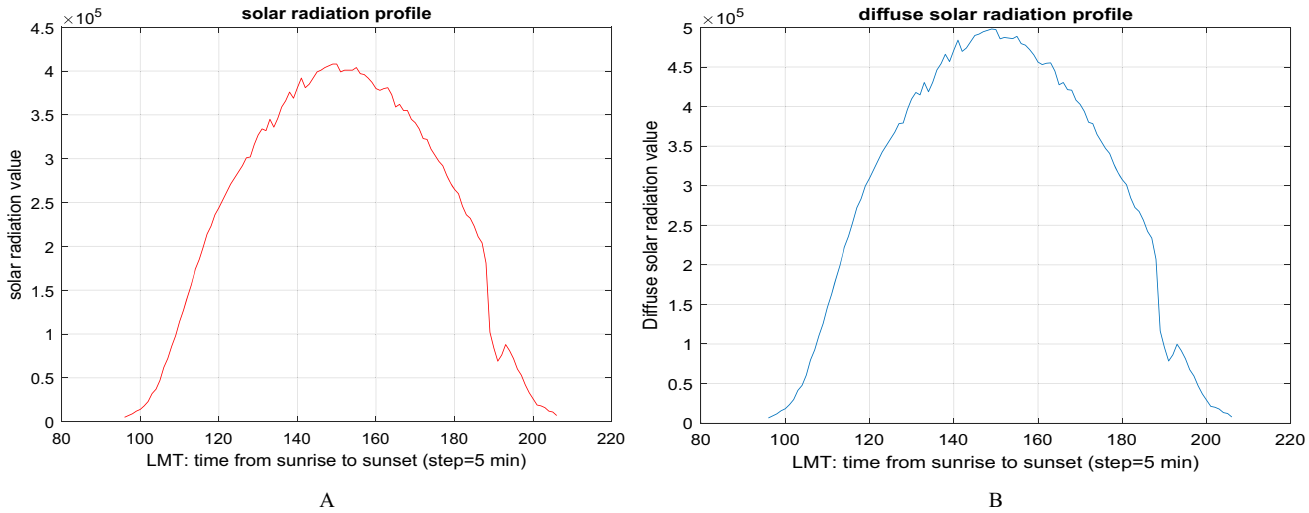

Fig. 7 The solar radiation for one day and Diffuse solar radiation

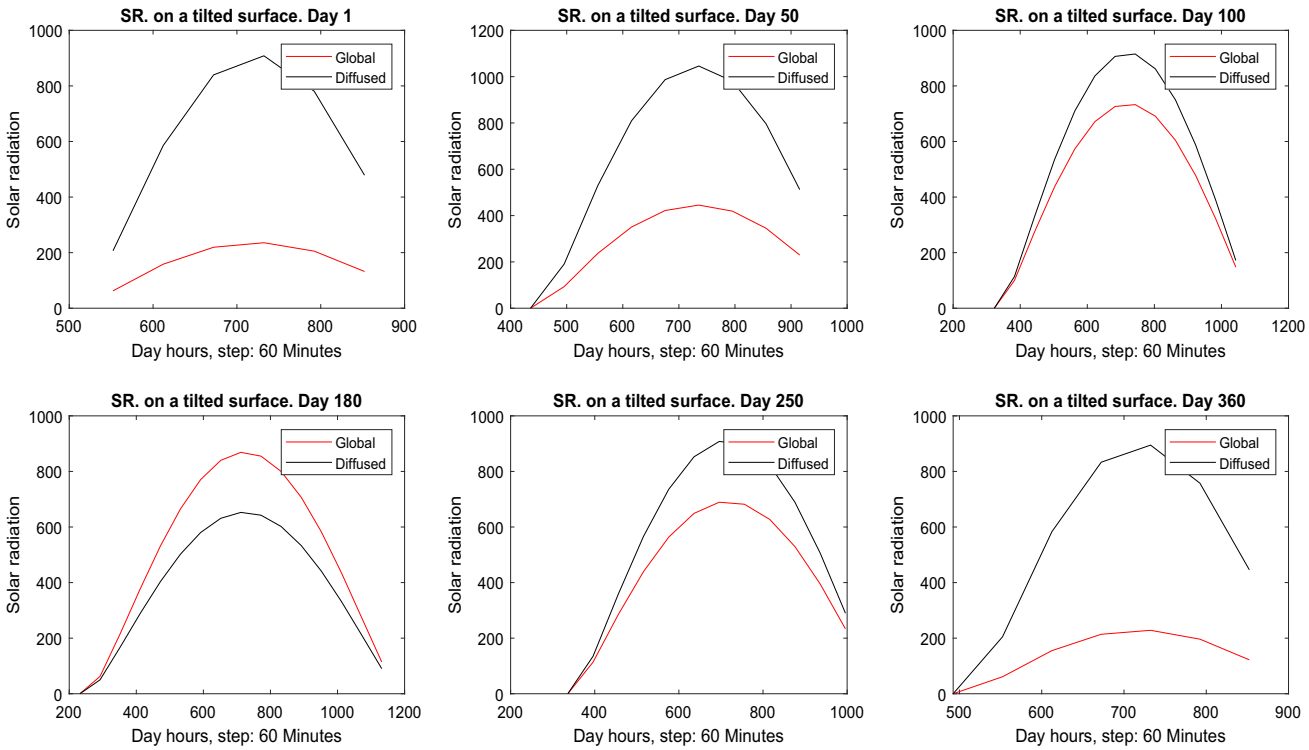

Fig. 8 Global solar radiation and diffuse solar radiation

contributes to this field by developing a methodology to allow the process of short-term prediction for the short-term future generated energy that is based on and affected by renewable resources and factors.

4.1 Data Generation and Montecarlo simulation

Figure 11 shows the generated signals that contains1000 normal and 1000 faulty signals, each of which contains 34 features. However, these signals have been also generated using the stochastic features of the Montecarlo simulation. To test the Prediction model Fig. 12.

In This phase we need to know the status of the pattern. Weather this signal is normal or a fault signal. However, in our problem we don't have target. There is no data set with a target. Also, the behavior of the normal traffic and that the behavior faulty traffic in the real life is a little bit of stochastic that does not follow a stable pattern. Therefore,

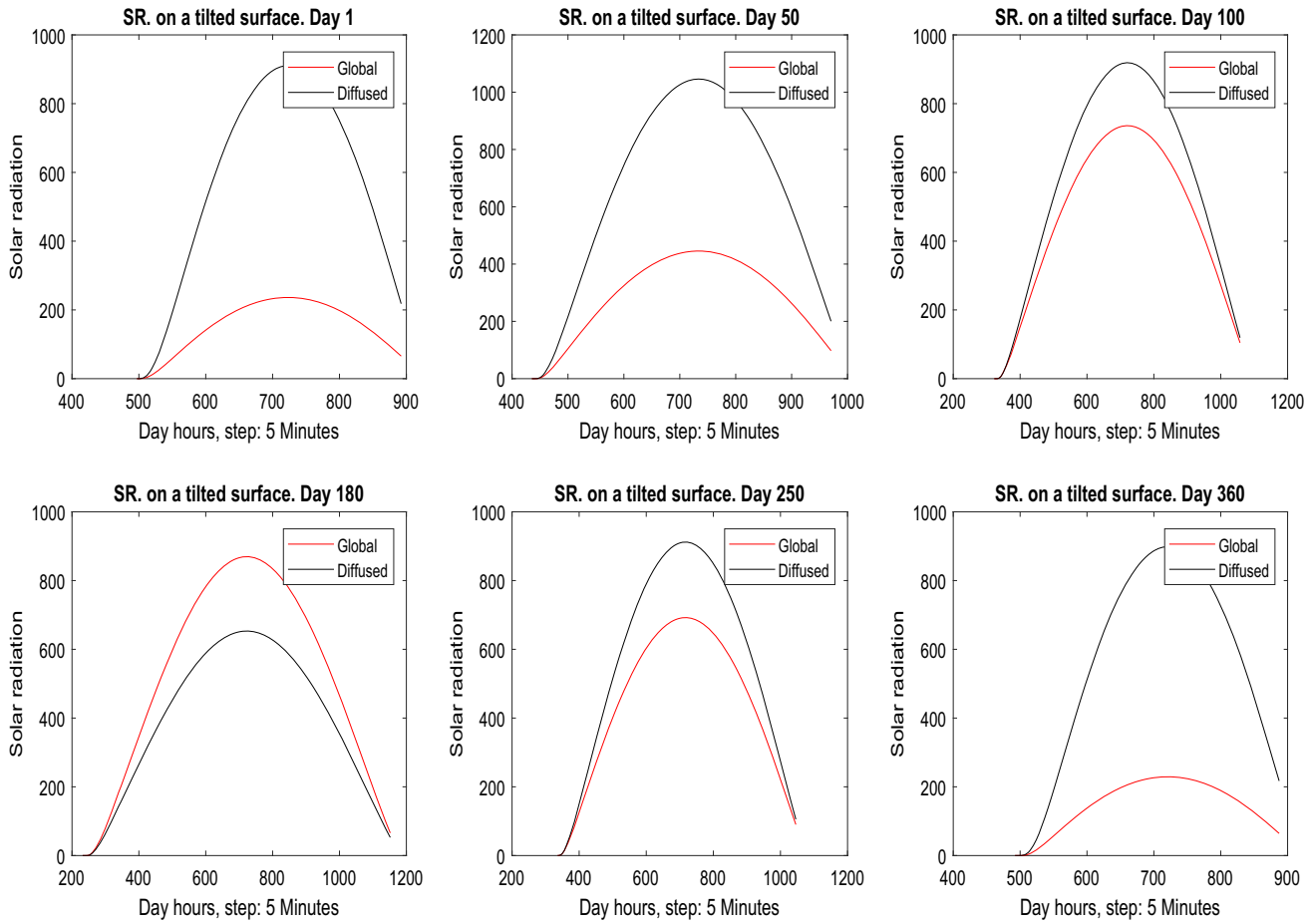

Fig. 9 5 min step solar radiation

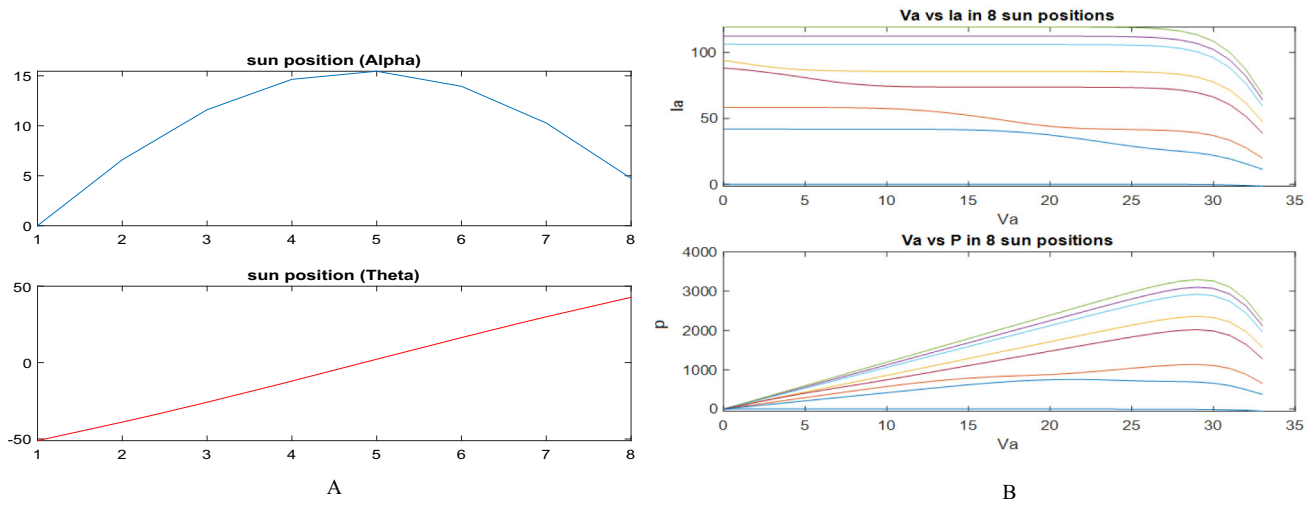

Fig. 10 The sun position alpha angle and theta angle and Sun position and generated power

we will not be able to identify the exact features of the input pattern and the target for them. Therefore, our problem can be defined as a clustering problem, In order to solve this clustering problem, We have used k-mean clustering algorithm the previously generated sample traffic have been sent to k-mean clustering algorithm then trained on it after that the clustering algorithm showed as a very clear recognition for the statues of traffic as shown in

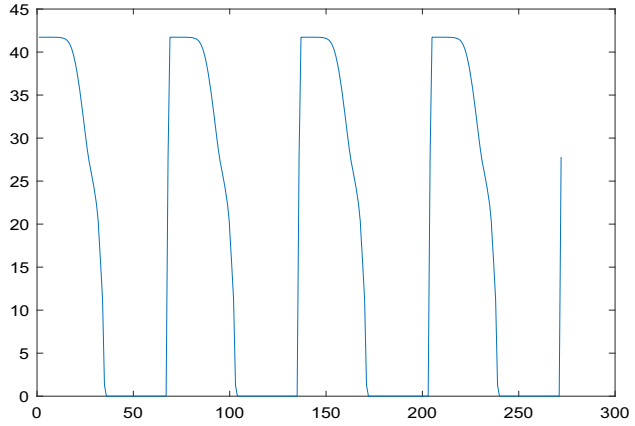

Fig. 11 Sample traffic from the generated big-data signals

this figure, the given pattern is found in the first column of the table and then the predicted pattern have been output by k-mean clustering algorithm and as we see here both of them are identical with no any missing values. Therefore, we calculate the loss we have found it very, close to 0 which means that the accuracy is almost 100% (Loss K-means Testing = 2.2204e-16).

4.2 Prediction final results (for LSTM)

the below Fig. 13 shows the plots of the input training and target training, input testing and targeted testing for the prediction process (LSTM). It also shows the output prediction process which is almost exactly as the target testing this also can be improved by showing accuracy, which is almost near to 99.5%.

the below Figs. 14 show the prediction finally this classified, or cluster recognized pattern should be sent to the prediction algorithm in order to prediction the Future traffic. For this we have used long short-term memory, which is best reignition technique and we have identified parameters using train and error as shown is this table, however, the training validation and testing process as shown in the figure, after the completion of the training validation and testing process is seen in this figure and as shown, the root mean square error is almost 0.71%.

While in the below Fig. 15, GRU is used to do the same prediction task, using the same inputs. This phase is made for a comparison purpose. Where it is clearly noticed that the GRU is slightly faster than the LSTM. But with higher error rate (1%) Table 3.

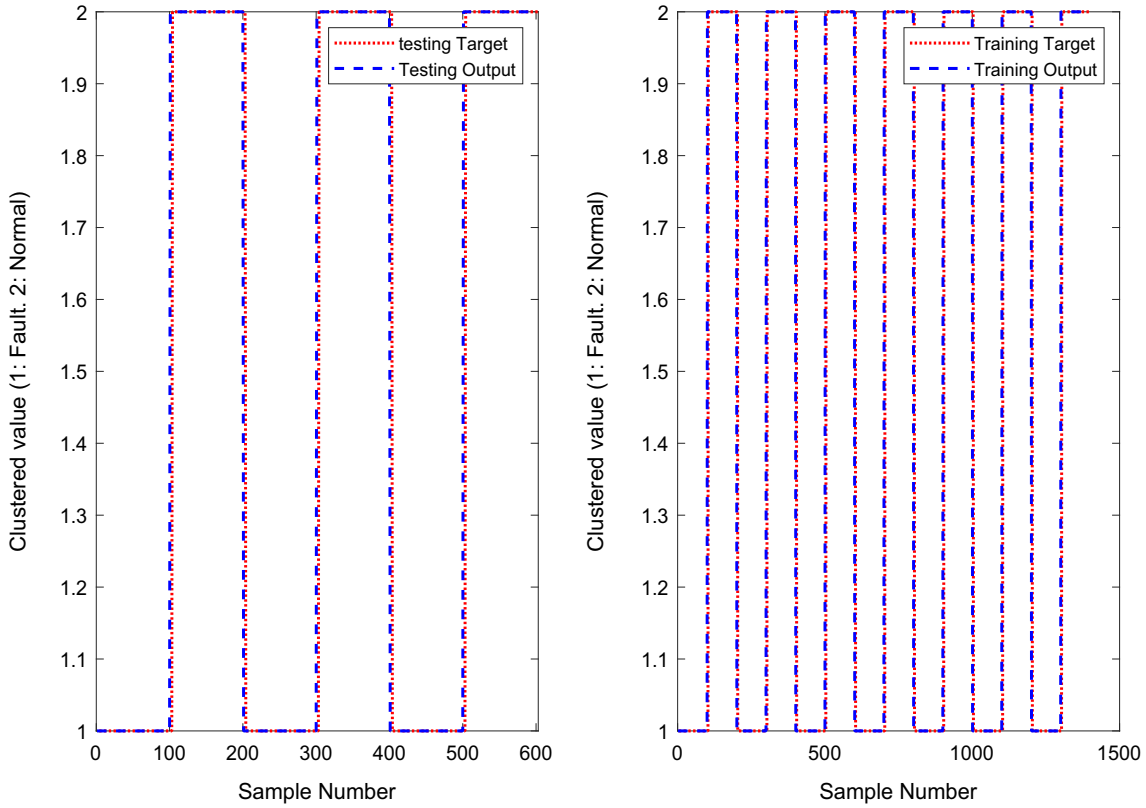

Fig. 12 Pattern Clustering with K-Mean

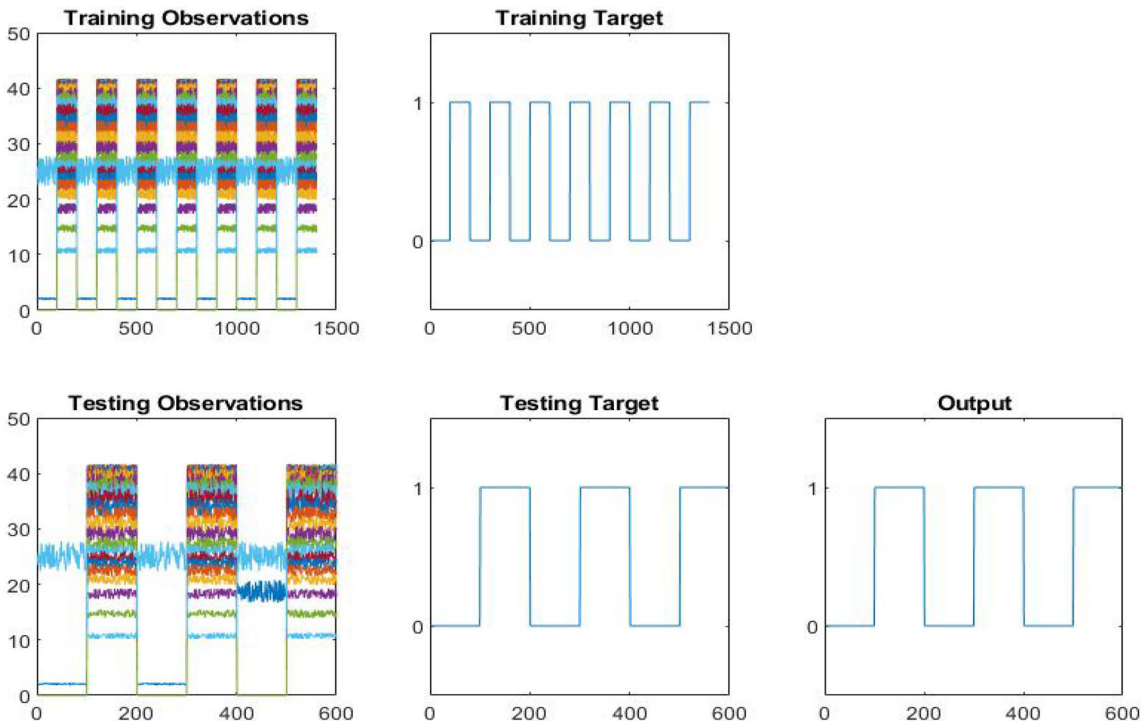

Fig. 13 The input training and target training input testing and targeted testing for the prediction process (LSTM)

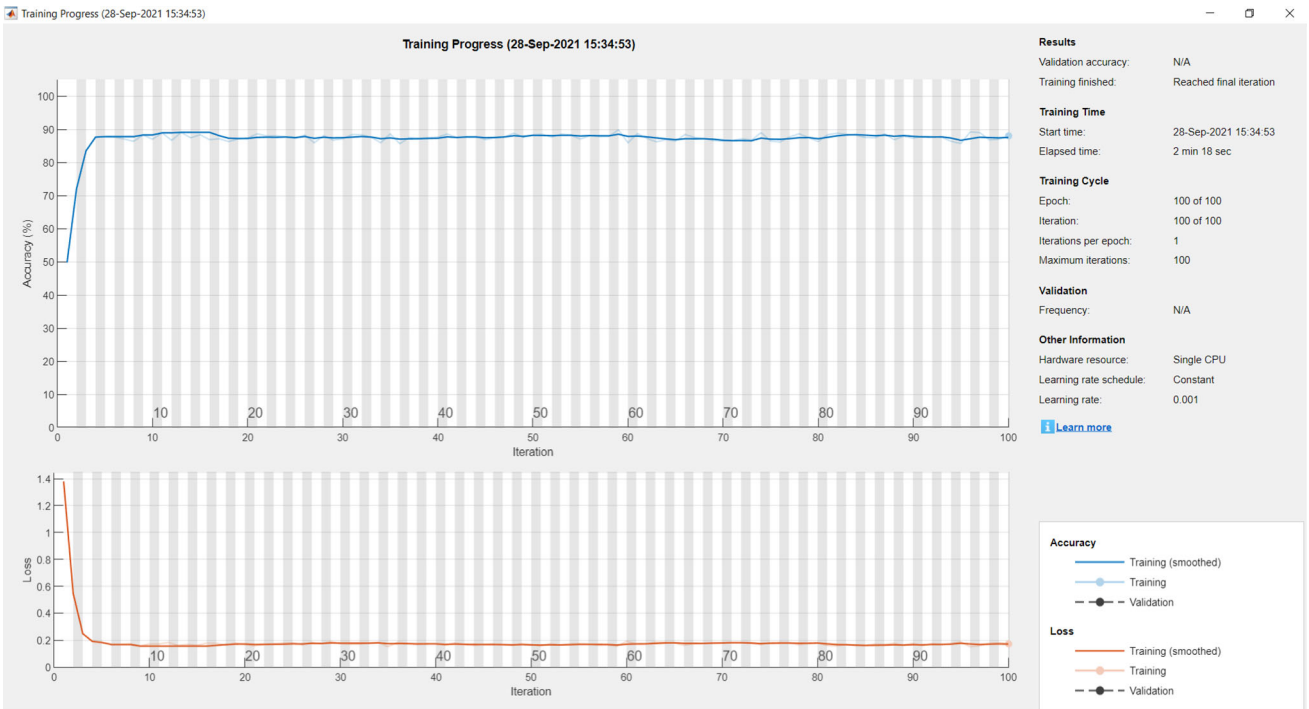

Fig. 14 Prediction (LSTM model and parameters)

4.3 Identify Qs value

This section illustrates how to identify the Q_S value. However, in (Castillejo-Cuberos and Escobar 0.2020) it was determined that the overall specific static and dynamic characteristics of solar irradiance for a determined time period, or state, are summarized by the clearness index (K_t), the diffuse fraction (K) and the variability of the solar

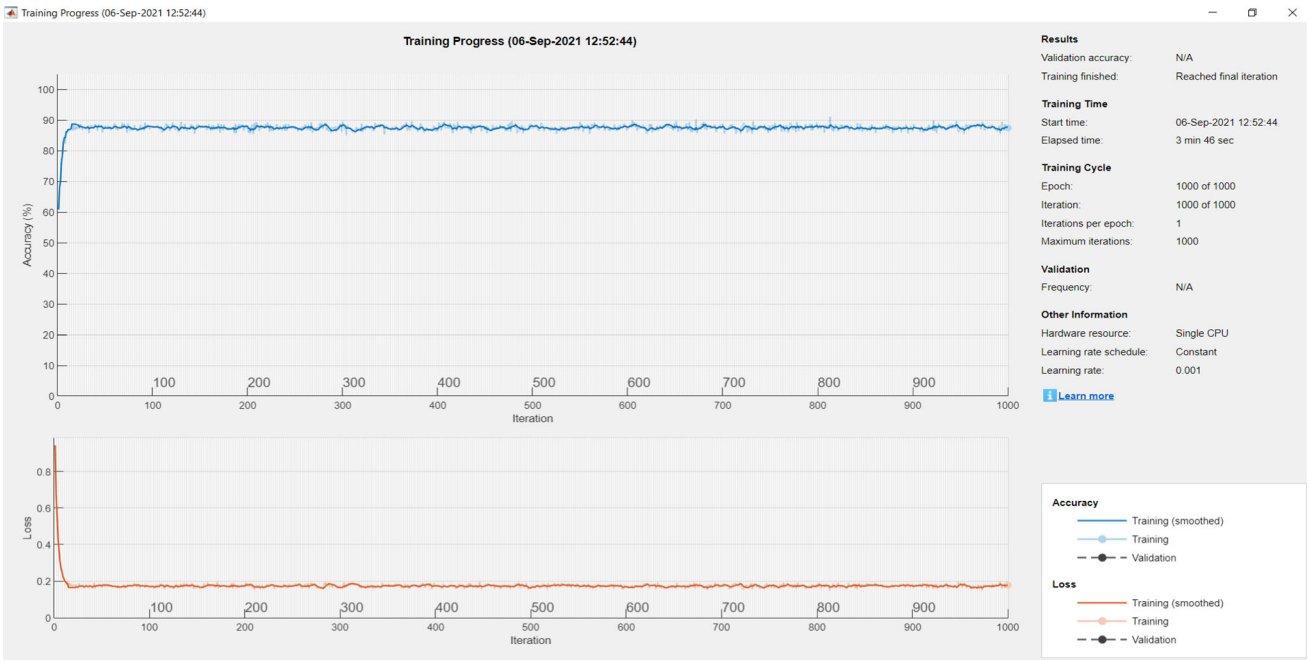

Fig. 15 Prediction (GRU model and parameters)

Table 3 LSTM and GRU final comparative results

Algorithm	LSTM	GRU
Elapsed time (s)	0.167865	0.104726
accuracy	0.9950	0.9900
RMSE	0.0071	0.0100
MAPE	0.0033	0.0075
MAE	0.0050	0.0100
MBE	0.0050	0.0100

resource. By defining a variable termed the solar resource quality score (Q_S), irradiance patterns can be classified in an explicit scheme comparable to clustering techniques but providing a straightforward approach and a simple analytical expression on which to perform the classification, easily translatable across studies, overcoming the issue of data specificity previously discussed.

The only variables needed are the global horizontal irradiance (GHI). The Q_S is a measure of how close a particular state is from the ideal of full atmospheric clearness (K_t = 1), no attenuation (K = 0) and constant irradiance (variability = 0).

Lave et al. (2015) was used, defined as the probability that a particular ramp rate $\left(\mathrm{RR} = \mu\left(\frac{GHI}{G\varepsilon\sin(\alpha)}\right) \geq 0.03\right)$ exceeds a given threshold ramp ($RR_0 = 1000W/m^2$ during an evaluation period, and$(\varepsilon = 0.02)$.

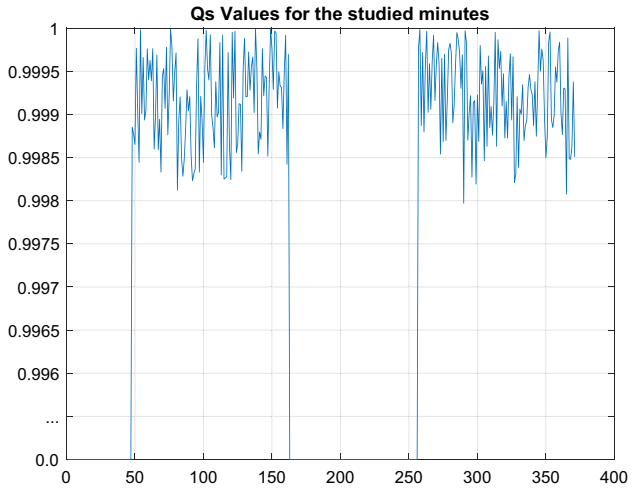

Fig. 16 Qs values using GRU

The Variability Score (CDF based) (VS_{cdf}) uses the normalized, which is the result of linearly mapping the VS_{cdf} from the range 0.002–0.4054 to 0–1 (Castillejo-Cuberos and Escobar 2020), (Δt) temporal resolution timestep

$$VS_{cdf} = min\left(\sqrt{\left(\frac{RR}{RR_0}\right)^2 + \left(P(|RR_{\Delta t}| > RR_0) - 1\right)^2}\right) \tag{72}$$

where,

$(\Delta t) = 5mints.$

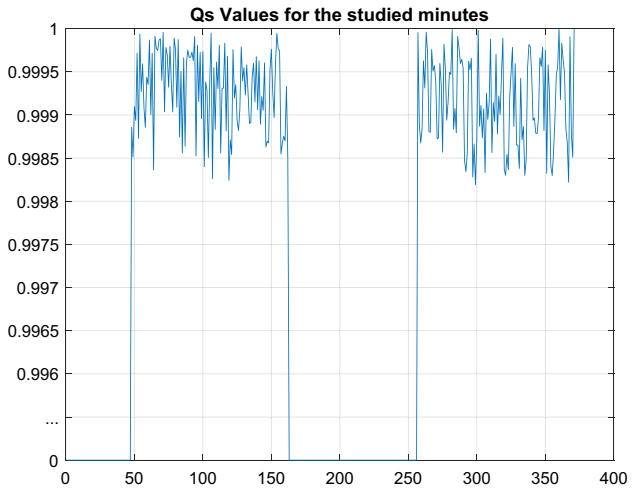

Fig. 17 Qs values using LSTM

Table 4 Maximum efficiency for GRU and LSTM algorithms:

LSTM maximum efficiency	0.4160
GRU maximum efficiency	0.4160

Table 5 Exact optimal values for the future predicted of the P, Qs, and efficiency for using LSTM, GRU algorithms

Optimization variables	Future predicted values of the P, QS, and the efficiency for LSTM	Future predicted values of the P, QS, and the efficiency for GRU
S	− 40.3791	-31.6842
β	52.5599	64.1509
γ	548.9758	689.3164
h_r	0.11198	0.060177
h_c	3.4146	4.5122
l	158.5407	156.2825
G_T	72.5209	367.3079
V_{WS}	34.9638	34.2958
φ	0.78	0.78
θ	2773.0819	2752.9003
∫	12.8308	12.5744
ʎ	176.2378	30.3052
£	0.2	0.2
Ƃ	16.9878	21.0931
α	28.4468	23.3764
μ	29.7625	26.4615
ξ	25.619	29.7124
ϑ	25.5178	24.1549
ϖ	23.7174	20
Ɣ	25.6212	27.932
Ⱦ	29.7412	20.9187

The quality score (Q_S):

$$Q_S = 1 - \left(\sqrt{\frac{(1-K_t)^2 + K^2 + (VS_{cdf})^2}{\sqrt{3}}} \right) \tag{73}$$

Van Haaren et al. (2014), the Daily Average Ramp Rate (DARR) is calculated as the sum of the absolute value of the changes in irradiance on a minute basis, normalized with an irradiance (G) of 1000 W/m2.

$$\text{DARR} = \sum_{k=1}^{k=371} \frac{|GHI_k - GHI_{k-1}|}{G} \tag{74}$$

where,

$$GHI = 0 < GHI < \left(100 + 1.5 * G * (sin\alpha)^{1.2}\right)$$

After implementation the Qs value are determine as shown in the below Figs. 16 and 17:

4.4 Identify efficiency Value:

The maximum power point is given by:

$$P_{max} = V_{max} I_{max} \tag{75}$$

The maximum efficiency of a solar cell is expressed as:

$$\eta = \frac{P_{max}}{P_{in}} = \frac{I_{max} V_{max}}{A G_a} \tag{76}$$

where,

A: Area of the PV module,G_a: Radiation.

After implementation maximum efficiency for the GRU and LSTM algorithms value are determine as shown in the below Table 4:

After implementing the optimization algorithm, we have obtained the optimal below values for using LSTM and GRU algorithms. The below tables Illustrate the exact optimal values for the future predicted values of the P, QS, and the efficiency Table 5.

5 Conclusions and future work

This work addresses the problem of future value estimation based on data with stochastic behaviour. The dependent variables used are the future active generated power, Qs, and the efficiency for the PV grid. While the independent variables are represented with the weather conditions and load behaviour. In this work we proposed a solution based on K-means clustering that converts the stochastic behaviour of the independent variables into a known pattern. Then it becomes easier for either the LSTM or the GRU to use it for the short-term future estimation of the dependant variable. A mathematical model in addition to dataset are

used to build a complete environment, with its operation and signals. And mote- Carlo method is used for simulation. As outcomes of a comparative study, we can confirm that although GRU is a fast algorithm, LSTM provides more accurate short-term predictions. Moreover, it's been concluded that stochasticity is an important factor to be considered in any future estimation process. However, the method proposed in this work shows clear outcomes in solving the stochasticity issue and enhancing the performance of the prediction. Uncertainty in this dynamic environment is an important factor that could be considered in future perspectives.

Acknowledgements The authors thank the anonymous reviewers forgiving their valuable comments and helping us to improve the quality of the paper.

Funding The funding is provided by Anglia Ruskin University.

Data Availability Dataset employed by this research can be retrieved from UK Smart Grid Industry 2021–2024. reportlinker.com/report-summary/Electric-Power/88256/UK-Smart-Grid-Industry.html.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Agarap AFM (2018) A neural network architecture combining gated recurrent unit (GRU) and support vector machine (SVM) for intrusion detection in network traffic data. In Proceedings of the 2018 10th International Conference on Machine Learning and Computing (pp 26–30).

Al-Housani M, Bicer Y, Koç M (2019) Assessment of various dry photovoltaic cleaning techniques and frequencies on the power output of CdTe-type modules in dusty environments. Sustainability 11(10):2850

Ali AS and Azad S (2013) Demand forecasting in smart grid. In Smart Grids (pp 135–150). Springer, London

Bessa RJ (2014) Solar power forecasting for smart grids considering ICT constraints. In Proceedings of the 4th Solar Integration Workshop. Berlin-Germany

Bui V, Kim J and Jang YM (2020) Power demand forecasting using long short-term memory neural network based smart grid. In 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 388–391). IEEE

Cao E (2010) Heat transfer in process engineering. McGraw-Hill Education

Castillejo-Cuberos A, Escobar R (2020) Understanding solar resource variability: An in-depth analysis, using Chile as a case of study. Renewable Sustain Energy Rev 120:109664

Chemetova S, Santos P and Ventim-Neves M (2017) Load forecasting as a computational tool to support smart grids. In 2017 12th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1–6). IEEE

Durisch W, Bitnar B, Mayor J, Kiess H, Lam K, Close J (2007) Efficiency model for photovoltaic modules and demonstration of its application to energy yield estimation. J Solar Energy Mater Solar Cells 79:91

Hasan M, Toma RN, Nahid AA, Islam M, Kim JM (2019) Electricity theft detection in smart grid systems: A CNN-LSTM based approach. Energies 12(17):3310

Huang X, Shi J, Gao B, Tai Y, Chen Z, Zhang J (2019) Forecasting hourly solar irradiance using hybrid wavelet transformation and Elman model in smart grid. IEEE Access 7:139909–139923

Jakkula V, Cook DJ (2007) Mining sensor data in smart environment for temporal activity prediction. Poster session at the ACM SIGKDD, San Jose, CA

Lamnatou C, Chemisana D, Mateus R, Almeida MGD, Silva SM (2015) Review and perspectives on Life Cycle Analysis of solar technologies with emphasis on building-integrated solar thermal systems. Renewable Energy 75:833–846

Lauer M, Jaddivada R and Ilic M (2019) Household energy prediction: methods and applications for smarter grid design. In 2019 8th Mediterranean Conference on Embedded Computing (MECO) (pp. 1–4). IEEE

Lave M, Reno MJ, Broderick RJ (2015) Characterizing local high-frequency solar variability and its impact to distribution studies. Sol Energy 118:327–337

Lee EK, Shi W, Gadh R, Kim W (2016) Design and implementation of a microgrid energy management system. Sustainability 8(11):1143

Li J, Deng D, Zhao J, Cai D, Hu W, Zhang M, Huang Q (2020) A novel hybrid short-term load forecasting method of smart grid using mlr and lstm neural network. IEEE Trans Industr Inf 17(4):2443–2452

Liu BY, Jordan RC (1963) The long-term average performance of flat-plate solar-energy collectors: with design data for the US, its outlying possessions and Canada. Sol Energy 7(2):53–74

Luo H, Wang M, Wong PKY, Tang J, Cheng JC (2021) Construction machine pose prediction considering historical motions and activity attributes using gated recurrent unit (GRU). Autom Construct 121:103444

Mills B, Schleich J (2012) Residential energy-efficient technology adoption, energy conservation, knowledge, and attitudes: An analysis of European countries. Energy Policy 49:616–628

Parvez I, Sarwat A, Debnath A, Olowu T, Dastgir MG, and Riggs H (2020) Multi-layer perceptron based photovoltaic forecasting for

rooftop PV applications in smart grid. In 2020 SoutheastCon (pp. 1–6). IEEE

Patil SV, Babu PV (2012) Experimental studies on mixed convection heat transfer in laminar flow through a plain square duct. Heat Mass Transf 48(12):2013–2021

Petrican T, Vesa AV, Antal M, Pop C, Cioara T, Anghel I, and Salomie I (2018) Evaluating forecasting techniques for integrating household energy prosumers into smart grids. In 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 79–85). IEEE

Saloux E, Teyssedou A, Sorin M (2011) Explicit model of photovoltaic panels to determine voltages and currents at the maximum power point. Sol Energy 85(5):713–722

Sarvi M, Ahmadi S, Abdi S (2015) A PSO-based maximum power point tracking for photovoltaic systems under environmental and partially shaded conditions. Prog Photovoltaics Res Appl 23(2):201–214

Silvestre S, da Silva MA, Chouder A, Guasch D, Karatepe E (2014) New procedure for fault detection in grid connected PV systems based on the evaluation of current and voltage indicators. Energy Convers Manage 86:241–249

Sun Y, Wang F, Wang B, Chen Q, Engerer NA, Mi Z (2017) Correlation feature selection and mutual information theory based quantitative research on meteorological impact factors of module temperature for solar photovoltaic systems. Energies 10(1):7

Tealab A (2018) Time series forecasting using artificial neural networks methodologies: A systematic review. Future Comput Informatics J 3(2):334–340

Tornai K, Kovács L, Oláh A, Drenyovszki R, Pintér I, Tisza D, Levendovszky J (2016) Classification for consumption data in a smart grid based on the forecasting time series. Electric Power Syst Res 141:191–201

Van Haaren R, Morjaria M, Fthenakis V (2014) Empirical assessment of short-term variability from utility-scale solar PV plants. Prog Photovoltaics Res Appl 22(5):548–559

Villalva MG, Gazoli JR and Ruppert Filho E (2009) Modeling and circuit-based simulation of photovoltaic arrays. In 2009 Brazilian Power Electronics Conference (pp. 1244–1254). IEEE

Xu H, Yao S, Li Q and Ye Z (2020). An improved k-means clustering algorithm. In 2020 IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS) (pp. 1–5). IEEE.

Yahyaoui I, Segatto ME (2017) A practical technique for on-line monitoring of a photovoltaic plant connected to a single-phase grid. Energy Convers Manage 132:198–206

Yu W, An D, Griffith D, Yang Q, Xu G (2015) Towards statistical modeling and machine learning-based energy usage forecasting in the smart grid. ACM SIGAPP Appl Comput Rev 15(1):6–16

Yu W, An D, Griffith D, Yang Q, and Xu G (2014) On statistical modeling and forecasting of energy usage in smart grid. In Proceedings of the 2014 Conference on Research in Adaptive and Convergent Systems (pp. 12–17)