DEVELOPMENT OF DEEP LEARNING PREDICTIVE MODELS FOR SHORT-TERM SOLAR RADIATION FORECASTING: CASE STUDY IN VIETNAM

A Thesis submitted by

Anh Ngoc Lan Huynh

BSc

For the award of

Master of Science Research (MSCR)

2020

ABSTRACT

Vietnam is a developing country with a projected high economic growth. The energy sector plays an essential role in its socio-economic development, with an average of 5% increase in annual energy demands. As a result of its growing energy consumption, the nation is hugely reliant on imported fossil fuels. In the foreseeable future, oil-based fuels are expected to become somewhat limited and even too exhausted. Coal resources, in general, are not easy to exploit or utilise due to notable climate change, economic and technical limitations. This situation increases pressure on national energy security in Vietnam. Considering the high solar radiation footprint in Vietnam, future assessments of the availability of renewable energy resources, through research and development initiatives and modelling studies, such as the one performed in this research study, can effectively provide meaningful information to Vietnamese policymakers in seeking alternative energy resources to replace the finite oil and coal reserves. Freely available renewable resources, such as solar energy, can be ideal sustainable solutions to satisfy the needs of energy security into the future. Research into solar energy forecasting has great potential in the search for alternative energy resources.

Since the availability of solar radiation at the earth's surface is directly proportional to the harvested solar power at any specified location, the forecasting of global solar radiation is essential for continuous monitoring and supply management of solar energy through solar generation systems. Therefore, the forecasting of solar radiation has been studied extensively in literature, using many forecasting methods, but these methods have been largely based on physical and statistically driven models. In spite of their usage, these methods come with certain limitations, such as the underlying assumptions of initial conditions, mainly with physics-based models. The concept of persistence to forecast future solar radiation, as used in previous methods, can also be a challenging factor to attain good accuracy. In contrast, physical models may not adequately address issues of stochasticity *(i.e.* rapid change in solar radiation or its predictor variable, *e.g.* cloud cover) datasets, or the issues of data non-stationarity. As an alternative forecasting method, machine learning approaches, which capture historical behaviour of solar radiation or its predictor variables to model utilising artificial intelligence algorithms are now becoming prominent. Such a model is constructed and trained to learn the patterns in historical solar radiation (and related) datasets, to build a non-linear mapping scheme between the antecedent (*i.e.* lagged) inputs and the target (*i.e.* solar radiation) dataset. This thesis focuses on developing and evaluating latest machine learning models for global solar radiation forecasting in the context of Vietnam, where the potential utility of machine learning models has not yet been fully explored.

The primary aim of this Master of Science (MSCR) thesis is to develop new and scientifically verified models to resolve the challenges in solar radiation forecasting, addressing the issues of complexity in predictor and the target datasets while also capturing the non-linear behaviour of solar radiation. To pursue this aim, the study is based on Long Short-Term Memory (LSTM) network algorithm built for global solar radiation forecasting. In some other

studies, LSTM has attained superior capability in learning the long- and short-term dependencies in historical datasets, by employing memory cells that determine the importance and distinguish between the important and unimportant data features through their input, forget and output network gates. Consequently, the essential data features are then used to build the feature engineering and data pattern learning process in the resulting LSTM model. This research, therefore, utilises the LSTM model to forecast the hourly solar radiation and also attempt to capture the dependence between consecutive hours on the same day, as well as the long-term (*e.g*. seasonal) behaviours to be learned efficiently.

This MSCR thesis, presented as a synthesis of two journal publications, has adopted an LSTM model incorporating a data pre-processing technique based on the Robust Local Mean Decomposition (RLMD). The research study area has focussed on solar energy belts, which are the critical zones in Vietnam for major solar energy projects. The aims of this research thesis are as follows: (1) To develop a near real-time solar radiation forecasting model using the LSTM algorithm applied to multiple time-step forecast horizons (*this work has been reported in Journal Paper 1*); (2) To further improve the method in the first aim and build a hybrid forecasting model utilising a data pre-processing technique based on the RLMD method, and further evaluate the forecasting performance of the resulting hybrid RLMD-LSTM model using half-hourly global solar radiation forecasting (*this work has been reported in Journal Paper 2 – under review*). The overall results of this study show the LSTM model can be used as a useful utility in global solar radiation forecasting at near real-time horizons. The findings of this study can have important implications for renewable energy feasibility studies, and also help in several areas where data-driven decisions may require some of the best practice forecasting techniques.

In synopsis, the predictive models developed in this MSCR thesis will provide significant benefits to solar energy generators, authorities for energy operation and distribution, through new and improved solar radiation forecasting tools. Energy forecasters can therefore adopt these novel methods, to address the issues of non-linearity and the non-stationarity in energy usage, by constructing real-time forecasting tailored for energy industries and other stakeholders.

CERTIFICATION OF THESIS

This thesis is entirely the work of **ANH NGOC LAN HUYNH** except where otherwise acknowledged, with the majority of the authorship of the papers presented as a Thesis by Publication being undertaken by the student. The work is original and has not previously been submitted for any other award, except where acknowledged.

Professor Ravinesh C. Deo

Principal Supervisor

Dr Shahab Abdulla

Associate Supervisor

Dr Nawin Raj

Associate Supervisor

Dr Mumtaz Ali

External Supervisor

Student and supervisors' signatures of endorsement are held at USQ.

STATEMENT OF CONTRIBUTION

The following provides details of the agreed percentage share of contributions of the candidate and the co-authors in the publications presented in this research thesis.

Article I: Chapter 4

Huynh, A. .-L.; Deo, R.C.; An-Vo, D.-A.; Ali, M.; Raj, N.; Abdulla, S. Near Real-Time Global Solar Radiation Forecasting at Multiple Time-Step Horizons Using the Long Short-Term Memory Network. *Energies* 2020, 13, 3517. [Impact Factor: 2.822, SNIP: 1.154, Scopus Ranked: Q1].

DOI: https://doi.org/10.3390/en13143517.

Author	Percentage Contributions	Task Performed
Anh Ngoc-Lan Huynh (MSCR Candidate)	65%	Method development, programming, data analysis, preparation of tables and figures, compilation, writing, and revision of the manuscript.
Ravinesh C. Deo Principal Supervisor	15%	Supervised and assisted in model concepts, provided detailed comments on the manuscript, edited and prepared for the submission.
Mumtaz Ali External Supervisor	5%	Provided statistical support including implementation of statistical concepts, analyses and interpretation of results.
Shahab Abdulla Associate Supervisor	5%	Supervised and provided comments on the manuscript.
Nawin Raj Associate Supervisor	5%	Supervised and provided comments on the manuscript.
Duc-Anh An-Vo	5%	Comments on the final manuscript.

Article II: Chapter 5

Huynh, A. .-L.; Deo, R.C.; Ali, M.; Raj, N.; Abdulla, S. A novel short-term solar predictive hybrid model based on Long Short-Term Memory integrated with Robust Local Mean Decomposition. *Applied Energy, In Review.* [Impact Factor: 8.848, SNIP: 2.865, Scopus Ranked: Q1].

Author	Percentage Contributions	Task Performed
Anh Ngoc-Lan Huynh (MSCR Candidate)	70%	Method development, programming, data analysis, preparation of tables and figures, compilation, writing, and revision of the manuscript.
Ravinesh C. Deo Principal Supervisor	15%	Supervised and assisted in model concepts, provided detailed comments on the manuscript, edited and prepared for the submission.
Mumtaz Ali External Supervisor	5%	Provided statistical support including implementation of statistical concepts, analyses and interpretation of results.
Shahab Abdulla Associate Supervisor	5%	Supervised and provided comments on the manuscript.
Nawin Raj Associate Supervisor	5%	Supervised and provided comments on the manuscript.

ACKNOWLEDGEMENTS

Please accept my great appreciation for these individuals and organisations for their whole-hearted assistance, scientific guidance, valuable and useful advice, and for generously sharing their expertise. Undoubtedly, without their support, motivation and encouragement, it would not have been possible for me to complete this study. In particular, I would like to express my gratitude and sincere appreciation to:

- **Professor Ravinesh C. Deo**, my Principal Supervisor, for his devotion and motivation in guiding me to become an excellent student. His quick response and straightforward approach helped me to carry out my study with a proper schedule. Several sentences cannot express my appreciation of his contribution to my achievement.
- **Dr Shahab Abdulla**, my associate supervisor, for his scientific advice and his useful comments in the journal articles and in my final thesis.
- **Dr Nawin Raj,** my associate supervisor, for his scientific advice and his useful comments in the journal articles and in my final thesis.
- **Dr Mumtaz Ali,** my external supervisor, to whom I am grateful for sharing his valuable suggestions, and plentiful comments at different stages of my research. He played an important role in building up my research foundation and contributing additional research funding.
- **Dr An-Vo Duc-Anh**, to whom I am grateful for his comments and advice in the journal articles.
- **Dr Barbara Harmes,** the key representative of the English Angels program within the USQ College (English Language), for her special support in proofreading and improving my writing skills.
- **Dr Douglas Eacersall**, an enthusiastic consultant from the HDR Learning advice consultation, for his special support in proofreading.
- **Advanced Data Analytics:** Environmental Modelling and Simulation Group, led by the Principal Supervisor, for contributions and sharing ideas.
- **My family and friends,** who have always supported and encouraged me in difficult times and hard situations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ARIMA	Autoregressive Integrated Moving Average	LSTM	Long Short-Term Memory
DL	Deep Learning	*MAE*	Mean Absolute Error
DNN	Deep Neuron Network	*MAPE*	Mean Absolute Percentage Error
GSR	Global Solar Radiation	*MSE*	Mean Squared Error
GSR_{ACTUAL}	Actual Global Solar Radiation	*MLP*	Multilayer Perceptron Network
GSR_{N}	Normalised Global Solar Radiation	*PACF*	Partial Auto-Correlation Function
GSR_{MIN}	Minimum value of Global Solar Radiation	$\lvert FE \rvert$	Absolute Forecasted Error
GSR_{OBS}	Observed Global Solar Radiation	*r*	Pearson's Correlation Coefficient
$< \mathrm{GSR}_{OBS} >$	Average value of Observed Global Solar Radiation	R^2	Coefficient of determination
GSR_{FOR}	Forecasted Global Solar Radiation	*RMSE*	Root Mean Square Error
$< \mathrm{GSR}_{FOR} >$	Average value of Forecasted Global Solar Radiation	*RRMSE*	Relative Root Mean Square Error
E_{NS}	Nash-Sutcliffe Efficiency	RNN	Recurrent Neural Networks
LM	Legate & McCabe's Index	*N*	Number of values in a data series
CNN	Convolutional Neural Network	SVR	Support vector regression
NWP	Numerical Weather Prediction	*ELM*	Extreme Learning Machine

CHAPTER 1: INTRODUCTION

1.1 Foreword

This chapter first introduces the background of the research topic, which is a basis of the statement of the problems. This chapter also highlights the two main objectives and briefly describes the organisation of this research thesis.

1.2 Background

Vietnam is a developing country with an annual average gross domestic product of around 7%. The energy sector has played a vital role in this growth. Both the primary and final energy demand has increased over 5% each year, with the primary fuels accounting for such growth being coal and oil, while the share of renewables has been negligible. Due to this growing energy consumption, Vietnam has relied on energy imported, including coal and oil, from foreign countries, which puts its national energy security at severe risk.

Extensive progress in science and technology has improved the comfort of human life. However, this energy generates environmental risks and energy crises, such as increasing environmental pollutants and reducing energy resources, which are considered threats to human life. Hence, the exploitation of renewable energy sources has generated increasing interest in the pursuit of environmental protection worldwide (Tan et al., 2012). In 2017, cumulative solar photovoltaic power capacity reached almost 398 GW and generated over 460 TWh, representing around 2% of global power. This growth is projected to continue at a similar rate in the future. Following the global trends in energy exploration (Luong, 2015; IEA, 2012; Shem et al., 2019) and recommendations of the United Nations Sustainable Development Goal that advocates the need for cleaner, affordable and accessible energy in all nations and regions, Vietnam has recently commenced capacity development for solar energy resources. With its geographical location close to the solar energy belt, Vietnam can harvest this energy from freely available sunlight, theoretically providing 60–100 GWh year-1 of concentrated solar power and 0.8–1.2 GWh year-1 photovoltaic power (Polo et al., 2015). These figures indicate that consumer power shares drive the continuous growth of solar energy.

Solar energy represents one of the promising options, particularly in photovoltaic power generation (Farivar and Asaei, 2011). However, solar radiation can be heavily influenced by rapid changes in natural conditions; hence, it is stochastic (Perez et al., 2014). This variability has a substantial impact on power grid security in connection with large-scale photovoltaic

power generation. Therefore, developing an accurate short-term solar radiation forecasting model is crucial to ensure the optimum dispatch and management of power systems, particularly with photovoltaic power generation (Mostafavi et al., 2013).

The commonly available computational models used to forecast short-term solar radiation include linear regression models, satellite data-based models, and neural network-based models (*e.g.* artificial neural network (ANN)) using meteorological parameters (*e.g.* air temperature, related humidity). Of these, there are two main classes: physical models and statistical models. Physical models are based on mathematical equations that describe the physical state and dynamic motions of the atmosphere. These models are complex non-linear equations that need strong computing power to solve them. Numerical methods are used to obtain the approximate solutions of these equations, and these models are also known as numerical weather prediction (NWP) models. The NWP models are used in forecasting but are not always available. The errors of solar radiation forecasts based on NWP vary significantly and depend on the atmosphere's different climate and dynamic motion at the study location (Wittmann et al., 2008; Zamora et al., 2005; Lorenz et al., 2009). The statistical models based on satellite data and sky images detect the motion of cloud structures using motion vector fields. Cloud structure motion is determined from two consecutive cloud index images from satellite data or sky camera images. The literature on this issue has proposed the errors of satellite data and sky images-based forecasts (Nova et al., 2005). However, those parameters are not always available due to high measuring costs and climate conditions, so it is necessary to develop a new model that can generate meaningful predictions in these circumstances.

With increasing computational techniques, machine learning-based forecasting models have been successfully proven to provide better performance over the physically-based models (Alzahrani et al., 2017; Paulescu and Blaga, 2016; Coulibaly and Ouedraogo, 2016). Machine learning models have also outperformed many of the time series methods (Benmouiza and Cheknane, 2016; Sfetsos and Coonick, 2000; Hocaoğlu, 2011; Rigler et al., 2004; Bracale et al., 2013; Ming and Ningzhou, 2011; Wenbin et al., 2002) that have utilised historical patterns to predict solar radiation (Ruiz-Arias et al., 2010; Martín et al., 2010; Moreno-Munoz et al., 2008; Colak et al., 2015; Rao et al., 2012; Kaplani and Kaplanis, 2012; Lauret et al., 2013; Ramedani et al., 2014). However, common weaknesses have been reported, such as those causing bias when extending data volume (Al-Musaylh et al., 2018a; De Leone et al., 2015) and over-fitting (Voyant et al., 2017). The latest advancements in the deep learning techniques of machine learning can effectively and accurately solve the above issues. However, they have

not been fully explored yet. Notably, the application of these techniques has not yet been performed in the context of Vietnam.

A better understanding of solar radiation's historical patterns, particularly the highly chaotic and intermittent properties of short-term signals, can help to provide an accurate forecast of solar energy. Capturing the patterns of variables by using novel deep learning-based models, this study implements this forecast of solar energy at different levels with sufficient time ahead. This knowledge is required to evaluate the stability and regulation of photovoltaic power, including reverse power management and dispatching to a grid schedule and unit commitment provided by energy industries. The developed deep learning model can provide better support for energy security and operation while extending this implementation to other fields.

This thesis aims to develop new machine learning models based on Long Short-Term Memory (LSTM) network algorithms, which have a good ability to model time-sequential datasets. Current literature shows that these models have not been used for solar radiation short-term forecasting in the specific solar belt locations, in particular, Bac-Ninh (21.2013°N, 106.0629°E); Da-Nang (16.0544° N, 108.2022° E); Central Highlands (24.0330° S, 148.7374° E); Song-Binh (10.3752° N, 106.4111° E) and Tri-An (11.0843° N, 106.9772° E) in Vietnam. Moreover, LSTM-based models have been used in multiple forecasting horizons, especially from one-minute to six-hourly scales in the present study region. This research thesis addresses these gaps, and so makes a vital scientific contribution towards solar radiation forecasting in Vietnam, where projected growth in solar energy projects is expected over the next few decades.

1.3 Statement of the problem

The research into solar radiation forecasting should be taken for several reasons, such as making proper operations and security, growing the capacity of the existing system and scheduling distributions (Zhang et al., 2017). In addition, solar radiation is highly variable because it is driven by synoptic and local weather patterns. This high variability presents challenges to meeting power production and the demand curve. Solar radiation forecasting is vital for energy industries, allowing for the set up of contingency mechanisms to mitigate any deviation from the required production. The underestimation and overestimation of solar radiation forecasting cause either very high or adverse prices on the intraday market, and intraday forecast errors determine the need for costly balancing power (Zhang et al., 2017).

Consequently, a robust predictive model with minimal inaccuracies is necessary for the context of the national solar energy market.

Vietnam is now following the global trend of developing renewable energy with the aim of reducing fossil fuel usage and environmental issues. With a geographic location close to the equator, Vietnam takes advantage of a high volume of solar radiation; hence solar energy is one of the significant energy sectors. The national estimated energy consumption is relatively high and will account for 60% of the national demand in the next five years. Currently, there are still several errors in forecasts for solar energy from the old forecast technique. An accurate solar radiation forecasting model is essential to minimize the errors, which can lead to a better performance of the energy market. Nevertheless, at present, none is available in Vietnam.

Solar radiation can be predicted from historical time-series data with several independent variables (*e.g.* temperature, rainfall). Also, the non-stationarity features of solar radiation can influence the precision and accuracy of forecasting models. To address these issues, a robust forecasting model from data-driven models is used, which can significantly deal with complex temporal behaviour of solar radiation.

Data-driven models can manage uncertainties as a common issue in forecasting problems where physical (*i.e.* mathematically based) models can be less effective. A review of previous and recent studies showed that various data-driven models had been successfully adopted, in which machine learning is the most common technique. Such machine learning includes the Multivariate Adaptive Regression Splines (MARS) (Srivastava et al., 2019) and Support Vector Regression (SVR) (Ghimire et al., 2019c; Deo et al., 2016). Moreover, recently deep learning techniques are more popular due to their advanced benefits for different data forecasting, such as a Recurrent Neural Network (RNN) (Yona et al., 2013), Long Short-Term Memory (LSTM) (Ghimire et al., 2019b) and Multilayer Perceptron (MLP) (Azimi et al., 2016).

Data decomposition methods, such as a robust version of local mean decomposition, are also required to address the non-stationarity features of those input variables (Liu et al., 2017b). Furthermore, optimisation and selecting the best parameters are other challenges for solar radiation forecasting that can be achieved by grid search and trial-and-error techniques. Therefore, this study investigates the ability of the methods above for solar radiation forecasting.

An integration of forecasting approaches and decomposition techniques, which can be used to improve the model's forecast accuracy, has been rare in short-term solar radiation forecasting. Motivated by the reasons mentioned above, this study developed a hybrid LSTM model by integrating LSTM with RLMD, which can analyse the erratic and chaotic solar radiation better. This thesis assesses the effectiveness of the developed model over multiple horizons and locations. However, it was essential first to identify relevant research problems. While these established questions are extensively represented in the following sections, the research gaps have also been identified at the same time, based on published literature.

1.4 Research aims and objectives

The main objectives of this master thesis are to develop and evaluate the utility of machine learning algorithms in real-life solar radiation forecasting problems. This thesis also aims to test these models for their usage in five specific solar-rich locations in Vietnam by presenting a set of two high-quality publications. Therefore, this research study will examine the viability and explore the most optimal machine learning model for solar radiation forecasting from a minute to half-hourly horizons.

The primary purpose of this thesis is to develop Long Short-Term Memory (LSTM) based models for analysing solar radiation in Vietnam. Subsequently, the LSTM is applied along with Robust Local Mean Decomposition (RLMD) to analyse solar radiation's stochastic patterns.

This thesis adopts a deep learning technique within a forecasting model framework to achieve the following specific objectives:

i. To develop a near real-time solar radiation prediction model using LSTM for multiple time horizons (*i.e.* 1-, 5-, 10-, 15-minute) in the context of Vietnam and to compare their accuracies with other machine learning models by using visual analysis and performance metric measures. *This objective is achieved and the results are published in Paper 1* (Chapter 4).

ii. To evaluate the LSTM model's forecasting performance incorporated with a pre-processing technique – RLMD – in multiple locations in Vietnam in a longer short-term horizon (*i.e.* half-hourly). *This objective is achieved and will be published in Paper 2* (Chapter 5).

These objectives have been fulfilled through two journal papers produced as an outcome of the MSCR Thesis by Publication.

1.5 Organisation of the Thesis

This thesis comprises two parts focusing on two objectives, presented as an MSCR Thesis by Publication, which covers the two objectives, and with a conclusion that summarises the challenges, findings, significance and scientific contributions of this study and recommendations for future work.

Chapter 1 presents the introduction with the background, research gaps and research questions and also presents the objectives of this study.

Chapter 2 overviews the literature based on LSTM and RLMD and their applications in terms of solar radiation forecasting. This chapter firstly explains the requirement of short-term solar radiation forecasting. Secondly, it provides the literature review of Deep Learning and LSTM, in particular in terms of solar radiation forecasting. Then, a general review of RLMD regarding its application is presented. Finally, this section summarises a statement of existing gaps from the above literature review.

Chapter 3 describes the data sources and methodology. This chapter provides general viewpoints while the specific study area, data and methods are presented in respective chapters. It also presents the model development process and the equations of the different methods used in this research.

Chapter 4 presents a published journal article in the journal Energies (DOI: https://doi.org/10.3390/en13143517). It presents the research results of very short-term *(i.e.* multi-horizon) solar radiation forecasting, which covers this research's first objective. It also explores the application of data-driven models, including LSTM, ARIMA, SVR, MLP, DNN.

Chapter 5 presents as a journal article submitted for a review in the journal Applied Energy. It describes half-hourly solar radiation forecasting results. It also shows an improvement of forecast accuracy and reduction in forecast error using the RLMD method.

Chapter 6 covers conclusions and limitations, and recommendations for future research in solar radiation forecasting.

CHAPTER 2: LITERATURE REVIEW

2.1 Foreword

This chapter provides a literature review of methodology for forecast horizon in terms of solar radiation forecasting. This chapter also provides a brief statement of existing gaps, which will be addressed in this study.

2.2 The importance of understanding the properties of solar radiation and the short-term forecasting model

The increase in fossil fuel prices and climate change concern has spurred the demand for renewable energy resources, which have many advantages, including being environmentally friendly and sustainable. Solar energy is one of the renewable energies, which uses solar photovoltaic systems to convert sunlight into electricity. These sources are highly intermittent and have chaotic properties. Typically, the output of solar photovoltaic is highly dependent on solar radiation, temperature and different weather parameters. A generative power generally depends on a numerical weather prediction, which is called a physical model. The quality of the power forecasting algorithm heavily depends on numerical weather prediction, which means it decreases with increasing time horizons. Forecasting algorithm horizons range from short-term forecasts in hourly ranges, mid-term forecasts up to some days, to long-term forecasts in the range of some weeks. Different time horizons are of interest for different market participants that help improve the various application of power systems. In particular, there are several models such as a long-term forecast, a mid-term forecast (*e.g.* a day-ahead forecast) (Yang et al., 2014; Pierro et al., 2015), and a short-term forecast (*e.g.* an hourly-ahead forecast) (Qing and Niu, 2018; Mellit and Pavan, 2010).

Based on the interest of users from scheduling to management, different time horizons – namely, very short-term forecasting, short-term forecasting, medium-term forecasting and long-term forecasting – are considered. In detail, very short-term forecasting concentrates on a few seconds up to a 30-minutes interval, which is useful in immediate actions. In contrast, short-term forecasting, ranging from 30-minute intervals to 6 hours ahead, is suitable for load dispatch planning and operational security. For example, medium forecasting, ranging from a day ahead forecast, is helpful while achieving significant operational management and cost optimization; long-term forecasting, from one day to a longer time horizon, is considered.

Solar radiation plays a vital role in the success of solar energy operation, which is shown in various studies. The solar radiation-based forecasting models provide the best accuracy compared to other methods (*e.g.* cloud-based, temperature-based or other meteorological parameter-based models). However, solar radiation is not always available for various specific usage purposes (*e.g.* long-term, medium-term, short-term) due to high measuring cost and weather condition. It is even more challenging when analysing solar radiation in the short-term since minute solar radiation is rare. While the use of hourly radiation to evaluate the overall energy delivery of a solar system usually provides accurate results, hourly data cannot predict the short-term behaviour. Although it has been known for over 30 years that hourly radiation data are not representative of actual instantaneous or minute radiation, models based on hourly radiation measurements have continued to be used due to lack of available detailed radiation data (Mccormick and Suehrcke, 2018). Therefore, there is an urgent need to develop a proper short-term forecasting model that can reduce previously noted limitations. Moreover, the intermittent and non-controllable characteristics of solar production bring several other problems, such as voltage fluctuations, low power quality and stability issues (Moreno-Munoz et al., 2008); (Anderson and Leach, 2004). In short, the new modelis also necessary for: estimating the reserves; scheduling the power system; congestion management; optimal management of the storage with the stochastic production; trading the produced power in the electricity market; and, finally, to achieve a reduction in the costs of electricity production (Espinar et al., 2010); (Moreno-Munoz et al., 2008).

2.3 Limitations of common forecasting methods

Regarding solar radiation forecasting, there is no one-fits-all approach. Particularly, the forecast horizon decides the suitability of alternative models, such as deciding on operational management. For example, short-term forecasting models will forecast solar radiation from 5 minutes to a few hours ahead. A very short-term horizon mentioned in the literature is meaningful from an economic perspective, in that a rise in the accuracy of solar energy forecasts may facilitate major cost savings (Martinez-Anido et al., 2016). The main purpose of short-term forecasts is to maintain operation security (Voyant et al., 2017).

Specifically, a review of solar radiation forecasts using machine learning algorithms with a forecast horizon of 30 minutes and below are given. For instance, the review by Raza et al. (2016) is based on a comprehensive study of several machine learning-based forecasting methods. A design of solar radiation forecasting offers several research views, and these complicate cross-study comparisons. Although studies often employ specific spatio-temporal

data in a unique context of weather characteristics, there is no guarantee that a method can be successful in all places and with the specific time horizons. The following table provides a summary of the previous related studies in terms of solar radiation's forecast horizon, corresponding data, and the employed forecasting methods.

Table 2.1. Prior studies on near real-time solar radiation forecasting

Study	Date source	Time resolution	Forecast horizon	Number of data points		Method	
				Training set	Testing set	Proposed method	Benchmark
Sivaneasan et al. (2017)	GSR	5 minutes	5-minute	24,260	8,640	BPNN	Fuzzy Logic-BPNN
Golestaneh et al. (2016)	GSR	10 minutes	10-minute	52,560	52,560	ELM	Persistent, BELM
Sun et al. (2018)	TSI	1 minute	15-minute	68,833	8,075	CNN	N/A
Khelifi et al. (2020)	GSR	1 minute	10-minute	N/A	N/A	MLP	RBF
Paulescu and Paulescu (2019)	TSI	10 minutes	20-minute	38,371	33,644	ARIMA	RW, MA, ES
(Vaz et al., 2016)	GSR	15 minutes	15-minute	21,170	1,798	ANN	ANN, MLP, NARX
Nobre et al. (2016)	GSR	15 minutes	15-minute	35,040	35,040	ARIMA	Persistence
Ryu et al. (2019)	TSI, GSR	1 minute	5,10,15,20 minute	2,016	864	CNN	Persistence
Zhou et al. (2019)	GSR	7.5 minutes	7.5,15, 30,60	201,480	67,160	LSTM	ARIMAX, MLP, Persistence

As shown in Table 2.1, several methods have been applied to different datasets with different spatio-temporal scales and time resolution, ranging from one minute to 7.5 minutes resolution. Furthermore, several forecast horizons have been tested, from a minute timescale to few minutes ahead forecasting. Moreover, Table 2.1 reveals that few studies involve an evaluation of models across several forecast horizons and big data context. Except for (Zhou et al., 2019), the authors devise a model for multiple forecast horizon with training set greater than 100,000 points.

This study adopts deep learning to capitalize its benefits for short-term forecasting; primarily, to overcome limitations of conventional data-driven models (*e.g.* machine learning), since they are unable to capture short- and long-term dependency between a target (*e.g.* future solar radiation) and the corresponding historical variables. Firstly, machine learning cannot be used directly with big data due to the issues of scalability, and it may not have a sufficiently sophisticated architecture to extract complex patterns. A similar conclusion can be found in (Voyant et al., 2017). Secondly, no study considers the efficiency of a forecasting model for more than one forecast horizon due to the shortage of data in the real-time horizon. Finally, no study considers the ability of LSTM in dealing with data efficiency (*e.g.* data volume). Therefore, a restriction of prior studies is that they solely explain solar radiation behaviour and forecasting efficiency in a specific context.

2.4 LSTM-based forecasting techniques

Some sequential models (*e.g.* Markov models, Kalman filters and conditional random fields) are commonly used tools to address the raw sequential input data in previous works. However, the drawback of these traditional sequential models is that they cannot adequately capture long-term temporal dependencies. Many indiscriminative or even noisy signals in the sequential input during an extended period may ignore informative and discriminative signals in applying day-ahead solar radiation, and this can lead to the failure of these above sequence models. Recently, RNN has emerged as one useful model for sequence learning, which has already been successfully applied in various fields, including image captioning, speech recognition, genomic analysis and natural language processing (Längkvist et al., 2014).

For a problem in which time-series are forecasted using historical patterns within a forecasting framework, a Long Short-Term Memory (LSTM) model, the advanced version of RNN, has been embraced as a powerful new tool. The LSTM model (Hochreiter and Schmidhuber, 1997) has shown its ability to handle predictive issues (*e.g.* over-fitting) (Alzahrani et al., 2017) by absorbing long-term dependencies of historical data and is extensively employed in image recognition (Zhu et al., 2017), automatic speech recognition (Sak et al., 2014) and natural language processing (Wang and Jiang, 2015). In terms of solar radiation forecasting, LSTM is expected to explore the temporal and spatial dependence of data while utilising contextual information (Yu et al., 2019; Zhou et al., 2019). Hence, LSTM has been studied in solar forecasting during the last 5 years (Wang et al., 2019; Caballero et al., 2018; Alzahrani et al., 2017; Qing and Niu, 2018; Yu et al., 2019; Mukherjee et al., 2018; Ghimire et al., 2019b; Lee and Lee, 2018; Zhou et al., 2019; Gensler et al., 2016; Wang et al.,

2018a; Muhammad et al., 2019; Siddiqui et al., 2019; Lee et al., 2018). The study by (Qing and Niu, 2018) used LSTM to forecast hourly solar energy on an island of Santiago, Cape Verde. In another study, LSTM was used by (Gensler et al., 2016) to forecast solar radiation base on weather-based data in Germany. In another study (Srivastava and Lessmann, 2018), LSTM was used to present solar energy forecasting in Europe and US.

A developed Principal Component Analysis (PCA) was incorporated with LSTM for solar forecasting in Europe. An LSTM standalone model was developed by (Ghimire et al., 2019b) and (Ghimire et al., 2019a), where a hybrid Convolutional Neural Network (CNN) is incorporated with LSTM for solar forecasting among Australian solar-rich cities. Despite their growing applications in solar forecasting, an LSTM-based model has not been fully explored in terms of near real-time solar forecasting in general, and in the context of Vietnam.

2.5 Robust Local Mean Decomposition (RLMD)

Although standalone models and hybrid models achieve good predictions, they may provide unsatisfactory forecasting results in some circumstances where actual solar radiation is non-stationary (Voyant et al., 2017). Therefore, decomposition-based methods, for example Local Mean Decomposition (LMD) (Wang et al., 2018c), Empirical Mode Decomposition (EMD) (Li et al., 2018), Discrete Wavelet Transform (DWT) (Ren et al., 2015), Ensemble (EMD) (Wang et al., 2015), Empirical Wavelet Decomposition (EWT) (Dash et al., 2020) and Variational Mode Decomposition (VMD) have been successfully exploited in solar radiation forecasting (Wang and Wu, 2016). Nevertheless, they still contain inherent disadvantages (Wang and Wu, 2016), such as the end effect and mode-mixing problems of LMD. A Robust LMD (or RLMD) was developed by Liu et al. (2017b) to address weaknesses of LMD in terms of signal processing. These techniques of RLMD and LMD are applied to adapt time-frequency by demodulating amplitude and frequency modulated signals into a set of sub-series. Therefore, they can help to avoid the negative frequency of the decomposed sub-series, increasing the precision in further processing steps. A study with this idea (Jiang et al., 2019) first applied RLMD in wind forecasting that indicates better performance in short-term wind forecasting. Table 2.2 provides an overview of the common hybrid deep learning methods in terms of short-term solar radiation forecasting. Most models were integrations of pre-processing techniques and deep learning forecasting models; however, they have not been fully explored. In short, from Table 2.2 and to the best of our knowledge, RLMD has not been studied previously in terms of short-term solar radiation forecasting.

Table 2.2. Prior studies of hybrid deep learning method in terms of short-term solar radiation forecasting

Reference	Horizon	Method
Zang et al. (2018)	Short-term	VMD-CNN
Wang et al. (2019)	Short-term	CNN-LSTM
Ghimire et al. (2019b)	Short-term	CNN-LSTM
Khodayar et al. (2019)	Long-term	CGAE
Gensler et al. (2016)	Long-term	Auto encoder LSTM
Wang et al. (2018b)	Long-term	DWT-CNN-LSTM
Siddiqui et al. (2019)	Long-term	CNN-LSTM
Lee et al. (2018)	Long-term	CNN-LSTM

2.6 Statement of gaps in the existing literature

In terms of forecasting technique, the mentioned studies show that machine learning has good potential in very short-term solar energy forecasting. However, a noted limitation of ML algorithms is the insufficient learning capability for high dimensional datasets (Goodfellow et al., 2016), which directly influences the precision and accuracy of forecasting models by over-fitting and extrapolation (Jieni and Zhongke, 2008). By considering that the GSR time series often have the long-term and short-term dependency in the low-frequency approximate parts, the LSTM network, a special type of RNN, is employed to predict the decomposed low-frequency sub-layer. Also, to deal with the non-stationarity of solar radiation, Robust LMD is adopted to decompose the high temporal data, to improve the precision of the forecasting model and reduce forecasting errors.

To address gaps in knowledge that advocate a need to assist in integrating solar energy variability behaviour into real-time systems, the novelty of this research is to design a new deep learning predictive model based on an integration of the Long Short-Term Memory and Robust Local Mean Decomposition algorithms tailored for short-term solar radiation predictions. This

study also aims to emulate the RLMD-LSTM model at multi-step forecast horizons. Following the studies mentioned in previous sections, the RLMD algorithm is incorporated to extract intrinsic features of the solar radiation series, while in the second phase LSTM utilises all relevant features for prediction. This research designs, and then evaluates, the RLMD-LSTM hybrid predictive model with non-linear data pre-processing and mapping capability, integrated with an RLMD, to obtain accurate solar radiation predictions.

CHAPTER 3: DATA AND METHODOLOGY

3.1 Foreword

This chapter provides an overview of the study area and the datasets used in developing the hybridised, data intelligent, solar radiation forecasting models in this MSCR thesis. In Vietnam, there were different chosen study sites depending on the availability of the data, to achieve the objectives of this thesis. The following sections describe these locations in detail and provide a summary for these study areas with relevant data. This chapter also provides a brief methodology, whereas the specific methodology of the model development techniques is presented in detail in other relevant chapters. In summary, this chapter provides an overview of the procedure that was used to develop the hybridised solar radiation forecasting.

3.2 Data locations

The study area includes five different locations from the North to the South in Vietnam, where the solar radiation is potentially more solar rich than other parts of the country with higher solar radiation volume. These locations, namely Bac-Ninh (21.2013°N, 106.0629°E), Da-Nang (16.0544°N, 108.2022°E), Central Highlands (24.0330°S, 148.7374°E), Song-Binh (10.3752°N, 106.4111°E), and Tri-An (11.0843°N, 106.9772°E) are shown in Figure 3.1.

A brief description of these study sites is as follows:

(i) Bac-Ninh has a subtropical dry winter climate characterized by hot and humid summers with frequent tropical downpours of short duration and warm and frequently dry winters. This province is undergoing revitalization with more sustainable future solar energy systems, partly funded by the World Bank. The province also meets the criteria set for selection for the future installation of solar measurement stations, (*i.e.* homogeneous landscape and land-usage without large water bodies). Thus, the development of a solar radiation forecasting model at multiple forecast horizons, especially in Bac-Ninh, is a justified research endeavour to support the United Nation's Sustainable Development Goal #7, related to the access to affordable renewable energies for all populations.

(ii) Da-Nang includes a steep Annamite mountain range that dominates Da-Nang to the North and North-west, and which features peaks ranging from 700 to 1,500 meters (2,300 to 4,900 feet) in height, and low-lying coastal plains with some salting to the South and East, with several white sand beaches along the coast.

(iii) Central Highlands has a lower altitude and higher temperatures than other sub-regions. Influenced by the equatorial climate, the annual average temperature is about 20°C with a moderate climate all year round. The climate in Central Highlands Vietnam has two distinct seasons: a dry season and a rainy season.

(iv) Song-Binh is a solar rich city. On average, the temperatures are high. The rainy season is from May to October. The warmest month is April with an average maximum temperature of 35°C (95°F), and the coldest month is December with an average maximum temperature of 30°C (86°F).

(v) Tri-An is in South-east Vietnam. In Tri-An, the wet season is oppressive and overcast. The dry season is muggy and partly cloudy. Over the year, the temperature typically varies from 30°C to 35°F and is rarely below 24°C or above 40°C.

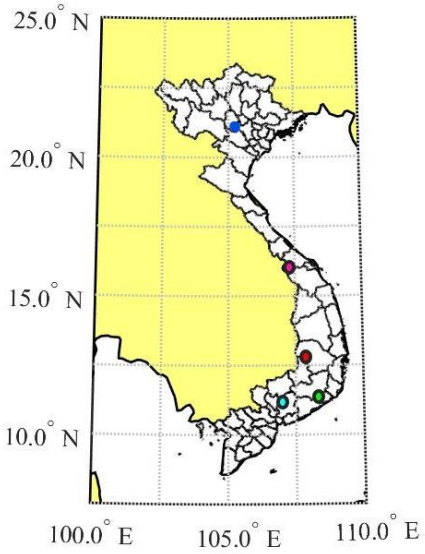

Figure 3.1 Location of study sites in this project. Bac-Ninh is in blue, Da-Nang is in pink, Central Highlands is in red, Song-Binh is in green, Tri-An is in cyan.

3.3 Data sources

The data for model development has been acquired from the World Bank (Peters et al., 2019) for ground-based solar radiation measurement. The purpose of this study was to forecast multiple time horizons (*i.e.* 1-minute, 5-minute, 10-minute, 15-minute and 30-minute) (very short-term) and half-hourly (short-term) solar radiation forecasting in five specific regions in Vietnam. However, the primary data were not available at every timescale, since it is rare to have specific time horizon data (*e.g.* 5-minute or half-hourly). Therefore, the original solar

radiation data had been converted into appropriate timescale data (*e.g.* from 1-minute to 5-minute). Notably, global solar radiation measurements were performed simultaneously, 24 h a day, at equidistant time intervals of 1 minute. Only the data from 06:00 to 18:00 were used for designing the predictive model as these times represent a period of meaningful daylight hours. The details of these data for each objective are presented in Table 3.1.

For Objective 1, the aggregated data from Bac-Ninh were used and extracted from the World Bank. Besides, five forecast horizons of 1-minute, 5-minute, 10-minute, 15-minute and 30-minute were presented utilising data from the period Sep-2017 to June-2019. Chapter 4 provides more details of these data.

For Objective 2, the aggregated data from four different locations (*i.e*. Da-Nang, Central Highlands, Song-Binh, Tri-An) were extracted from the same source as Objective 1. Chapter 5 provides more details of these data.

Table 3.1 Details of data used in this study

Objective	**Data type**	**Source**	**Forecast horizon**	**Study period (dd-mm-yyyy)**
Objective 1 (Paper 1)	Global solar radiation	The World Bank	1-minute (1M)	1-6-2019 to 30-6-2019
			5-minute (5M)	1-1-2019 to 30-6-2019
			10-minute (10M)	27-9-2017 to 30-6-2019
			15-minute (15M)	27-9-2017 to 30-6-2019
			30-minute (30M)	27-9-2017 to 30-6-2019
Objective 2 (Paper 2)			30-minute (30M)	27-9-2017 to 30-6-2019

3.4 General Methodology

Various necessary tasks were applied to the data before the model development step. Converting the original data including the inputs and relative targets into their required forecast horizons, using partial correlation function (PACF) to select best statistically significant lags from the target variable were the first two steps in all objectives of this thesis. Secondly, the data in Objective 2 were decomposed using a robust version of Local Mean Decomposition (RLMD) to address the non-stationarity issues associated with the data. Also, to avoid large numeric ranges from the higher frequency data from inputs and target datasets of all objectives, this study normalised these data between zero and one using Eq. (1,2). Finally, the best parameters and boundary conditions of the models developed in the respective objectives of

this study were selected using optimisation and trial-and-error methods. Respective chapters describe specifically these processes.

$$GSR_N = \frac{GSR_{ACTUAL} - GSR_{MIN}}{GSR_{MAX} - GSR_{MIN}} \quad (1)$$

$$GSR_{ACTUAL} = GSR_N (GSR_{MAX} - GSR_{MIN}) + GSR_{MIN} \quad (2)$$

This thesis considers various forecasting models to evaluate and compare their abilities of solar radiation forecasting over different horizons. The models are: deep recurrent neural network (DNN); multivariate adaptive regression splines (MARS); support vector regression (SVR); autoregressive integrated moving average (ARIMA); multilayer perceptron (MLP); and, Long Short-Term Memory (LSTM). The pre-processing approach used to handle the non-stationarity features in this study is robust local mean decomposition (RLMD). A trial-and-error method was also utilised in this study to select the appropriate parameters.

In the literature review, the LSTM algorithm was introduced by Hochreiter and Schmidhuber (1997) to address the vanishing gradient issue. The ARIMA model was popularised by the work of Box and Jenkins(Box, 1994) Support vector regression (SVR) is a regression version of the SVM model that is common in solar energy forecasting (Fentis et al., 2017; Alfadda et al., 2017). The deep neural network is a machine learning method that has been advanced based on artificial neural networks (ANN), and is capable of trained complex input and learning procedures (Liu et al., 2017a). Similar applications of DNN in solar energy forecasting can be found in (Díaz–Vico et al., 2017; Alzahrani et al., 2017). The multilayer perceptron network (MLP) is the most common type of feed-forward network (Werbos, 1974). MLP has three layers: an input layer, an output layer and a hidden layer. In a range of forecasting applications, the LSTM model is considered a fast and precise data intelligent approach that can offer better performance compared to other algorithms (Ghimire et al., 2019a).

Objective 1 provides forecasting model developments and evaluations for multiple time horizons (*i.e.* 1-minute, 5-minute, 10-minute, 15-minute and 30-minute). In this objective (Chapter 4), LSTM was used and developed against SVR, ARIMA, DNN, MLP. In Objective 2 (Chapter 5), the hybrid RLMD-LSTM was developed and compared with RLMD-MARS, RLMD-SVR for short-term horizon (*i.e.* 30-minute). Each related chapter presents the details

of the theoretical backgrounds and methods. To sum up, the novel and hybrid models developed in this study were:

1- The LSTM model for multiple forecast horizon (*i.e.* 1-minute, 5-minute, 10-minute, 15-minute and 30-minute) presented in Chapter 4. The study applied the PACF technique to select the significant lag (input)s variables from the target variables.

2- The two phases RLMD-LSTM for half-hourly forecasting horizons is presented in Chapter 5. The RLMD was adopted to decompose the target data into product functions (PFs) and a residual component.

   In this thesis, several methods are adopted to evaluate the model's performance (Steyerberg et al., 2010) including correlation coefficient (r), Root Mean Square Error (*RMSE*), Mean Absolute Error (*MAE*), Willmott (WI) and Nash-Sutcliffe (E_{NS}) Legate & McCabe's Index (*LM*). The details and mathematical equations for these statistical indices are shown in each chapter of this thesis. Additionally, this thesis also provides visual analysis by box plots, scatter plots, time series plots and Taylor plots.

CHAPTER 4: MULTIPLE VERY SHORT-TERM HORIZON FORECASTING

Article I: Near Real-Time Global Solar Radiation Forecasting at Multiple Time-Step Horizons Using the Long Short-Term Memory Network

4.1 Foreword

This chapter proposes the Long Short-Term Memory (LSTM) network modelling strategy based on deep learning algorithms for the context of the very short-term global solar radiation (GSR) forecasting. This study uses five different time-step solar radiation data at a solar-rich study site (Bac-Ninh, Vietnam) from 2017 to 2019.

Before developing the LSTM model, the partial autocorrelation function was applied to the high-resolution 1-minute scaled solar radiation dataset that generates statistically significant lagged predictor variables. Those lagged variables describe the historical behaviour of GSR. After that, the LSTM algorithm is applied to capture the short- and long-term dependencies within the GSR data series patterns to predict future GSR at multiple forecasting horizons. The benchmarked models are: other deep learning, a statistical model, a single hidden layer and a machine learning-based model. As a result, the LSTM model provides a satisfactory result with a correlation coefficient exceeding 0.90 and also outperforming the benchmark models.

In accordance with robust statistical metrics (*e.g.* root-mean-square error, mean absolute error, Willmott's Index, Nash-Sutcliffe coefficient and the Legates & McCabe's Index) and visual analysis of model performance in the testing period, this study ascertains the practicality of the proposed LSTM deep learning approach to generate reliable GSR forecasts for all five timescales. While the predictive performance is seen to reduce as the forecast horizon is increased (from a minute to 30-minute interval) the relative root-mean-square error remained profoundly lower for the LSTM (12–44%) compared with the ANN (23–50%) and ARIMA models (54–137%) in the testing phase. The Legates & McCabe's Index, yielding a value of approximately 0.8194–0.9545 for LSTM, compared with 0.8051–0.9255 and 0.4065–0.8005, respectively, for the ANN and ARIMA models, also confirms the superiority of the LSTM approach for GSR forecasting at multiple forecast horizons. Accordingly, the present study confirms the utility of LSTM for its potential application in solar energy monitoring, including its usage in renewable energy resource evaluations and modern management systems tailored for continuous monitoring of energy variables (*e.g.* ocean waves, hydropower and wind). Further exploration of LSTM, based on its excellent performance, is also encouraged in other

areas such as energy sustainability, or demand and supply modelling, energy pricing and policy development areas.

4.3 Published article I

energies

MDPI

Article

Near Real-Time Global Solar Radiation Forecasting at Multiple Time-Step Horizons Using the Long Short-Term Memory Network

Anh Ngoc-Lan Huynh [1,*], Ravinesh C. Deo [1,*], Duc-Anh An-Vo [2], Mumtaz Ali [3], Nawin Raj [1] and Shahab Abdulla [4]

[1] School of Sciences, Institute of Life Sciences and the Environment, University of Southern Queensland, Queensland 4350, Australia; nawin.raj@usq.edu.au
[2] Centre for Applied Climate Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia; Duc-Anh.An-Vo@usq.edu.au
[3] Deakin-SWU Joint Research Centre on Big Data, School of Information Technology, Deakin University, Burwood VIC 2134, Australia; mumtaz.ali@deakin.edu.au
[4] Open Access College, University of Southern Queensland, Queensland 4500, Australia; Shahab.Abdulla@usq.edu.au
* Correspondence: anh.huynh@usq.edu.au (A.N.-L.H.); ravinesh.deo@usq.edu.au (R.C.D.)

Received: 15 May 2020; Accepted: 26 June 2020; Published: 8 July 2020

Abstract: This paper aims to develop the long short-term memory (LSTM) network modelling strategy based on deep learning principles, tailored for the very short-term, near-real-time global solar radiation (GSR) forecasting. To build the prescribed LSTM model, the partial autocorrelation function is applied to the high resolution, 1 min scaled solar radiation dataset that generates statistically significant lagged predictor variables describing the antecedent behaviour of GSR. The LSTM algorithm is adopted to capture the short- and the long-term dependencies within the GSR data series patterns to accurately predict the future GSR at 1, 5, 10, 15, and 30 min forecasting horizons. This objective model is benchmarked at a solar energy resource rich study site (Bac-Ninh, Vietnam) against the competing counterpart methods employing other deep learning, a statistical model, a single hidden layer and a machine learning-based model. The LSTM model generates satisfactory predictions at multiple-time step horizons, achieving a correlation coefficient exceeding 0.90, outperforming all of the counterparts. In accordance with robust statistical metrics and visual analysis of all tested data, the study ascertains the practicality of the proposed LSTM approach to generate reliable GSR forecasts. The Diebold–Mariano statistic test also shows LSTM outperforms the counterparts in most cases. The study confirms the practical utility of LSTM in renewable energy studies, and broadly in energy-monitoring devices tailored for other energy variables (e.g., hydro and wind energy).

Keywords: solar radiation; long short-term memory network; near real-time solar radiation forecasting

1. Introduction

Conventional energies (e.g., fossil fuel) have been a primary energy resource for many decades [1–3]; however, these resources are being replaced gradually by various renewable resources as a pivotal solution that aims to meet the future energy crisis caused by their depleting nature and the environmental damage caused by greenhouse gas emissions through the burning of carbon-positive fuel [4,5]. Following the global trends in energy exploration [6–8] and the recommendations of the

United Nations Sustainable Development Goal that advocates a dire need for cleaner, affordable and accessible energy in all nations and regions. Thus, Vietnam has recently commenced capacity development for solar energy resources. With its geographical location close to the solar energy belt, Vietnam can harvest this energy from freely available sunlight, theoretically, providing 60–100 GWh·year^{-1} of solar concentrated power and 0.8–1.2 GWh·year^{-1} as photovoltaic power [9]. These figures advocate a continuous growth of solar energy which will meet the increasing consumer power demand. As such, it is important to develop modern technologies for energy management systems that purposely support real-time energy integration in a power grid or a distribution system [10]. An accurate near-real-time forecasting tool, especially tailored for solar energy management and proportional dispatching to and from a grid system is therefore a scientific contrivance for the update of solar energy into a real grid system [11].

Solar energy forecasting is typically based on consumer usage to provide greater stability and energy regulation, reverse management and dispatching, scheduling and unit commitments [11]. For each of the consumer's usages, the forecasting timescales can vary from being a long-term forecast (e.g., a monthly forecast) [12] to a mid-term forecast (e.g., a day-ahead forecast) [13,14] and to a short-term forecast (e.g., an hour-ahead) period [15,16]. However, studies on a very short-term or near-real-time forecast are relatively scarce, thus, the present research work aims to fulfil this need.

There are many approaches in solar radiation forecasting, divided roughly into data-driven (or artificial intelligence) and physical (atmospheric dynamic) models [17]. Many existing studies, however, reveal limitations in forecasting techniques (e.g., computational resources to calibrate a huge volume of data, thus encountering unexpected errors) and challenges arising from complexity of predictor variables (e.g., intermittent and chaotic properties of consumer demands, meteorological and geographical data) [11,18]. To overcome these issues, the present research work is focused on developing a new modelling strategy for near-real-time solar radiation forecasting by implementing the latest deep learning techniques.

The construction of a solar radiation-forecasting model in general and the global solar radiation (GSR) model, in particular, have been intensively explored. With the recent advances of computational data science, machine learning-based forecasting models typically provide distinct advantages over physical models [17,19,20] and time-series models [21–35]. Models based on machine learning and neural networks have evolved over recent decades. However, common weaknesses have been reported, such as those causing bias when extending data volume, and over-fitting [11]. Deep learning techniques, the latest advancement of machine learning, can solve the above issues but have not been fully explored. On the other hand, solar forecasting is relatively new in Vietnam, although there were previous studies of solar radiation [36–38] and solar potential mapping [39] in recent decades. Studies implementing machine learning methods for solar forecasting can be found in other Asian countries [40–43]. However, the application of these techniques has not been performed in the context of Vietnam, although a recent study with a similar approach was undertaken in Australia [44]. Nonetheless, to the best of the author's knowledge, the present study is the first exploring the predictive power of a deep learning method—the long short-term memory (LSTM) network model for minute-ahead solar energy forecasting, particularly in the context of Vietnam where the prospect for solar powered energy systems is relatively high.

This study adopts long short-term memory (LSTM) networks, a branch of deep neural networks, which has shown an excellent ability to handle predictive issues and has been extensively employed in image recognition, automatic speech recognition and natural language processing [45–52]. LSTM is believed to overcome limitations of conventional data-driven models in capturing short- and long-term dependency between a target (e.g., future solar radiation) and corresponding historical variables and big data issues. In addition, due to its ability of removing abundant information to resolve vanishing gradient issues, LSTM is appropriate to represent the learning data over different temporal domains [17]. Hence, LSTM has been studied in solar forecasting during the past five years [15,17,44,53–63]. For instance, the first study regarding LSTM [64] demonstrated its forecasting skills for one-day ahead utilizing remote-sensing data under various topographical conditions with the best root mean square error (RMSE) ~24% and mean absolute error (MAE) ~17%. In [15,65], LSTM

performed well for one-day-ahead forecasting under multiple seasonal and weather conditions. Generally, although forecasting methods show their predictive skills in a different context, optimizing the forecasting methods has still been an important problem of interest. Similarly, developing an optimal LSTM model in terms of solar energy forecasting is also under consideration. Most recent papers have focused on data pre-processing techniques to optimize predictive results [59,60,66,67]; hence, there is a lack of intensive papers relating to optimizing the LSTM technique itself. Moreover, a huge dataset volume was discussed in the context of traditional time-series methods which pointed out that performance efficiency decreased against the increase of data volume. Thus, the present work aims to extensively explore optimization and performance assessment of the LSTM forecasting technique in the near-real-time case by also considering multiple performance metrics (i.e., relative prediction errors, including Willmott's index, the Nash–Sutcliff index, and the Legates and McCabe index) adopted for multiple forecast horizons ranging from 1 min to ½ hour periods.

In terms of the model performance evaluation, despite higher levels of model assessment skill in the error measurement approaches compared with the correlation coefficient (r) which represents the relationship between observed and predicted values [68], it is not totally sensible when applying RMSE and MAE alone [44,69], especially in deep learning method evaluation. Therefore, it is reasonable to apply multiple metrics in model performance evaluation to avoid their specific weaknesses [70]. For this reason, applying multiple evaluation metrics to assess the predictive performance of the LSTM method in near-real-time forecasts is a novelty in this paper.

Moreover, due to geographical location and weather condition, in some circumstances (e.g., Vietnam), it is difficult and costly to obtain such meteorological variables in a near-real-time horizon (i.e., minute interval). To address this issue, the present work will employ historical global solar radiation (GSR) time-series, data which is hardly seen in the literature.

Since the model's accuracy is expected to decrease over the passage of time, the timescale of the forecasts encompasses the next minute of the GSR data in advance, to verify the persistence skill of the LSTM model. Therefore, to address the gaps in knowledge and also to advocate the need for a sustainable real-time energy management, the novelty of this paper is to firstly develop a near-real-time solar forecasting based on the integration of the LSTM algorithm. The paper also aims to emulate the LSTM model at multiple forecast horizons (i.e., 1, 5, 10, 15, 30 min) to ensure it is validated over a much longer period.

To perform this, a time series of the GSR data measured at the minute interval at a selected location (Bac-Ninh, Vietnam) is obtained. To demonstrate the advantages of the LSTM model in terms of near-real-time solar forecasting, this paper also compares LSTM performances against those of the traditional forecasting method, autoregressive integrated moving average (ARIMA), and the well-known machine learning methods of multilayer perceptron (MLP) network, support vector regression (SVR), and a deep learning method, deep neural network (DNN), in GSR near-real-time solar forecasting. As a representative of traditional forecast modelling, ARIMA and SVR are chosen for the modelling due to the non-stationary properties of the collected data [71]. Meanwhile, the MLP and DNN models are representative of neural network algorithms, which have been widely employed in recent decades [72,73]. To explore the predictive skill of the proposed method, the minute interval data is evaluated at multiple time horizons: 1 min (1M), 5 min (5M), 10 min (10M), 15 min (15M) and 30 min (30M) forecast.

The main contributions of this study are as follows.

1. Development and optimization of a near-real-time GSR forecasting method by implementing the LSTM algorithm for 1 minute using lagged combinations of the aggregated GSR data as the predictor variables.
2. Evaluation of the performance of the proposed model against benchmarked models (DNN, MLP, ARIMA, SVR) by a range of model evaluation metrics.
3. Implementation of the proposed models for multi-minute ahead (e.g., 5M, 10M, 15M, 30M) and evaluation of the performance of LSTM over multiple forecast horizons.

To reach these objectives, this paper is organized as follows: Section 2 reviews previous literature. Section 3 presents a theoretical overview of the objective models. In Section 4, the dataset considered is introduced and explained, detailing model tuning and benchmark algorithms. Section 5 presents model performance metrics. A discussion of empirical results is available in Section 6 before the paper concludes in Section 7.

2. Related Work

In terms of solar irradiance forecasting, there is no one-fits-all modelling approach; in particular, the forecast horizon determines the suitability of alternative models (e.g., to support decision-making in operational management). Previous research has studied short-term models which forecast solar irradiance from 5 min to a few hours ahead. The focus of this paper is minute ahead forecasting (e.g., 1 min, 5 min, 10 min, 15 min and 30 min). A minute horizon is established in the literature and meaningful from an economic perspective since a rise in accuracy of solar energy forecasts may facilitate major cost savings [74]. The main purpose of minute-ahead forecasts is to maintain operational security [11].

In the following, previous studies with comparable forecast horizons from the past decade are discussed. Specifically, a review of solar irradiance forecasts using machine learning algorithms with forecast horizon of 30 min and below are given. The review [75–77] is based on a comprehensive study of several solar energy forecasting methods. A design of solar irradiance forecasts offers several research views and these complicate cross-study comparisons. Although studies often employ a specific spatio-temporal data in a unique context of weather characteristics, there is no guarantee that a method can be successful in all places and with all different time horizons. To depict a review of minute solar energy forecasting, the following table classifies and summarizes the previous related studies in terms of forecast horizon, corresponding data, and the employed forecasting methods.

As shown in Table 1, several methods have been applied to different data sets with different spatio-temporal scales and time resolution, ranging from 1 to 7.5 min resolution. In addition, several forecast horizons have been tested, from a minute timescale to few-minutes-ahead forecasting. Moreover, Table 1 reveals few studies involving an evaluation of models across several forecast horizons and in a big data context. Except for [58], the authors devise a model for multiple forecast horizon with training sets greater than 100,000 points.

Table 1. Prior studies of near-real-time solar irradiance forecasting.

Study	Data Source	Time Resolution	Forecast Horizon	The Number of Data Points		Method	
				Training Set	Testing Set	Proposed Method	Benchmark
[78]	GSR	5 min	5 min	24,260	8640	BPNN	Fuzzy Logic-BPNN
[79]	GSR	10 min	10 min	52,560	52,560	ELM	Persistent, BELM
[80]	TSI	1 min	15 min	68,833	8075	CNN	N/A
[81]	GSR	1 min	10 min	N/A	N/A	MLP	RBF
[82]	TSI	10 min	20 min	38,371	33,644	ARIMA	RW, MA, ES
[83]	GSR	15 min	15 min	21,170	1798	ANN	ANN, MLP, NARX
[84]	GSR	15 min	15 min	35,040	35,040	ARIMA	Persistence
[46]	TSI, GSR	1 min	5, 10, 15, 20 min	2016	864	CNN	Persistence
[58]	GSR	7.5 min	7.5, 15, 30, 60	201, 480	67, 160	LSTM	ARIMAX, MLP, Persistence

In terms of forecasting technique, these studies show that machine learning (ML) has good potential in very short-term solar energy forecasting. However, a limitation of ML algorithms is the insufficient learning models for high dimensional datasets [85], which directly influences the precision and accuracy of forecasting model by over-fitting and extrapolation [86]. By considering that the GSR time series often have long-term and short-term dependency in the low-frequency approximate parts, the long short-term memory (LSTM) network, a special type of recurrent neural network (RNN), is employed to predict the decomposed low-frequency sub-layer in this study.

Multiple conclusions emerge from Table 1. Firstly, the potential of LSTM has not been fully explored, especially in analysing high dimensional data at 1 min intervals. The similar conclusion can be found in [11]. Secondly, no study considers the efficiency of forecasting models for more than one forecast horizon. This might be due to the shortage of data in real time horizons. Finally, no study considers the ability of LSTM in dealing with data efficiency (e.g., data volume). Therefore, a restriction of prior studies is that they solely explain the GSR behaviour and forecasting efficiency in a specific context.

In this paper, the aim is to overcome these issues. Firstly, the forecasting ability across multiple forecast horizons through employing GSR accelerated from 1 min interval data for each horizon is demonstrated. This technique can resolve the shortage of available datasets, which facilitates a broader demonstration of the forecasting model efficiency across multiple forecast horizons. In addition, the bias toward data efficiency through employing different partition proportions in training and testing sets is addressed. An appropriate data proportion in forecasting models is thus carefully chosen. Finally, a high number of data points (over 100,000) is used for validation to test model efficiency in terms of big data.

As the objective technique in this study, LSTM is employed to prove its potential in real-time solar radiation forecasting, as well as dealing with high dimensional real-time GSR. As shown in Table 1, this approach has not yet been fully explored and previous studies on LSTM-based solar energy forecasting faced a risk of over-tuning [87]. Overcoming the limitation of the available dataset and facilitating multiple forecast horizons, the approach implemented in this study allows the mitigation of this risk using appropriate hyperparameter testing to optimize the LSTM model. With respect to benchmark methods, another deep learning technique (i.e., DNN), two machine learning techniques (i.e., MLP, SVR) and a statistical technique (i.e., ARIMA) are developed for comparison.

3. Theoretical Overview

3.1. Objective Predictive Model: Long Short-Term Memory (LSTM) Network

The LSTM algorithm, used recently in solar radiation modelling is a branch of the deep recurrent neural network (RNN) (Figure 1a), which is an advanced method of machine learning, feed-forward neuron networks (FFNNs) (Figure 1b) [88]. Both models apply the idea of the human brain neuron network in which each neuron (blue colour) is an information processing unit. The improvement of RNN over FFNN is feedback loops (in red colour), which are units with memory. These units can remember, re-incorporate and update information from patterns learnt from previous steps, thus, RNN can learn progressively, rather than randomly, as is the case with FFNN. The previous state of the neuron, that is, the parameters of the previous time step, can be re-incorporated and taken into account when updating the memory. However, this property of RNN causes the vanishing gradient problem that prevents RNNs learning from deep sequences of broad contexts [89].

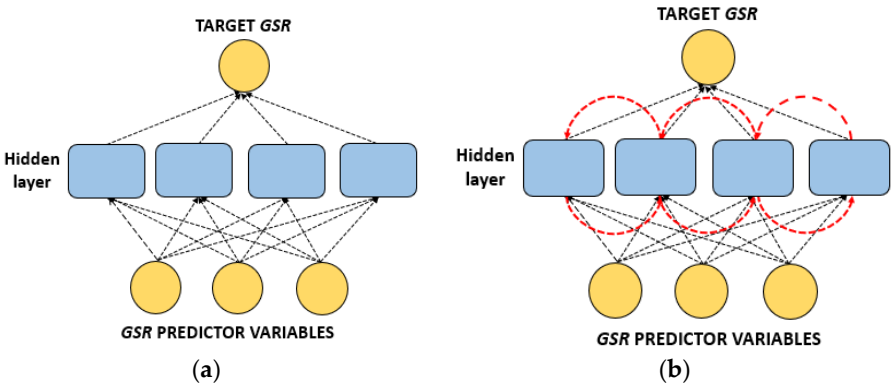

Figure 1. Descriptive flowchart of: (**a**) conventional feed-forward neuron network (FFNN), (**b**) recurrent neural network (RNN) based on deep learning methodology. The GSR predictor variables

could be either the exogenous variables related to the target GSR or statistically significant lagged GSR_{t-n} values capturing historical dependence of antecedent GSR to forecast the future values.

The long short-term memory (LSTM) algorithm was introduced by Hochreiter and Schmidhuber [90] in 1997 to address the vanishing gradient issue. Figure 2 illustrates the internal structure of LSTM with the innovative memory blocks called cells from which LSTM outperforms RNNs. From Figure 2, the transmission stage is between the previous hidden layer, the cell state and the next hidden layer. The cell state is the main chain of data flow, which allows the data to flow forward essentially unchanged. However, in this cell, there are specific gates, which allow some linear transformations to occur. The main utility of the LTSM model applied in real-time modelling contexts is its capability to learn long-term dependencies among the consecutive events on a relative timestamp through incorporating self-connected "gates" in the hidden units. In the context of GSR, especially at multiple forecasting horizons in this study, this model is likely to capture more accurately the real-time dependence of the historical, the current and the future GSR values, to finally create a more representative modelling framework. The gates enable LSTM units to read, write and remove information in the memory. Thus, they allow LSTM to remember the relevant data patterns while removing the irrelevant data, hence, sustaining a constant and relatively low error level, unlike the ARIMA (and other time-series) statistical model that uses its error to propagate into the future timescale, and potentially induces the inherent inaccuracies in the testing phase.

In terms of solar energy forecasting and applications in real time, the LSTM model is expected to exploit the temporal and spatial dependence of antecedent GSR data, while utilizing the contextual information. Consequently, in recent years, the LSTM has been implemented in many fields, including solar energy prediction [55,58], although the present study is the first of its kind to develop and apply this model for Vietnam, and, in particular, at multiple forecasting horizons.

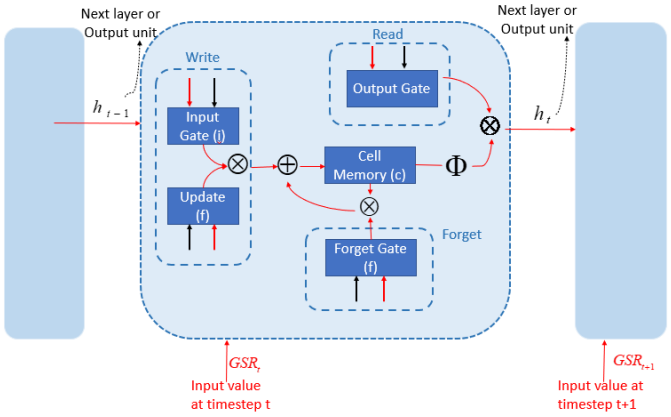

Figure 2. Descriptive flowchart of a long short-term memory (LSTM) unit in the first layer for time step t which is an advancement of the hidden layers of the RNN, as adopted for the proposed LSTM model. Reproduced from [91].

3.1.1. Computational Aspects of LSTM Network Model

To gain an in-depth understanding of LSTM, Figure 2 illustrates a single localized LSTM cell in the first layer of a network at the timestep t. Symbols $\otimes$ and $\oplus$ represent point-wise scalar multiplication and summation, respectively. The colour arrows $\downarrow\downarrow$ show direction of input to the systems. Φ is an activation function which sets the ReLU (rectified linear units) in this experiment. These units are known as the Input Gate, Update Gate, Output Gate, and Forget Gate, and they represent the output value at the separate gates [92]. The gates normally receive an input of the same LSTM unit's output, but obtained at a previous time step (h_{t-1}). These gates also receive the input data related to the current time step (x_t) in order to emulate the future value of GSR at any given timestep. With the same structure as RNN, a novel Forget Gate function enables inappropriate information to

be removed and forgotten, which resolves the gradient vanishing issues of the RNN algorithm when applied to a large dataset context.

Firstly, based on the last hidden state (h_{t-1}) and the new input x_t , LSTM possibly selects the information, which is to be processed from the cell state represented by the Forget Gate (f):

$$f_t = \sigma\left(w_f \times \left[h_{t-1}, x_t\right] + b_f\right) \quad (1)$$

Secondly, the next step is to determine the information that is stored in the cell state. There is a new candidate $\widetilde{c_t}$ which is generated by x_t and h_{t-1} through a tanh layer. This new candidate is then scaled by the Input Gate i:

$$\widetilde{c_t} = \tanh(w_c \times \left[h_{t-1}, x_t\right] + b_c) \quad (2)$$

$$i_t = \sigma\left(w_i \times \left[h_{t-1}, x_t\right] + b_i\right) \quad (3)$$

Then, by combining the previous cell state C , both C_{t-1} and c_t , in which the former is determined in the Forget Gate (f) and the latter is identified by the Input Gate (i) as Equation (4):

$$c_t = f_t \times c_{t-1} + i_t \times \widetilde{c_t} \quad (4)$$

The above three kinds of gates are not static. The recent state information h_{t-1} and the current input x_t are jointly determined by non-linear activation after linear combination.

Finally, in the output process, there are two steps. The Output Gate is known as a new gate which is responsible for deciding appropriate parts from the cell state to be outputted. The cell state c_t is activated by tanh function, which is then filtered through multiplying by o_t . The multiplication result is the desired output h_t :

$$o_t = \sigma(w_o \times \left[h_{t-1}, x_t\right] + b_o) \quad (5)$$

$$h_t = o_t \times \tanh(c_t) \quad (6)$$

where w_f, w_i, w_c, w_o are weight matrices, b_f, b_i, b_c, b_o are bias vectors. σ () is the sigmoid activation function.

3.1.2. Benchmark Model: Autoregressive Integrated Moving Average (ARIMA)

This study also adopts the autoregressive integrated moving average (ARIMA) model to further validate the efficacy of the LSTM Network model. The ARIMA model was popularized by the work of Box and Jenkins [93]. ARIMA analyses a set of (univariate) predictor data partitioned into a subset of input/target to validate the LSTM and other models. Using its own time-lagged information and the respective model errors, ARIMA can identify the intermittent and chaotic patterns of original GSR time-series data, which is an alternative effective skill when other methods (e.g., LSTM or others) are not available.

An ARIMA model includes three parameters (*p*, *d*, *q*), with *p* as the number of autoregressive terms, *d* as non-seasonal differences and *q* as the number of lagged errors. The ARIMA process generally involves model identification, estimation and forecasting, defined as follows:

$$\psi_p(B)(1-B)^d Y_t = \delta + \theta_q(B)_{\varepsilon_t} \quad (7)$$

in which ψ_p —the autoregressive parameter of order *p*; *B*—the backshift operator; Y_t —the original predictor dataset; δ —constant value; θ_q —the moving average parameter *q*; and *d* is the differencing order used for the regular or non-seasonal part of the series.

In the identification of an ARIMA model, the differencing parameter (*d*) is analysed by autocorrelation and partial autocorrelation to decide whether the differencing effect should be performed in a non-stationary dataset. Furthermore, *p* and *q* terms are identified for the model by

analysing maximum likelihood estimation, which determines parameters maximising the probability of data by least squares. There are various terms (e.g., log likelihood, Akaike's information criterion (*AIC*), Bayesian information criterion (*BIC*), *r*, RMSE) to determine the maximum combination of (*p*, *d*, *q).* Expressions of *AIC* and *BIC* are as follows:

$$AIC = -2\log(L) + 2(p + q + k + 1) \quad (8)$$

$$BIC = -2\log(L) + (p + q + k) * (p + q + k + 1)\log(n) \quad (9)$$

where L is the log likelihood of data, $k = 1$ if $c = 0$ and $k = 0$ if $c = 0$; n is the sample size. The last term in brackets is the number of parameters (including the variance of the residual)

A detailed description of the ARIMA model can be found elsewhere and further applications of this method can be found in other's works [31,94,95]. Generally, the ARIMA model assumes a scenario where there is no change in consecutive periodical measurements or the readings used to construct a model. Given that previous studies have employed an ARIMA model for GSR forecasting, this technique is also employed in this study in the interest of its ability in capturing historical patterns from the present time-series data.

3.1.3. Benchmark Model: Support Vector Regression (SVR)

Support vector regression (SVR) is a regression version of the SVM model that is usually applied in solar energy forecasting [96,97]. SVR transforms the original feature space into a high-dimensional one using a hyperplane which employs kernel functions (e.g., Gaussian, linear) to effectively separate data [98]. Herein, SVR is implemented using Python environment version 3.6 using the Sklearn library which is optimized using a grid search procedure.

3.1.4. Benchmark Model: Deep Neural Network (DNN)

The deep neural network is a machine learning method that has been advanced based on artificial neural networks (ANN), and is capable of trained complex input and learning procedures [45]. Similar applications of DNN in solar energy forecasting can be found in [17,99]. Herein, the deep learning library of Python retrieving solar radiation is applied. Like other neural network methods, the employed model comprised one input/output layer and multiple hidden layers. Various structures of the deep neural network were analysed to determine an optimal training model.

3.1.5. Benchmark Model: Multilayer Perceptron Network (MLP)

The multilayer perceptron network (MLP) is the most common type of feed-forward network [100]. MLP has three layers: an input layer, an output layer and a hidden layer. In this paper, MLP is implemented by the Python environment version 3.6 using the deep learning library. As for LSTM and DNN, various structures of MLP were analysed to determine an optimal training model.

4. Materials and Method

4.1. Study Region

The data utilized to build and evaluate the proposed LSTM network model comprised the minute interval time-series of global solar radiation (GSR), acquired from the reliable source of World Bank repositories from September 2017 to June 2019. The Vietnam government aims to develop large solar plants near its capital city that can reduce load emissions and avoid a downwind situation. A chosen location, the Bac-Ninh region (Figure 3), is a small city with about 100,000 people. The city is located in the North of Vietnam, not far from the capital, at latitude 21.2013° N and longitude 106.0629° E, and elevation of 60 m above sea level. The Bac-Ninh site has a subtropical dry winter climate characterized by hot and humid summers with frequent tropical downpours of short duration, and warm and frequently dry winters [101]. This province is undergoing revitalization in terms of more sustainable future solar energy systems, which are partly funded by the World Bank. The province also meets the criteria set for the selection of the present study location for the future

installation of solar measurement stations, i.e., it is solar-rich with terrain either flat or characterised by low obstructions, homogeneous landscape and land-usage (clearly represented by satellite pixels for validation), without large water bodies, mountains, dirt roads, industrial pollution, open-pit mining operations, or a danger of flooding [102,103]. Thus, the development of a solar forecasting model at multiple forecast horizons, especially in Bac-Ninh, is a justified research endeavour to support the United Nation's Sustainable Development Goal #7 related to the access to affordable renewable energies for all populations.

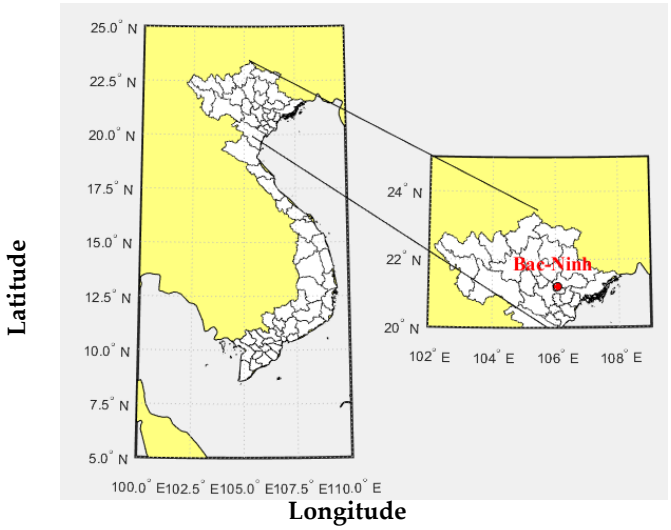

Figure 3. The study site in Bac-Ninh—Vietnam's solar city—where the proposed LSTM model was developed for multiple forecast horizons of global solar radiation (GSR).

4.2. Data Preparation

This section details the activities of related to data preparation, including the construction of multiple time-scale datasets using 1 min original measurements, the handling of missing data, and the input of those data structures into the LSTM network. Notably, GSR measurements were performed simultaneously, 24 h a day, at equidistant time intervals of 1 min. Only the data from 06:00 to 18:00 were used for designing the predictive model as these times represent a period of meaningful daylight hours.

With the aim to construct a framework for a near-real-time prediction model, the raw 1 min time-series data were firstly used to generate the data at 5-, 10- 15- and 30 min-ahead time-series data, which were then used as the target variable. Details of those data are presented in Table 2.

Table 2. Descriptive statistics of the global solar radiation (GSR) dataset aggregated at various timescales for the Bac-Ninh region in Vietnam used to develop the prescribed LSTM model tested at multiple forecast horizons.

Forecasting Horizon	Data Period	Minimum Wm^{-2}	Maximum Wm^{-2}	Mean Wm^{-2}	Standard Deviation Wm^{-2}
1 min (1M)	1 June 2019 to 30 June 2019	0	1376	207	283
5 min (5M)	1 March 2019 to 30 June 2019	0	6047	730	1157
10 min (10M)	27 September 2017 to 30 June 2019	0	11,889	1416	2297re
15 min (15M)	27 September 2017 to 30 June 2019	0	16,574	2124	3430
30 min (30M)	27 September 2017 to 30 June 2019	0	32,042	4249	6804

With respect to the missing data, it is noted that missing values represent only 0.15% of the time-series data, and were due to equipment faults or site closures in the measured period. We imputed those missing values by the mean value of the whole period [95,104]. Clearly, more powerful techniques (e.g., step-wise linear regression fit, Kalman filters) could be considered and might facilitate better imputation but given the relatively low percentage of missing data, these may not be

required in this study. Moreover, these data are employed for the comparison of LSTM with the other models. Consequently, the imputation method should not influence the relative performance of the alternative forecasting methods.

To prepare the suitable number of inputs for each time-scale horizon (based on historical behaviour of short-term solar radiation measurements), the autocorrelation coefficient and the partial correlation coefficient (PACF) were employed. The detailed procedures can be found in [94]. Explicitly, the PACF function computes a time-series regression against its *n*-timescale lagged values by removing the dependency on intermediate elements and identifying those patterns potentially prevalent in the future GSR data that are correlated to the antecedent GSR data. This procedure aims to develop forecast models that consider the role of memory (i.e., antecedent GSR) adopted in forecasting the current GSR value, and possibly, considers several other atmospheric factors that could potentially influence ground level GSR. Consequently, the input vector $GSR_{t-1}, GSR_{t-2}, ..., GSR_{t-n}$, called the *n*-lagged set of inputs deduced from the PACF method, was then used as the LSTM model's input to predict the GSR as the target. Figure 4 shows the PACF plot of GSR time series with lagged inputs as predictor variables for the LSTM model applied at 1M, 5M, 10M, 15M, and 30M forecast horizons.

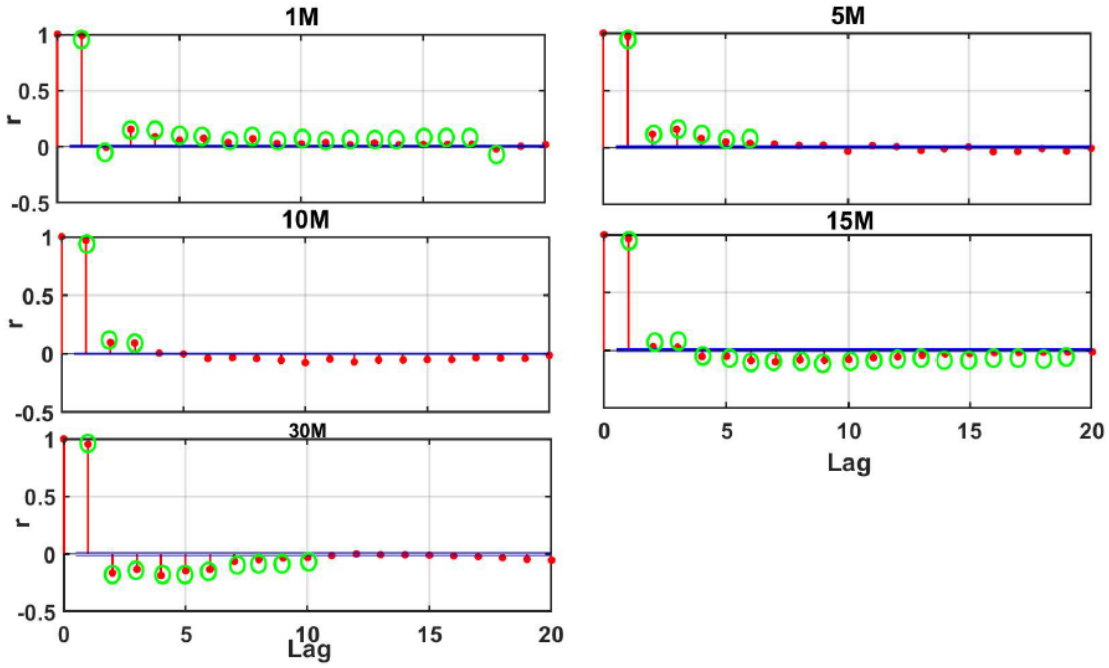

Figure 4. Partial autocorrelation function (PACF) plots of the GSR time series showing lagged inputs autocorrelation of GSR as predictor variables used for GSR models at different forecast horizons (i.e., 1M, 5M, 10M, 15M, and 30M timescales). The blue lines denote the statistically significant boundary at the 95% confidence interval. The green circles are used to show the zone outside which the target GSR has statistically significant correlations with its antecedent values.

The primary scope of this study is to design, for the first time in the present study region, an LSTM model that has the capability to forecast near real-time GSR using minute interval data, applied for multi-step forecasting horizons. To expand the practical scope of the modelling techniques, the developed model was applied at 1 min (1M), 5 min (5M), 10 min (10M), 15 min (15M) and ½ hourly (30M) forecasting horizons, to enable LSTM to generate a more granular interval GSR, as required in real-life decisions, for example, through constant monitoring of solar energy resources. Hence, the primary task is to construct a matrix of a training and testing dataset that can reliably be applied to the proposed LSTM model.

The normalization of modelled data was accomplished by statistical rules to overcome the numerical difficulties caused by the data features, patterns and fluctuations using the conventional methods of feature scaling [105]. Normalization is applied to the *n*-lagged inputs [9] to be in the range of [0–1] by the following formula:

$$GSR_N = \frac{GSR_{ACTUAL} - GSR_{MIN}}{GSR_{MAX} - GSR_{MIN}} \quad (10)$$

$$GSR_{ACTUAL} = GSR_N(GSR_{MAX} - GSR_{MIN}) + GSR_{MIN} \quad (11)$$

After normalization, a major task was to determine the training data, to construct the predictive model, and the testing data, and thus achieve the highest performance. The partitioning of data followed the notion that researchers use different divisions between testing and training sets, which generally vary with the problem. There is no 'rule of thumb' for data divisions. In [58], the authors employed about 75% of inputs for training and the remainder for the testing set, while in [44] the partition proportion for training and testing sets was approximately 80:20. Subsequently, the normalized data are then divided into the training (80%) and testing (20%) sets (Table 3a). Noticeably, the number of data points is significantly higher than any of the previous relevant papers [15,64].

Table 3. (**a**) Model designations and various input combinations based on the partial autocorrelation function used to identify antecedent lagged GSR for the objective LSTM and the comparative counterparts adopted at multi-step forecasting horizons. (**b**) Training phase results of LSTM with different partition proportions.

(a)

Forecast Horizon	Significant Lagged GSR	Number of Data Point	Training	Validation	Testing
			80%	Percentage of Training Data	20%
1M	18	43,197	34,558		8637
5M	6	35,133	49,518		12,376
10M	3	92,874	24,757	10%	6187
15M	19	61,897	34,558		8637
30M	10	30,946	28,106		7025

(b)

Model	Training-Testing Proportion	*r*	*RMSE* (Wm^{-2})	*MAE* (Wm^{-2})
LSTM	80–20	0.9957	32.086	13.670
-	70–30	0.9901	43.7088	15.715
-	60–40	0.9799	60.9179	23.402

4.3. LSTM Model Implementation

Prior to developing the proposed LSTM-based solar radiation forecasting model, the historical GSR data were pre-processed at multiple forecasting time horizons. The proposed model-based LSTM was developed under the Python environment on an Intel Core i5 and 16 GB RAM computer.

The development and validation of the proposed method, as shown in Figure 5, is presented in the following steps:

Step 1: Construct the data matrix which is used as the input in the first layer. The statistically significant lags were calculated from the original GSR time series data using the partial autocorrelation function (PACF). In addition, to demonstrate data efficiency in this model, we also used different partition proportion in dividing training and testing sets (Table 3b). As a result, the scale of 80:20 represents the highest performance of the LSTM, therefore, this scale is appropriate in this study.

Step 2: After incorporating the significant lagged inputs as the input predictor, the LSTM was implemented using the Keras deep learning package in Python [106]. The input layer of the trained LSTM network had four timesteps; hidden neurons were set to 80; and the output layer with a linear activation function had one neuron. In addition to these fixed values, we ran the LSTM model with

different combinations of hyperparameters (epoch-drop rate-batch size) which were selected through a grid search. Table 3a summarizes the general architecture of different types of LSTM by various combinations of hyperparameters. However, only those experiments for 1M are shown.

Step 3: To select the optimal model for each case, the LSTM algorithm begins with the change of each hyperparameter. Then, based on the recorded evaluation metrics (*r*-value, RMSE) in the training phase, we selected the optimal LSTM model with the highest *r*-value and the lowest RMSE. Table 4b presents the experimental results in the training phase with the optimal models highlighted in red. However, only those experiments for 1M are shown. After conducting all experiments, the summarized results of the optimal model for all forecasting time horizons (1M, 5M, 10M, 15M and 30M) are shown in Table 4c.

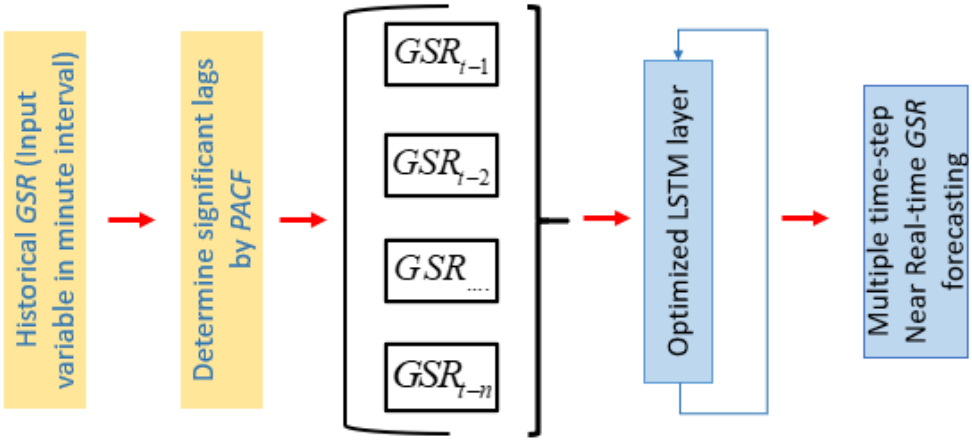

Figure 5. Descriptive flowchart for the relevant steps in the LSTM model design phase.

For the case of an LSTM network model, the computational cost is considered to be an important aspect in terms of learning process [107], which is directly influenced by the dataset size in the training phase [107] and the respective hyperparameters [108]. To reduce the high computational cost in the learning process of the objective (i.e., LSTM) model, the hyperparameters for the model are chosen through a grid search process for the optimal parameters; however, this can be relatively time-consuming. For instance, for each LSTM model, the search took about 11–12 h; however, when the optimal hyperparameters are determined prior to running the primary LSTM model, the computational time of the model was reduced to a much lower value (<15 min).

Generally, a hyperparameter is a parameter whose value is set before the learning process commences. There are two types of hyperparameters, namely, model hyperparameters (e.g., the size of the neural network and the number of input layers in FFNNs) and algorithm model hyperparameters (e.g., dropout and batch size). Model hyperparameters cannot be inferred during the training process since this must be referred to the model selection task. The latter, algorithm hyperparameters, in principle, can increase the speed and quality of the learning process. Therefore, determining the most appropriate hyperparameters is essential for the success of a deep learning model such as the LSTM model adopted in this study. Depending on the model types, strategies for choosing hyperparameters may vary. While some of the hyperparameters are model-specific, some common hyperparameters that can be used in any deep learning model, and that were also adopted in this study, are:

- Epoch defines the number of times that the learning algorithm will work through the entire training dataset. The number of epochs is usually hundreds or thousands, allowing the learning algorithm to run until the error from the learning model is minimized. In this study, the number of epochs is set to a maximum of 2000 (Table 4).
- Batch size defines the number of data points that are propagated through the network. The batch size can be seen as a for-loop iterating over one or more data points. At the end of each batch,

the predicted values are compared to the actual values and the errors are calculated. From these errors, the update algorithm is used to improve the model. Depending on data length, to determine whether a greater batch size can provide the better performance, the batch size is set as in Table 3.

- Dropout is a regularization layer that blocks a random set of cell units in one iteration of LSTM training. Since over-fitting is prone during training, the dropout layer creates blocked units which can remove connections in the network. Therefore, it possibly decreases the number of free data points to be predicted and the complexity of the network. The dropout rate is often set between 0 and 1. In this study, this parameter was tested between two values, 0.1 and 0.2, to determine whether a greater value of dropout rate improves LSTM performance (Table 4a).
- Least absolute deviations and least square error (L1 and L2 regulation): In addition to dropout, the L1 and L2 regularization parameter is also used such that the L1 and L2 penalization parameter decreases the sum of absolute differences and the sum of square of differences between observed and forecasted values. In principle, adding a regularization term to the loss will facilitate a better network mapping (by penalizing large values of parameters which minimize the amount of nonlinearity of GSR values).
- Activation function: With the exception of the output layer, all the layers within a network typically use the same activation function known as the rectified linear unit (ReLU).

Table 4. (**a**) The architecture of the objective LSTM model with various model design parameters. Note that ReLU stands for rectified linear unit. (**b**) Experimental results of LSTM model in 1M forecast horizon in the training phase. (**c**) The optimal architecture used in designing optimized LSTM vs. the comparative models (i.e., autoregressive integrated moving average (ARIMA), deep neural network (DNN), multilayer perceptron (MLP), support vector regression (SVR)) in the training phase.

(a)

Model	Model Hyperparameters	Search Space for Optimal Hyperparameters
LSTM	Hidden neurons	(100, 200, 300, 400, 500)
-	Epochs	(1000, 1200, 1500, 2000)
-	Optimizer	(Adam)
-	Drop rate	(0.1, 0.2)
-	Activation function	(ReLu)
-	Layer 1 (L1) and Layer 2 (L2), Layer 3 (L3)	(50, 40, 40)
-	Batch size	(400, 600, 700, 750, 800)

(**b**)

Sequence	Initial Set-Up Epoch	Actual Used Epoch	Drop Rate	Batch Size	r	*RMSE* (Wm^{-2})
1	2000	54	0.1	500	0.9874	33.201
2	2000	55	0.1	750	0.9875	33.178
3	2000	53	0.1	800	0.9884	33.098
4	2000	62	0.1	1000	0.9876	33.096
5	2000	64	0.2	800	0.9956	32.086

(**c**)

Time-Horizon	GSR Model	Design Parameter		r	*RMSE* (Wm^{-2})
1M	LSTM	Number of epochs-Drop rate-Batch size	64-0.1-800	0.9956	33.2012
	DNN	Number of epochs-Drop rate-Batch size	162-0.1-500	0.990	44.0424
	MLP	-	-	0.9821	61.7642
	ARIMA	*p-d-q*	0-1-0	0.9808	57.6876
	SVR	*Cost Function (C), Epsilon (ε)*	1.0-1.0	0.9846	59.2223
5M	LSTM	Number of epochs-Drop rate-Batch size	59-0.2-800	0.9714	265.5456
	DNN	Number of epochs-Drop rate-Batch size	199-0.1-500	0.9650	1338.4922
	MLP	-	-	0.9721	361.7641
	ARIMA	*p-d-q*	0-1-0	0.9724	287.7479

	SVR	*Cost Function (C), Epsilon (ε)*	1.0-1.0	0.9218	389.6317
10M	LSTM	Number of epochs-Drop rate-Batch size	59-0.2-800	0.9914	26.4411
	DNN	Number of epochs-Drop rate-Batch size	194-0.1-500	0.9871	26.6411
	MLP	-	-	0.9599	26.3175
	ARIMA	*p-d-q*	0-1-0	0.9205	22.622
	SVR	*Cost Function (C), Epsilon (ε)*	1.0-1.0	0.9514	51.6976
15M	LSTM	Number of epochs-Drop rate-Batch size	70-0.2-500	0.9653	76.9883
	DNN	Number of epochs-Drop rate-Batch size	162-0.1-500	0.9657	88.9887
	MLP	-	-	0.9547	220.7234
	ARIMA	*p-d-q*	0-1-0	0.9618	1033.4372
	SVR	*Cost Function (C), Epsilon (ε)*		0.8983	117.7362
30M	LSTM	Number of epochs-Drop rate-Batch size	62-0.2-500	0.9572	710.7477
	DNN	Number of epochs-Drop rate-Batch size	28-0.1-700	0.9067	709.0347
	MLP	-	-	0.9192	900.1132
	ARIMA	*p-d-q*	0-1-0	0.8859	952.8502
	SVR	*Cost Function (C), Epsilon (ε)*	1.0-1.0	0.8314	1270.7158

4.4. Benchmark Models Implementations

To comprehensively evaluate the optimal LSTM forecasting model, five other popular forecasting models based on the ARIMA, DNN, MLP, and SVR algorithms were developed under the Python environment, version 3.6, on an Intel Core i5 computer. For the purpose of brevity and conciseness, only the results at the 1 min (1M) forecasting horizon are shown here, but the results obtained at the other forecasting horizons resulted in relatively similar deductions. Finally, following the previous steps, the optimal models based on the LSTM versus the counterpart models are shown in Table 4c for a diverse range of forecasting horizons.

5. Model Performance Criteria

Several methods have been previously adopted to evaluate model performance [109]. In the present work, a popular set of statistical metrics (e.g., bias, mean square error, linear correlation coefficient) are employed to assess the model performance since each individual metric has its own strength and weakness [110]. For instance, due to the standardization of observed and forecasted means and variance, the robustness of Pearson's correlation coefficient (r), which exceeds 1 as the perfect model, may have limited meaning [70,95]. Moreover, while root mean square error (RMSE) is relevant for high values, mean absolute error (MAE) assesses all deviations of observed data both in the same manner and regardless of sign [111]. RMSE and MAE are recommended to address each other's weaknesses and to obtain accuracy in an absolute unit [111]. The performance of a model can be decreased because of partial peaks and higher magnitudes, which may cause large errors and be insensitive to small magnitudes. To solve this problem, efficiency measurement indexes, such as Willmott (WI) and Nash-Sutcliffe (E_{NS}) [112] are introduced with the advantage of overcoming insensitivity and over-dominance of significant errors over small errors [113,114]. Nevertheless, E_{NS} is relatively high even in the poorly-fitted models and vice versa, hence, it can confuse performance evaluation [115]. Therefore, WI is implemented to be incorporated with E_{NS} [112].

However, a certain degree of insufficiency still occurs with WI that can be improved by the Legate and McCabe index *(LM)* [116]. Since different forecasting horizons can lead to differences in data distribution, the relative root mean square error (RRMSE) [117] and mean absolute percentage error (MAPE) [118] are computed since they are also the benchmark of evaluating a "good" model. A model's precision level is excellent if RRMSE < 10%, good if 10% < RRMSE < 20%, fair if 20% < RRMSE < 30%, and poor if RRMSE > 30% [117]. Therefore, to properly assess model performance, in this paper, several statistical score metrics are exploited, such as the Pearson's correlation coefficient

(*r*) [119], root mean square error [120] (RMSE; $\mathrm{W\,m^{-2}}$), mean absolute error [121] (MAE; $\mathrm{W\,m^{-2}}$), including the relative error values: RMSE (RRMSE; %), MAE, MAPE, WI, and E_{NS}, which are adopted as the well-known metrics employed elsewhere (e.g., [79]).

$$r = \left(\frac{\sum_{i=1}^{n} (GSR^{OBS} - < GSR^{OBS} >)(GSR^{FOR} - < GSR^{FOR} >)}{\sqrt{\sum_{i-1}^{N} (GSR^{OBS} - GSR^{OBS})^2} \sqrt{\sum_{i-1}^{N} (GSR^{FOR} - GSR^{FOR})^2}} \right)^2 \tag{12}$$

$$MAE = \left(\sum_{N}^{t=1} \left| GSR^{OBS} - GSR^{FOR} \right| \right) / N \tag{13}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (GSR^{OBS} - GSR^{FOR})^2} \tag{14}$$

$$E_{NS} = 1 - \frac{\sum_{i-1}^{N} \left[GSR^{OBS} - GSR^{FOR} \right]^2}{\sum_{i-1}^{N} \left[GSR^{OBS} - GSR^{FOR} \right]^2}, -\infty \le E_{NS} \le 1 \tag{15}$$

$$WI = 1 - \frac{\sum_{i-1}^{N} \left[GSR^{OBS} - GSR^{FOR} \right]^2}{\sum_{i-1}^{N} \left[\left| (GSR^{FOR} - (GSR^{OBS})) \right| + \left| (GSR^{OBS} - (GSR^{FOR})) \right| \right]^2}, 0 \le WI \le 1 \tag{16}$$

$$LM = 1 - \frac{\sum_{i=1}^{n} \left| GSR^{OBS} - GSR^{FOR} \right|}{\sum_{i=1}^{n} \left| GSR^{OBS} - < GSR^{OBS} > \right|}, -\infty \le LM \le 1 \tag{17}$$

$$MAPE = (\sum_{N}^{t=1} \left| (GSR^{OBS} - GSR^{FOR}) / GSR^{OBS} \right|) / N \tag{18}$$

$$RRMSE = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (GSR_i^{FOR} - GSR_i^{OBS})^2}}{\frac{1}{N} \sum_{i=1}^{N} GSR_i^{OBS}} \times 100 \tag{19}$$

6. Statistical Significance Testing

Based on performance metrics it is difficult to conclude whether the results are due to chance or decisive. We possibly reject a factually good parallel model since the performance metrics might be generated stochastically. Consequently, from a statistical perspective, a significant difference of forecasting performance cannot be solely judged by traditional performance metrics. Therefore, this paper employed a modern statistic evaluation method, the Diebold–Mariano (DM) test [122], which can offer a quantitative method to evaluate the forecast accuracy of forecasting model. The DM test is applicable to nonquadratic loss functions, multi-period forecasts and forecast errors that are non-Gaussian and nonzero-mean. Details of the DM test can be found in [122] and some applications of the DM test can be found in [123,124]

Finally, with the help of the DM test, the interference by sample stochastic difference can be revealed, such that the better forecasting model can be identified statistically. To determine whether one forecasting model is better than another, we might first test the equal accuracy hypothesis. A null hypothesis (h_o) means that the observed differences between the performances of two forecasting models are not significant. An alternative hypothesis (h_1) means that the observed differences

between the performances of two forecasting models are significant. Since the DM statistics converge to a normal distribution, we can reject the null hypothesis at the 5% level of significance if |DM| > 1.96, otherwise, we cannot reject the null hypothesis (h_o). If the DM statistic value does not meet the acceptable criterion, then the null hypothesis cannot be accepted, i.e., the two forecasts are statistically not different. By comparing LSTM to each counterpart in turn, it can be concluded whether the LSTM model is superior than its counterparts.

7. Results and Discussion

In this section, the experimental results and the overall performance assessment at different forecasting horizons are presented. For each modelling experiment, five GSR forecasting models, including the proposed LSTM model and the counterpart models (i.e., ARIMA, DNN, MLP, and SVR) are employed. To demonstrate the merits of the LSTM model over the counterpart models in terms of their near-real-time solar forecasting capability, a plethora of model evaluation metrics for the testing phase, as described by Equations (9)–(16), is presented in Tables 5–7.

Table 5. The model performance in the testing period as measured in terms of correlation coefficient *(r)*, root mean squared error (RMSE), mean absolute error (MAE).

Predictive Model	*r*					RMSE Wm^{-2}					MAE Wm^{-2}				
	1M	**5M**	**10M**	**15M**	**30M**	**1M**	**5M**	**10M**	**15M**	**30M**	**1M**	**5M**	**10M**	**15M**	**30M**
LSTM	0.9920	0.9999	0.9999	0.9578	0.9531	40.9125	1400	18.6627	79.7273	731.7482	21.6428	1059	12.3368	43.8383	409.7196
MLP	0.9780	0.9266	0.9062	0.9246	0.8554	65.7511	1852	88.9537	218.7543	1254.3440	34.2960	1326	53.8914	106.8205	778.1039
DNN	0.9910	0.9606	0.9998	0.9568	0.9094	44.4086	1570	61.8762	86.6580	940.4280	25.5140	1134	40.9027	49.0178	576.9220
ARIMA	0.9902	0.9607	0.9989	0.9584	0.9094	52.9785	1589	37.8037	161.1655	937.1356	31.9632	1149	24.9898	100.5221	571.3325
SVR	0.9856	0.9266	0.9358	0.9247	0.8555	56.1271	1619	74.8298	99.9360	1244.6963	31.5000	1136	42.1483	70.2232	773.6047

Table 6. The model performance in the testing period as measured in terms of Willmott's index (WI*)*, Nash–Sutcliffe Efficiency (E_{NS}) and relative root mean square error (RRMSE).

Predictive Mode*l*	*WI*					E_{NS}					RRMSE (%)				
	1M	**5M**	**10M**	**15M**	**30M**	**1M**	**5M**	**10M**	**15M**	**30M**	**1M**	**5M**	**10M**	**15M**	**30M**
LSTM	0.9984	0.9409	0.9989	0.9770	0.9811	0.9831	0.6420	0.9920	0.8712	0.8931	9.9278	51.7123	10.1362	42.1591	41.0858
MLP	0.9959	0.8816	0.9721	0.9167	0.9347	0.9563	0.3737	0.8188	0.0306	0.6859	15.9581	68.3986	48.3132	115.6755	70.4282
DNN	0.9981	0.9227	0.9844	0.9717	0.9718	0.9800	0.5500	0.9123	0.8479	0.8235	10.7782	57.9785	33.6067	45.8240	52.8026
ARIMA	0.9972	0.9202	0.9947	0.9500	0.9700	0.9716	0.5386	0.9673	0.4738	0.8247	12.8582	58.7073	20.5322	85.2231	52.6178
SVR	0.9969	0.9179	0.9801	0.9635	0.9364	0.9681	0.5212	0.8718	0.7977	0.6907	13.6223	59.8060	40.6421	52.8454	69.8865

Table 7. The model performance in the testing period as measured in terms of the Legates and McCabe index (LM) and mean absolute percentage error (MAPE).

Predictive Model	LM					MAPE (%)				
	1M	**5M**	**10M**	**15M**	**30M**	**1M**	**5M**	**10M**	**15M**	**30M**
LSTM	0.9204	0.4658	0.9275	0.7575	0.7741	16	48	100	86	116
MLP	0.8739	0.3311	0.6832	0.4090	0.5710	47	54	127	143	49
DNN	0.9062	0.4279	0.7596	0.7288	0.6819	49	58	60	62	67
ARIMA	0.8825	0.4203	0.8531	0.4438	0.6850	233	267	151	127	85
SVR	0.8842	0.4272	0.7522	0.6115	0.5734	56	91	143	120	262

For all of the modelling experiments capturing the highest Pearson's correlation coefficient (*r*), the lowest root mean square error (RMSE), and the lowest mean absolute error (MAE), the proposed LSTM model achieves better results than the counterpart models executed for multiple time horizons. In particular, the LSTM model-simulated 1M forecast horizon outperforms all of the other developed models with the statistics r = 0.9920, RMSE = 40.9125 Wm^{-2}, and MAE = 21.6428 Wm^{-2} vs. r = [0.9920, 0.9780], RMSE = [40.9125, 65.7511], and MAE = [21.6428, 34.2960] where [] represents the upper and lower bounds of the metrics for the various models.

Figure 6 illustrates scatterplots for the observed and the forecasted GSR values of the developed models for the 1M horizon. In each panel, the coefficient of determination (R^2) and a linear fit equation ($GSR^{OBS} = mGSR^{FOR} + c$) are shown to demonstrate the coherence between forecasted and observed GSR [104]. Here the constants—'*m*' (gradient) and '*c*' (intercept on the *y*-axis)—and R^2 are utilized to outline the models' overall accuracy. Note that R^2 and *m* values close to 1.00 and *c* value close to 0 should be attained for a perfect forecasting model. Evidently, the LSTM model achieves a better degree of agreement than the corresponding counterpart models. Moreover, to demonstrate the LSTM model's outstanding performance in predicting the GSR data in the testing phase, Figure 6 also shows a time-series plot for all of the study cases, for which the forecasted values of LSTM (in red) appear to be closer to the observed GSR values (in blue) compared to those of the counterpart models.

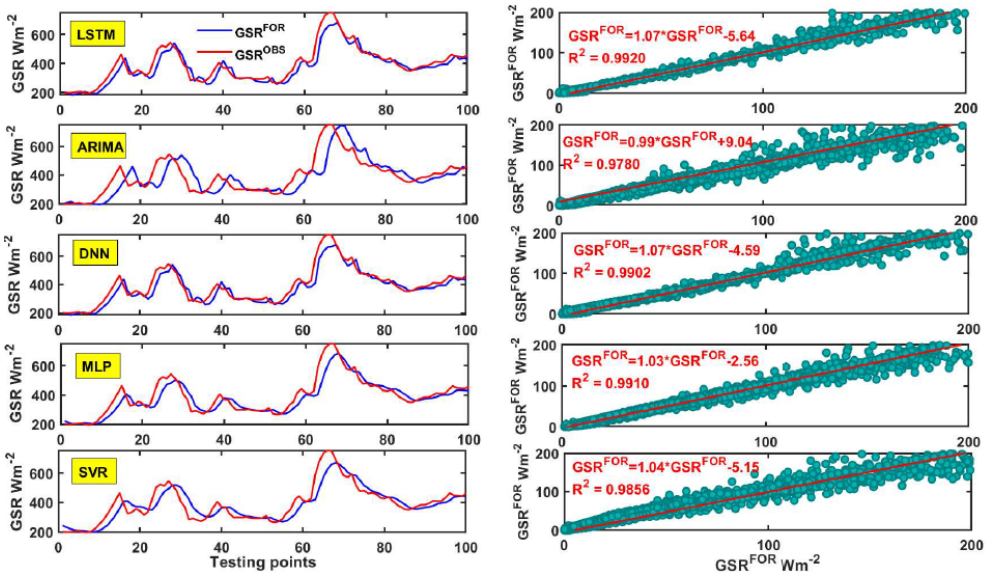

Figure 6. Times-series plot and scatter plot of the forecasted ('GSR_{FOR}') vs. observed ('GSR_{OBS}') data for each model in the testing period for 1M forecast horizons. A least squares regression line and the coefficient of determination (R^2) are shown in each panel.

To further explore the precision of the proposed LSTM model, Table 6 presents the metrics evaluating the forecasting errors (i.e., RRMSE, LM and MAPE). As can be seen, the proposed LSTM model is seen to outperform the counterpart models in all of the study cases in terms of the lowest RRMSE and MAPE and the highest value of the LM index. Evidently, the LM values produced by LSTM for all of the multiple forecasting horizons outperform those of both of the counterpart models. For instance, LM in the 1M model is 0.9204, whereas those of counterpart models (i.e., MLP, DNN, ARIMA, and SVR) are 0.8739, 0.9062, 0.8825, and 0.8842, respectively. While it is argued that RRMSE is limited in the context of a dataset with the same variance, in our case, the RRMSE value clearly shows us which model would be better in terms of producing fewer and relatively low-magnitude errors [125]. Thus, LSTM certainly performs better than the counterpart models as it generated an RRMSE that is lower than that of the comparative models. Meanwhile, in terms of the MAPE value,

the results of the proposed LSTM over multiple time horizons yield values of 16%, 48%, 100%, 86% and 116%, respectively, implying that the LSTM does not perform particularly well. However, the disadvantage of MAPE, which could perhaps explain this result, is that it generates a substantial percentage error for near-zero observed values as infinite MAPEs, and this effect can be quite pronounced if the observed GSR values are less than 1 [126]. This is a reasonable explanation for low performance in terms of MAPE since GSR time-series data, particularly at the very short-term time-scales used in this study, are expected to intermittently contain numerous near-zero values in the morning as observed elsewhere [125].

Figure 7 illustrates the boxplots for the case of the 1M model that depict the different forecasting skills regarding the absolute prediction error (i.e., forecasted—observed GSR values). The lower and upper lines of the boxplot denote the first and third quartile values (25th and 75th percentiles), respectively, and the median value (50th percentile) is represented by the central line. Additionally, two horizontal lines are also drawn from the first and third quartiles to the smallest and largest non-outliers, respectively. To concur with earlier results, the boxplot provides further justification that the distributed errors for the proposed LSTM model also acquire a much smaller spread with a correspondingly smaller magnitude of the quartile statistics and median values compared to the MLP, DNN, ARIMA, and SVR models.

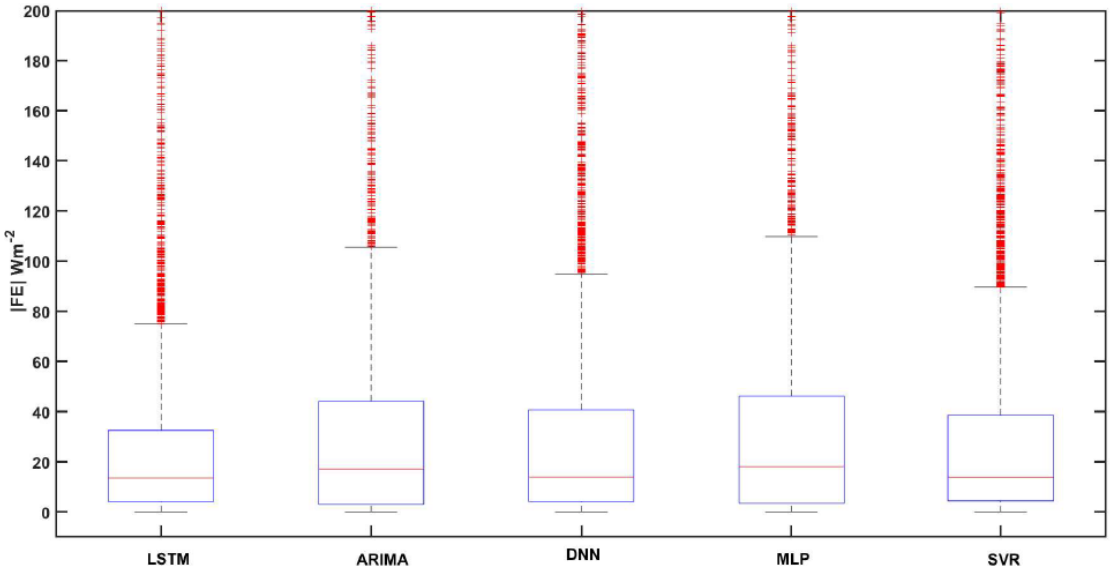

Figure 7. A boxplot of absolute values of the forecasted error in GSR generated by the LSTM model against the comparative models, i.e., ARIMA, DNN, MLP, and SVR, in the model's testing phase for the 1-min forecasting horizon.

Lastly, to consolidate the findings presented so far that demonstrate the efficacy of LSTM model, a Taylor diagram that determines the angular location to the inverse cosine of the correlation coefficient is presented in Figure 8 to show the closest model in respect to the observed data in the testing period. The correlation coefficient (*r*), on the radial axis, and the standard deviation, on the polar axis, are used simultaneously to judge the closest fitting model. For all different timescales, the LSTM generates the highest value of *r*, with the forecasted results being closest to the observed data compared to the other comparative approaches.

In addition, the forecasting performance of the five models is compared by the DM test (Section 5). The forecasting comparison of every pair of models is summarized in Table 8. The null hypothesis means that the two forecasts have the same accuracy, otherwise, the two forecasts have different levels of accuracy in the alternative hypothesis. The statistically significant better performance of LSTM over the counterparts is indicated as 'yes'. From Table 8, the conclusions of comparison of the LSTM model and the counterparts (i.e., DNN, MLP, ARIMA and SVR) can be drawn as follows.

Firstly, since the absolute value of the DM statistic in most cases is greater than 1.96, the null hypothesis is rejected at the 5% level of significance; that is to say, the observed differences are significant and the forecasting accuracy of LSTM models is better than that of the counterparts. The exceptions are the comparison of LSTM vs. DNN at the 1M forecast horizon and that of LSTM vs. SVR at the 15M forecast horizon, with corresponding absolute DM statistics of 0.272 and 0.268, respectively, which are less than 1.96. This shows the performance of LSTM vs. DNN and LSTM vs. SVR are not significant and might be due to stochastic interference. These clearly prove that the LSTM models receive more significance than the others. In addition, the *p*-value at a 5% level of significance is less than 0.05, which means all models are statistically significant.

In summary, by an evaluation of forecasting based on performance metrics and the DM test, the LSTM model was demonstrated to outperform the comparative models. Thus, it is found to be a versatile solar forecasting tool, especially over short-term, multiple timescale horizons.

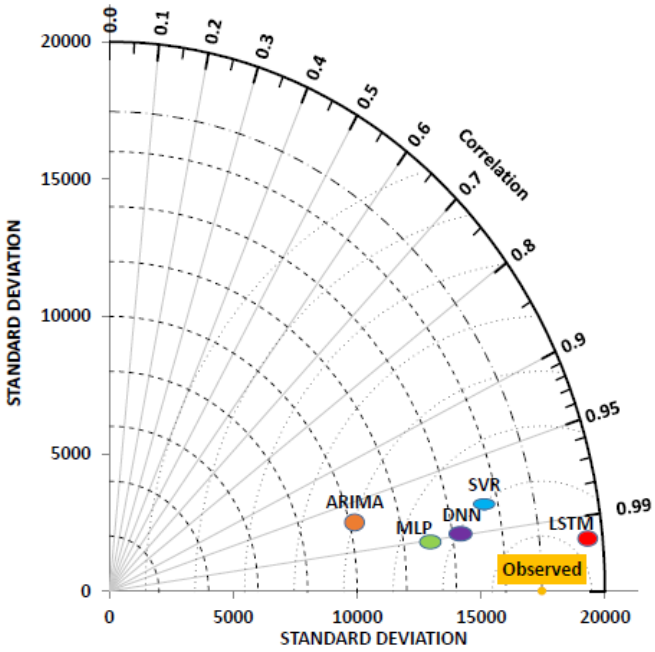

Figure 8. Taylor diagram illustrating the degree of statistical correlation and the root mean square centre difference between the observed and forecasted GSR values by the LSTM model relative to comparative models (i.e., MLP, DNN, ARIMA, and SVR) for the 1M timescale.

Table 8. Diebold–Mariano (DM) test adopted to compare the predictive accuracy of any two forecasting models. The objective model (i.e., LSTM) is compared against counterpart models (ARIMA, MLP, DNN and SVR) in the testing phase. The DM test is used to evaluate if the two forecasts are statistically different with a null hypothesis of no difference rejected if the test statistic falls outside DM = ±1.96 at a 5% level of significance. The statistically significant better performance of LSTM over the other models is indicated as 'yes'.

Diebold–Mariano (DM) Test Statistics Forecast Horizon	LSTM vs. DNN	LSTM vs. ARIMA	LSTM vs. MLP	LSTM vs. SVR
1 Minute (1M)	-	-	-	-
DM statistic	−0.272	−24.381	−25.824	−16.933
p-value	0.785	0.000	0.000	0.000
Reject Null Hypothesis	No	Yes	Yes	Yes
5 Minute (5M)	-	-	-	-
DM statistic	46.585	46.394	−50.779	43.614
p-value	0.000	0.000	0.000	0.000
Reject Null Hypothesis	Yes	Yes	Yes	Yes
10 Minute (10M)	-	-	-	-
DM statistic	62.231	27.318	62.231	32.816
p-value	0.000	0.000	0.000	0.000
Reject Null Hypothesis	Yes	Yes	Yes	Yes
15 Minute (15M)	-	-	-	-
DM statistic	−51.638	−29.999	39.581	−0.268
p-value	0.000	0.000	0.000	0.789
Reject Null Hypothesis	Yes	Yes	Yes	No
Half Hourly (30M)	-	-	-	-
DM statistic	−9.209	18.234	−19.558	17.000
p-value	0.000	0.000	0.000	0.000
Reject Null Hypothesis	Yes	Yes	Yes	Yes

8. Further Discussion, Limitations and Future Scope

Despite the excellent performance of the developed LSTM model, as evaluated by several statistical metrics and visualized model analysis, the proposed model is further evaluated by comparing the results in this study with those of previous studies. In one such study, an LSTM model was developed for 1-hourly day-ahead solar irradiance forecasting on the island of Santiago, Cape Verde. The study employed weather variables (i.e., temperature, dew point, humidity, visibility, wind speed, weather type) for 30 months (March 2011 to August 2012). In concurrence with the present study (Table 5), it was concluded that the LSTM model was the best model as it generated the lowest RMSE in comparison to the persistence, ANN, and linear least squares regression methods (76.245 Wm^{-2}, 209.2509 Wm^{-2}, 230.9857 Wm^{-2}, and 133.5313 Wm^{-2}, respectively).

Another relevant comparative study employed LSTM to estimate hourly and daily GSR in Atlanta, New York, and Hawaii using hourly meteorological data and cloud type information from 2013 to 2017 as a training and testing population. The proposed LSTM demonstrated excellent forecasting performance for hourly forecasts on all-weather (i.e., mixed days and cloudy days). The mean absolute percentage error (MAPE) of LSTM in measured locations (i.e., Atlanta, New York, Hawaii) on cloudy days was 14.9%, 20.1% and 18.1% respectively. All *r*-values of LSTM were above 0.85, outperforming comparative models (i.e., ARIMA, SVR, ANN, CNN, and RNN) with one exceptional case where the *r*-value of RNN was higher than that of LSTM (0.91 and 0.90, respectively). Overall, however, LSTM showed its outstanding performance. The study of Ghimire et al. [44] designed a hybridized framework that integrated a convolutional neural network with LSTM for half-hourly GSR forecasting in Australia; their model was superior, with over 70% of predictive errors lying below ±10 Wm^{-2}. The results from the last two studies are the only close available comparisons of solar forecasting studies employing LSTM. Regarding the evaluation in terms of statistical score metrics, in this study, model-based LSTM outperformed by a noticeable margin, with outstanding *r*, RMSE and MAE (Table 5) in all forecasting horizons (i.e., 1M, 5M, 10M, 15M, and 30M). Moreover, the two compared studies focused solely on a specific forecasting horizon, but this study presented LSTM performance over multiple time horizons in which the *r*-value was over 0.9 for all cases.

In terms of optimization of the LSTM model, an epoch can be set to various times for a given dataset, and is used in the training stage. In this study, the number of epochs (Table 4a) was set to be 2000 in every case. However, the training phase stopped when the evaluation metric MAE did not increase on the validation set, in other words, the number of epochs did not reach 2000 (Table 4b). To allow LSTM to perform better, the number of epochs should be set at a higher value to possibly reach the optimized model. Therefore, the number of training times or epochs does not influence the performance in the training phase since it demonstrates a random property in practice. Moreover, it is noticeable that the LSTM performance improved when the drop rate and batch size increased (Table 4b). This aspect is also a novelty of this study.

To summarize, the newly developed model-based LSTM can be considered to be superior for near-real-time solar forecasting modelling and future solar energy management to the compared previous machine learning methods (i.e., MLP, SVR), deep learning method (i.e., DNN) and time-series method (i.e., ARIMA).

This study supports the significant merits of a deep learning technique to attain better precision in near-real-time solar forecasting. Further, it also demonstrates the ability of models based on LSTM architecture in different forecasting horizons that can assist power generation companies in energy management. Since the *r*-value in very short-term horizons (i.e., 1M, 5M, 10M, 15M, 30M) is quite high, this model can be applied elsewhere with similar climatic conditions to Vietnam. However, the scope of this study could be further improved as it is restricted in terms of the forecasting horizon.

Further studies can study LSTM's ability in longer-term forecasting horizons, such as medium term (i.e., hourly, daily) and long term (i.e., weekly, monthly, yearly) to support specific application purposes. Moreover, further studies can also apply feature extraction and feature selection to develop a hybrid LSTM model. Since this study focuses on GSR from ground-based measurement, further

study can apply LSTM in the context of multiple weather variables or satellite-derived variables in multiple weather conditions (i.e., so that the LSTM capability can be thoroughly explored.

9. Conclusions

For the first time in this study region, this paper developed and demonstrated a forecasting model-based LSTM algorithm for a near-real-time horizon using only global solar radiation times series, which is also an alternative approach for those circumstances where there is a lack of available predictor variables. The model was evaluated over multiple time horizons utilizing antecedent lagged global solar radiation (GSR) data from 2017 to 2019 in Vietnam. Moreover, several types of evaluation metrics were employed to assess the performance of the forecasting models, from which it was shown that the LSTM model yielded the most accurate results.

The LSTM models were optimized by the combination of hyperparameters (Table 3a) and were then compared to the optimized counterpart models. Evidently, the performance of the LSTM models were better in all cases, and the LSTM model was found to be superior compared to its counterparts at a 1 min horizon (Tables 5 and 6) as evidenced by its low relative forecasting errors (i.e., RMSE = 40.9125 Wm^{-2} and MAE = 21.6428 Wm^{-2}) and high performance metrics (i.e., r = 0.9920, WI = 0.9984, and E_{NS} = 0.9831).

By assessing the performance of the LSTM model utilizing the Legates–McCabe (LM) metric, the LSTM model was found to have the highest agreement. The obtained LM performance between forecasted and observed global solar radiation values for various time horizons (1M, 5M, 10M, 15M, and 30M) were 0.9204, 0.4658, 0.9275, 0.7575, and 0.7741 respectively, whereas the relative percentage errors (RRMSE) were approximately 10, 15, 11, 13, and 14%, respectively.

In addition, the DM test was employed to provide an evaluation framework for different models and to provide a strict criterion to evaluate the forecasting accuracy. A meaningful evaluation conclusion of superior performance of LSTM over the counterpart models was reached when most absolute DM statistic values were greater than 1.96 at a 5% level of significance. The only two exceptions were those of LSTM vs. DNN at a 1M forecast horizon ($|DM| = 0.272 < 1.96$) and LSTM vs. SVR at a 15M forecast horizon ($|DM| = 0.268 < 1.96$) at a 5% level of significance. Moreover, the p-values at a 5% level of significance were less than 0.05, which means all models were statistically significant.

In short, this study provides a baseline investigation that is relevant to other potential models to be used in near-real-time solar forecasting in future studies. Examples include the hybridization of LSTM with other methods, such as a convolutional neural network for feature mapping, using a wrapper-based feature selection method employing several atmospheric predictor variables [42], other deep learning methods (e.g., deep neural networks), and data decomposition methods such as wavelets and ensemble mode decompositions [10,127]. While these methods can potentially improve the proposed LSTM model, the present study, as a first investigation of the near real-time forecasting of GSR, has set a clear future foundation for adopting these techniques in the context of solar radiation modelling in Vietnam. Nonetheless, the findings of this study ascertain that the standalone LSTM model could adequately capture the nonlinear dynamics of global solar radiation time-series data. The model-based LSTM can be employed in longer time horizon solar forecasting (i.e., long term, medium term, and short term). Furthermore, the government and electricity generator companies in Vietnam can use this model prior to generating solar energy to derive an optimal production strategy.

Author Contributions: Conceptualization, A.N.-L.H.; Data curation, A.N.-L.H.; Formal analysis, A.N.-L.H.; Investigation, A.N.-L.H.; Methodology, A.N.-L.H.; Software, A.N.-L.H.; Supervision, R.C.D., N.R., M.A. and S.A.; Validation, A.N.-L.H., R.C.D.; Visualization, A.N.-L.H. and R.C.D.; Writing—original draft, A.N.-L.H.; Writing—review & editing, A.N.-L.H., R.C.D., D.-A.A.-V., M.A., N.R. and S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The paper utilized minute-level solar radiation data from the World Bank, which are duly acknowledged. We also would like to thank Barbara Harmes (Language Centre, Open Access College,

University of Southern Queensland, Australia) for providing help in proof-reading this paper. Finally, we thank both reviewers for their constructive comments that improved the clarity of the final paper.

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

Abbreviations

ACF	Autocorrelation	LSTM	Long Short-Term Memory
AR	Autoregressive	MA	Moving Average
ARMA	Autoregressive Moving Average	MAE	Mean Absolute Error
ARIMA	Autoregressive Integrated Moving Average	MAPE	Mean Absolute Percentage Error
FFNN	Feed Forward Neural Networks.	MSE	Mean Squared Error
CPU	Central Processing Unit	MLP	Multilayer Perceptron Network
d	Degree of differencing in ARIMA	PACF	Partial Auto-Correlation Function
DL	Deep Learning	p	Autoregressive term in ARIMA
DNN	Deep Neuron Network	$\lvert FE \rvert$	Absolute Forecasted Error
GSR	Global Solar Radiation	q	Moving average term in ARIMA
GSR_{ACTUAL}	Actual Global Solar Radiation	r	Pearson's Correlation Coefficient
GSR_{N}	Normalised Global Solar Radiation	R^2	Coefficient of determination
GSR_{MAX}	Maximum value of Global Solar Radiation	ReLU	Rectified Linear Unit
GSR_{MIN}	Minimum value of Global Solar Radiation	RMSE	Root Mean Square Error
GSR_{OBS}	Observed Global Solar Radiation	RRMSE	Relative Root Mean Square Error
$<GSR_{OBS}>$	Average value of Observed Global Solar Radiation	RNN	Recurrent Neural Networks
GSR_{FOR}	Forecasted Global Solar Radiation	N	Number of values in a data series
$<GSR_{FOR}>$	Average value of Forecasted Global Solar Radiation	SVR	Support vector regression
E_{NS}	Nash-Sutcliffe Efficiency	BPNN	Back-Propagation Neural Networks
LM	Legate & McCabe's Index	ELM	Extreme Learning Machine
CNN	Convolutional Neural Network	ANN	Artificial Neural Network
NARX	Nonlinear autoregressive network with exogenous inputs	B-ELM	Bayesian extreme learning machine
RBF	Radial Basis Function	RW	Rescorla–Wagner
ARIMAX	ARIMA with exogenous variables	ES	Evolution strategy

References

1. Färe, R.; Grosskopf, S.; Tyteca, D. An activity analysis model of the environmental performance of firms—Application to fossil-fuel-fired electric utilities. *Ecol. Econ.* **1996**, *18*, 161–175.
2. Agarwal, A.K. Biofuels (alcohols and biodiesel) applications as fuels for internal combustion engines. *Prog. Energy Combust. Sci.* **2007**, *33*, 233–271.
3. Ezra, D. *Coal and Energy: The Need to Exploit the World's Most Abundant Fossil Fuel*; Wiley: Hoboken, New Jersey, USA, 1978.
4. Amponsah, N.Y.; Troldborg, M.; Kington, B.; Aalders, I.; Hough, R.L. Greenhouse gas emissions from renewable energy sources: A review of lifecycle considerations. *Renew. Sustain. Energy Rev.* **2014**, *39*, 461–475.

5. Meinel, A.B.; Meinel, M.P. Applied solar energy: An introduction. *NASA STI/Recon Tech. Rep. A* **1977**, *77*, 33445.
6. Luong, N.D. A critical review on energy efficiency and conservation policies and programs in Vietnam. *Renew. Sustain. Energy Rev.* **2015**, *52*, 623–634.
7. IEA. *World Energy Outlook 2016 Executive Summary*; International Energy Agency: Paris, France, 2012.
8. Shem, C.; Simsek, Y.; Hutfilter, U.F.; Urmee, T. Potentials and opportunities for low carbon energy transition in Vietnam: A policy analysis. *Energy Policy* **2019**, *134*, 110818.
9. Polo, J.; Bernardos, A.; Martínez, S.; Peruchena, C.F. *Maps of Solar Resource and Potential in Vietnam*; Ministry Of Industry and Trade of the Socialist Republic of Vietnam: Hanoi, Vietnam, 2015.
10. Al-Musaylh, M.S.; Deo, R.C.; Li, Y.; Adamowski, J.F. Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting. *Appl. Energy* **2018**, *217*, 422–439.
11. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.-L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* **2017**, *105*, 569–582.
12. Qin, J.; Chen, Z.; Yang, K.; Liang, S.; Tang, W. Estimation of monthly-mean daily global solar radiation based on MODIS and TRMM products. *Appl. Energy* **2011**, *88*, 2480–2489.
13. Yang, H.-T.; Huang, C.-M.; Huang, Y.-C.; Pai, Y.-S. A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output. *IEEE Trans. Sustain. Energy* **2014**, *5*, 917–926.
14. Pierro, M.; Bucci, F.; Cornaro, C.; Maggioni, E.; Perotto, A.; Pravettoni, M.; Spada, F. Model output statistics cascade to improve day ahead solar irradiance forecast. *Sol. Energy* **2015**, *117*, 99–113.
15. Qing, X.; Niu, Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy* **2018**, *148*, 461–468.
16. Mellit, A.; Pavan, A.M. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. *Sol. Energy* **2010**, *84*, 807–821.
17. Alzahrani, A.; Shamsi, P.; Dagli, C.; Ferdowsi, M. Solar irradiance forecasting using deep neural networks. *Procedia Comput. Sci.* **2017**, *114*, 304–313.
18. Wang, H.; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. A review of deep learning for renewable energy forecasting. *Energy Convers. Manag.* **2019**, *198*, 111799.
19. Paulescu, E.; Blaga, R. Regression models for hourly diffuse solar radiation. *Sol. Energy* **2016**, *125*, 111–124.
20. Coulibaly, O.; Ouedraogo, A. Correlation of global solar radiation of eight synoptic stations in Burkina Faso based on linear and multiple linear regression methods. *J. Sol. Energy* **2016**, *2016*, 9.
21. Benmouiza, K.; Cheknane, A. Small-scale solar radiation forecasting using ARMA and nonlinear autoregressive neural network models. *Theor. Appl. Climatol.* **2016**, *124*, 945–958.
22. Sfetsos, A.; Coonick, A. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. *Sol. Energy* **2000**, *68*, 169–178.
23. Hocaoğlu, F.O. Stochastic approach for daily solar radiation modeling. *Sol. Energy* **2011**, *85*, 278–287.
24. Rigler, E.; Baker, D.; Weigel, R.; Vassiliadis, D.; Klimas, A. Adaptive linear prediction of radiation belt electrons using the Kalman filter. *Space Weather* **2004**, *2*, 1–9.
25. Bracale, A.; Caramia, P.; Carpinelli, G.; Di Fazio, A.; Ferruzzi, G. A Bayesian method for short-term probabilistic forecasting of photovoltaic generation in smart grid operation and control. *Energies* **2013**, *6*, 733–747.
26. Ming, D.; Ningzhou, X. A method to forecast short-term output power of photovoltaic generation system based on Markov chain. *Power Syst. Technol.* **2011**, *35*, 152–157.
27. Wenbin, H.; Ben, H.; Changzhi, Y. Building thermal process analysis with grey system method. *Build. Environ.* **2002**, *37*, 599–605.
28. Ruiz-Arias, J.; Alsamamra, H.; Tovar-Pescador, J.; Pozo-Vázquez, D. Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. *Energy Convers. Manag.* **2010**, *51*, 881–893.
29. Martín, L.; Zarzalejo, L.F.; Polo, J.; Navarro, A.; Marchante, R.; Cony, M. Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning. *Sol. Energy* **2010**, *84*, 1772–1781.
30. Moreno-Munoz, A.; De la Rosa, J.; Posadillo, R.; Bellido, F. Very short term forecasting of solar radiation. In Proceedings of the 2008 33rd IEEE Photovoltaic Specialists Conference, San Diego, CA, USA, 11–16 May 2008; pp. 1–5.

31. Colak, I.; Yesilbudak, M.; Genc, N.; Bayindir, R. Multi-period prediction of solar radiation using ARMA and ARIMA models. In Proceedings of the 2015 IEEE 14th international conference on machine learning and applications (ICMLA), Miami, FL, USA, 9–11 December 2015; pp. 1045–1049.
32. Rao, K.S.K.; Rani, B.I.; Ilango, G.S. Estimation of daily global solar radiation using temperature, relative humidity and seasons with ANN for Indian stations. In Proceedings of the 2012 International Conference on Power, Signals, Controls and Computation, Thrissur, Kerala, India, 3–6 January 2012; pp. 1–6.
33. Kaplani, E.; Kaplanis, S. A stochastic simulation model for reliable PV system sizing providing for solar radiation fluctuations. *Appl. Energy* **2012**, *97*, 970–981.
34. Lauret, P.; Boland, J.; Ridley, B. Bayesian statistical analysis applied to solar radiation modelling. *Renew. Energy* **2013**, *49*, 124–127.
35. Ramedani, Z.; Omid, M.; Keyhani, A.; Shamshirband, S.; Khoshnevisan, B. Potential of radial basis function based support vector regression for global solar radiation prediction. *Renew. Sustain. Energy Rev.* **2014**, *39*, 1005–1011.
36. Nguyen, B.; Pryor, T. A computer model to estimate solar radiation in Vietnam. *Renew. Energy* **1996**, *9*, 1274–1278.
37. Nguyen, B.T.; Pryor, T.L. The relationship between global solar radiation and sunshine duration in Vietnam. *Renew. Energy* **1997**, *11*, 47–60.
38. Polo, J.; Gastón, M.; Vindel, J.; Pagola, I. Spatial variability and clustering of global solar irradiation in Vietnam from sunshine duration measurements. *Renew. Sustain. Energy Rev.* **2015**, *42*, 1326–1334.
39. Polo, J.; Bernardos, A.; Navarro, A.; Fernandez-Peruchena, C.; Ramírez, L.; Guisado, M.V.; Martínez, S. Solar resources and power potential mapping in Vietnam using satellite-derived and GIS-based information. *Energy Convers. Manag.* **2015**, *98*, 348–358.
40. Qin, W.; Wang, L.; Lin, A.; Zhang, M.; Xia, X.; Hu, B.; Niu, Z. Comparison of deterministic and data-driven models for solar radiation estimation in China. *Renew. Sustain. Energy Rev.* **2018**, *81*, 579–594.
41. Wang, L.; Kisi, O.; Zounemat-Kermani, M.; Zhu, Z.; Gong, W.; Niu, Z.; Liu, H.; Liu, Z. Prediction of solar radiation in China using different adaptive neuro-fuzzy methods and M5 model tree. *Int. J. Climatol.* **2017**, *37*, 1141–1155.
42. Salcedo-Sanz, S.; Deo, R.C.; Cornejo-Bueno, L.; Camacho-Gómez, C.; Ghimire, S. An efficient neuro-evolutionary hybrid modelling mechanism for the estimation of daily global solar radiation in the Sunshine State of Australia. *Appl. Energy* **2018**, *209*, 79–94.
43. Kabir, E.; Kumar, P.; Kumar, S.; Adelodun, A.A.; Kim, K.-H. Solar energy: Potential and future prospects. *Renew. Sustain. Energy Rev.* **2018**, *82*, 894–900.
44. Ghimire, S.; Deo, R.C.; Raj, N.; Mi, J. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Appl. Energy* **2019**, *253*, 113541.
45. Liu, W.; Wang, Z.; Liu, X.; Zeng, N.; Liu, Y.; Alsaadi, F.E. A survey of deep neural network architectures and their applications. *Neurocomputing* **2017**, *234*, 11–26.
46. Ryu, A.; Ito, M.; Ishii, H.; Hayashi, Y. Preliminary Analysis of Short-term Solar Irradiance Forecasting by using Total-sky Imager and Convolutional Neural Network. In Proceedings of the 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia), Bangkok, Thailand, 19–23 March 2019; pp. 627–631.
47. Manohar, M.; Koley, E.; Ghosh, S.; Mohanta, D.K.; Bansal, R. Spatio-temporal information based protection scheme for PV integrated microgrid under solar irradiance intermittency using deep convolutional neural network. *Int. J. Electr. Power Energy Syst.* **2020**, *116*, 105576.
48. Zang, H.; Cheng, L.; Ding, T.; Cheung, K.W.; Liang, Z.; Wei, Z.; Sun, G. Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network. *IET Gener. Transm. Distrib.* **2018**, *12*, 4557–4567.
49. Wang, F.; Zhang, Z.; Liu, C.; Yu, Y.; Pang, S.; Duić, N.; Shafie-Khah, M.; Catalao, J.P. Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting. *Energy Convers. Manag.* **2019**, *181*, 443–462.
50. Awan, S.M.; Khan, Z.A.; Aslam, M. Solar Generation Forecasting by Recurrent Neural Networks Optimized by Levenberg-Marquardt Algorithm. In Proceedings of the IECON 2018—44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, USA, 21–23 October 2018; pp. 276–281.

51. Mishra, S.; Palanisamy, P. Multi-time-horizon Solar Forecasting Using Recurrent Neural Network. In Proceedings of the 2018 IEEE Energy Conversion Congress and Exposition (ECCE), Portland, OR, USA, 23–27 September 2018; pp. 18–24.
52. Wang, M.; Zang, H.; Cheng, L.; Wei, Z.; Sun, G. Application of DBN for estimating daily solar radiation on horizontal surfaces in Lhasa, China. *Energy Procedia* **2019**, *158*, 49–54.
53. Wang, K.; Qi, X.; Liu, H. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Appl. Energy* **2019**, *251*, 113315.
54. Caballero, R.; Zarzalejo, L.F.; Otero, Á.; Piñuel, L.; Wilbert, S. Short term cloud nowcasting for a solar power plant based on irradiance historical data. *J. Comput. Sci. Technol.* **2018**, *18*, 186–192.
55. Yu, Y.; Cao, J.; Zhu, J. An LSTM Short-Term Solar Irradiance Forecasting Under Complicated Weather Conditions. *IEEE Access* **2019**, *7*, 145651–145666.
56. Mukherjee, A.; Ain, A.; Dasgupta, P. Solar Irradiance Prediction from Historical Trends Using Deep Neural Networks. In Proceedings of the 2018 IEEE International Conference on Smart Energy Grid Engineering (SEGE), Oshawa, ON, Canada, 12–15 August 2018; pp. 356–361.
57. Lee, H.; Lee, B.-T. Bayesian deep learning-based confidence-aware solar irradiance forecasting system. In Proceedings of the 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea, 17–19 October 2018; pp. 1233–1238.
58. Zhou, H.; Zhang, Y.; Yang, L.; Liu, Q.; Yan, K.; Du, Y. Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism. *IEEE Access* **2019**, *7*, 78063–78074.
59. Gensler, A.; Henze, J.; Sick, B.; Raabe, N. Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks. In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 October 2016; pp. 002858–002865.
60. Wang, F.; Yu, Y.; Zhang, Z.; Li, J.; Zhen, Z.; Li, K. Wavelet decomposition and convolutional LSTM networks based improved deep learning model for solar irradiance forecasting. *Appl. Sci.* **2018**, *8*, 1286.
61. Muhammad, A.; Lee, J.M.; Hong, S.W.; Lee, S.J.; Lee, E.H. Deep Learning Application in Power System with a Case Study on Solar Irradiation Forecasting. In Proceedings of the 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Okinawa, Japan, 11–13 February 2019; pp. 275–279.
62. Siddiqui, T.A.; Bharadwaj, S.; Kalyanaraman, S. A deep learning approach to solar-irradiance forecasting in sky-videos. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 7–11 January 2019; pp. 2166–2174.
63. Lee, W.; Kim, K.; Park, J.; Kim, J.; Kim, Y. Forecasting solar power using long-short term memory and convolutional neural networks. *IEEE Access* **2018**, *6*, 73068–73080.
64. Srivastava, S.; Lessmann, S. A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. *Sol. Energy* **2018**, *162*, 232–247.
65. Gao, M.; Li, J.; Hong, F.; Long, D. Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM. *Energy* **2019**, *187*, 115838.
66. Wang, Y.; Shen, Y.; Mao, S.; Chen, X.; Zou, H. LASSO and LSTM integrated temporal model for short-term solar intensity forecasting. *IEEE Internet Things J.* **2018**, *6*, 2933–2944.
67. Zaouali, K.; Rekik, R.; Bouallegue, R. Deep learning forecasting based on auto-LSTM model for Home Solar Power Systems. In Proceedings of the 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Exeter, UK, 28–30 June 2018; pp. 235–242.
68. Deo, R.C.; Şahin, M.; Adamowski, J.F.; Mi, J. Universally deployable extreme learning machines integrated with remotely sensed MODIS satellite predictors over Australia to forecast global solar radiation: A new approach. *Renew. Sustain. Energy Rev.* **2019**, *104*, 235–261.
69. Mohammadi, K.; Shamshirband, S.; Tong, C.W.; Arif, M.; Petković, D.; Ch, S. A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. *Energy Convers. Manag.* **2015**, *92*, 162–171.
70. Ghimire, S.; Deo, R.C.; Raj, N.; Mi, J. Deep learning neural networks trained with MODIS satellite-derived predictors for long-term global solar radiation prediction. *Energies* **2019**, *12*, 2407.
71. Geurts, M. Time Series Analysis: Forecasting and Control. *JMR J. Mark. Res. (pre-1986)* **1977**, *14*, 269.

72. Ballabio, D.; Consonni, V.; Todeschini, R. The Kohonen and CP-ANN toolbox: A collection of MATLAB modules for self organizing maps and counterpropagation artificial neural networks. *Chemom. Intell. Lab. Syst.* **2009**, *98*, 115–122.
73. Hui, C.L.P. *Artificial Neural Networks: Application*; IntechOpen: London, UK, 2011.
74. Martinez-Anido, C.B.; Botor, B.; Florita, A.R.; Draxl, C.; Lu, S.; Hamann, H.F.; Hodge, B.-M. The value of day-ahead solar power forecasting improvement. *Sol. Energy* **2016**, *129*, 192–203.
75. Diagne, M., David, M., Lauret, P., Boland, J., Schmutz, N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renew. Sustain. Energy Rev*, **2013**. *27*, 65–76.
76. Inman, R.H.; Pedro, H.T.; Coimbra, C.F. Solar forecasting methods for renewable energy integration. *Prog. Energy Combust. Sci.* **2013**, *39*, 535–576.
77. Raza, M.Q.; Nadarajah, M.; Ekanayake, C. On recent advances in PV output power forecast. *Sol. Energy* **2016**, *136*, 125–144.
78. Sivaneasan, B.; Yu, C.; Goh, K. Solar forecasting using ANN with fuzzy logic pre-processing. *Energy Procedia* **2017**, *143*, 727–732.
79. Golestaneh, F.; Pinson, P.; Gooi, H.B. Very short-term nonparametric probabilistic forecasting of renewable energy generation—With application to solar energy. *IEEE Trans. Power Syst.* **2016**, *31*, 3850–3863.
80. Sun, Y.; Venugopal, V.; Brandt, A.R. Convolutional neural network for short-term solar panel output prediction. In Proceedings of the 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)(A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC), Waikoloa Village, HI, USA, 10–15 June 2018; pp. 2357–2361.
81. Khelifi, R.; Guermoui, M.; Rabehi, A.; Lalmi, D. Multi-step-ahead forecasting of daily solar radiation components in the Saharan climate. *Int. J. Ambient Energy* **2020**, *41*, 707–715.
82. Paulescu, M.; Paulescu, E. Short-term forecasting of solar irradiance. *Renew. Energy* **2019**, *143*, 985–994.
83. Vaz, A.; Elsinga, B.; Van Sark, W.; Brito, M. An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht, the Netherlands. *Renew. Energy* **2016**, *85*, 631–641.
84. Nobre, A.M.; Severiano, C.A., Jr.; Karthik, S.; Kubis, M.; Zhao, L.; Martins, F.R.; Pereira, E.B.; Rüther, R.; Reindl, T. PV power conversion and short-term forecasting in a tropical, densely-built environment in Singapore. *Renew. Energy* **2016**, *94*, 496–509.
85. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT press: Cambridge, Massachusetts, USA, 2016.
86. Jieni, X.; Zhongke, S. Short-time traffic flow prediction based on chaos time series theory. *J. Transp. Syst. Eng. Inf. Technol.* **2008**, *8*, 68–72.
87. Hand, D.J. Classifier technology and the illusion of progress. *Stat. Sci.* **2006**, *21*, 1–14.
88. Kubat, M. Neural networks: A comprehensive foundation by Simon Haykin, Macmillan, 1994, ISBN 0-02-352781-7. *Knowl. Eng. Rev.* **1999**, *13*, 409–412.
89. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166.
90. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780.
91. Yang, J.; Kim, J. An accident diagnosis algorithm using long short-term memory. *Nucl. Eng. Technol.* **2018**, *50*, 582–588.
92. Gers, F.A.; Schraudolph, N.N.; Schmidhuber, J. Learning precise timing with LSTM recurrent networks. *J. Mach. Learn. Res.* **2002**, *3*, 115–143.
93. Box, G.E.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control, Revised Edition*; Holden Day: San Francisco, CA, USA, 1976.
94. Al-Musaylh, M.S.; Deo, R.C.; Adamowski, J.F.; Li, Y. Short-term electricity demand forecasting using machine learning methods enriched with ground-based climate and ECMWF Reanalysis atmospheric predictors in southeast Queensland, Australia. *Renew. Sustain. Energy Rev.* **2019**, *113*, 109293.
95. Deo, R.C.; Şahin, M. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. *Renew. Sustain. Energy Rev.* **2017**, *72*, 828–848.
96. Fentis, A.; Bahatti, L.; Mestari, M.; Chouri, B. Short-term solar power forecasting using Support Vector Regression and feed-forward NN. In Proceedings of the 2017 15th IEEE International New Circuits and Systems Conference (NEWCAS), Strasbourg, France, 25–28 June 2017; pp. 405–408.

97. Alfadda, A.; Adhikari, R.; Kuzlu, M.; Rahman, S. Hour-ahead solar PV power forecasting using SVR based approach. In Proceedings of the 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 23–26 April 2017; pp. 1–5.
98. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.J.; Vapnik, V. Support vector regression machines. In Proceedings of the Advances in Neural Information Processing Systems 9, Denver, CO, USA, 3–5 December 1996; pp. 155–161.
99. Díaz–Vico, D.; Torres–Barrán, A.; Omari, A.; Dorronsoro, J.R. Deep neural networks for wind and solar energy prediction. *Neural Process. Lett.* **2017**, *46*, 829–844.
100. Werbos, P. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Ph. D. dissertation, Harvard University, Cambridge, MA, USA, 1974.
101. Schmidt-Thomé, P.; Nguyen, T.H.; Pham, T.L.; Jarva, J.; Nuottimäki, K. *Climate Change Adaptation Measures in Vietnam: Development and Implementation*; Springer: Berlin/Heidelberg, Germany, 2014.
102. Nguyen, T.N.; Wongsurawat, W. Multivariate cointegration and causality between electricity consumption, economic growth, foreign direct investment and exports: Recent evidence from Vietnam. *Int. J. Energy Econ. Policy* **2017**, *7*, 287–293.
103. Kies, A.; Schyska, B.; Viet, D.T.; von Bremen, L.; Heinemann, D.; Schramm, S. Large-scale integration of renewable power sources into the Vietnamese power system. *Energy Procedia* **2017**, *125*, 207–213.
104. Ghimire, S.; Deo, R.C.; Downs, N.J.; Raj, N. Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cities of Queensland Australia. *J. Clean. Prod.* **2019**, *216*, 288–310.
105. Kotsiantis, S.; Kanellopoulos, D.; Pintelas, P. Data preprocessing for supervised leaning. *Int. J. Comput. Sci.* **2006**, *1*, 111–117.
106. Keras-team, K.D. The Python Deep Learning Library. Available online: https://keras.io (accessed on 5 May 2019).
107. Nawi, N.; Atomi, W.; Rehman, M. The effect of data pre-processing on optimized training of artificial neural networks. *Procedia Technol.* **2013**, *11*, 32–39.
108. Yu, R.; Gao, J.; Yu, M.; Lu, W.; Xu, T.; Zhao, M.; Zhang, J.; Zhang, R.; Zhang, Z. LSTM-EFG for wind power forecasting based on sequential correlation features. *Future Gener. Comput. Syst.* **2019**, *93*, 33–42.
109. Steyerberg, E.W.; Vickers, A.J.; Cook, N.R.; Gerds, T.; Gonen, M.; Obuchowski, N.; Pencina, M.J.; Kattan, M.W. Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology (Cambridge Mass.)* **2010**, *21*, 128.
110. Tian, Y.; Nearing, G.S.; Peters-Lidard, C.D.; Harrison, K.W.; Tang, L. Performance metrics, error modeling, and uncertainty quantification. *Mon. Weather Rev.* **2016**, *144*, 607–613.
111. Willmott, C.J.; Matsuura, K.; Robeson, S.M. Ambiguities inherent in sums-of-squares-based error statistics. *Atmos. Environ.* **2009**, *43*, 749–752.
112. Willmott, C.J. On the evaluation of model performance in physical geography. In *Spatial Statistics and Models*; Springer: Dordrecht, The Netherlands, 1984; pp. 443–460.
113. Wilcox, B.P.; Rawls, W.; Brakensiek, D.; Wight, J.R. Predicting runoff from rangeland catchments: A comparison of two models. *Water Resour. Res.* **1990**, *26*, 2401–2410.
114. Legates, D.R.; McCabe, G.J., Jr. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **1999**, *35*, 233–241.
115. Garrick, M.; Cunnane, C.; Nash, J. A criterion of efficiency for rainfall-runoff models. *J. Hydrol.* **1978**, *36*, 375–381.
116. Willmott, C.J.; Robeson, S.M.; Matsuura, K. A refined index of model performance. *Int. J. Climatol.* **2012**, *32*, 2088–2094.
117. Ertekin, C.; Yaldiz, O. Comparison of some existing models for estimating global solar radiation for Antalya (Turkey). *Energy Convers. Manag.* **2000**, *41*, 311–330.
118. Tayman, J.; Swanson, D.A. On the validity of MAPE as a measure of population forecast accuracy. *Popul. Res. Policy Rev.* **1999**, *18*, 299–322.
119. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4.
120. Faber, N.K.M. Estimating the uncertainty in estimates of root mean square error of prediction: Application to determining the size of an adequate test set in multivariate calibration. *Chemom. Intell. Lab. Syst.* **1999**, *49*, 79–89.

121. Coyle, E.J.; Lin, J.-H. Stack filters and the mean absolute error criterion. *IEEE Trans. Acoust. Speechand Signal Process.* **1988**, *36*, 1244–1254.
122. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **2002**, *20*, 134–144.
123. Chen, H.; Wan, Q.; Wang, Y. Refined Diebold-Mariano test methods for the evaluation of wind power forecasting models. *Energies* **2014**, *7*, 4185–4198.
124. Diebold, F.X. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *J. Bus. Econ. Stat.* **2015**, *33*:1, 1. doi: 10.1080/07350015.2014.983236
125. Makridakis, S.; Wheelwright, S.C.; Hyndman, R.J. *Forecasting Methods and Applications*; John Wiley & Sons: Hoboken, New Jersey, USA, 2008.
126. Kim, S.; Kim, H. A new metric of absolute percentage error for intermittent demand forecasts. *Int. J. Forecast.* **2016**, *32*, 669–679.
127. Deo, R.C.; Wen, X.; Feng, Q. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Appl. Energy* **2016**, *168*, 568–593.

CHAPTER 5: SHORT-TERM HORIZON SOLAR RADIATION FORECASTING

Article 2: Novel short-term solar radiation hybrid model based on Long Short-Term Memory Networks integrated with Robust Local Mean Decomposition

5.1 Foreword

Due to the non-stationary property of solar radiation, traditional methods might find it hard to obtain an accurate and reliable solar prediction and solar energy generation. Therefore, a data-driven model tailored for short-term forecasting can generate meaningful information for various usages. The second paper introduces a new artificial intelligence hybrid model by employing a robust version of the local mean decomposition (RLMD) and Long Short-Term Memory (LSTM) networks, denoted as RLMD-LSTM, trained and evaluated on half-hourly ground-based solar radiation (GSR) from 2017 to 2019 for four solar-rich metropolitan cities in Vietnam, namely Da-Nang, Central Highlands, Song-Binh and Tri-An.

RLMD is implemented robustly to decompose data inputs from a complex set of predictive variables, whereas LSTM is employed for prediction with the results benchmarked by classical approaches (*i.e.* Support Vector Regression (SVR), Long Short-Term Memory (LSTM), Multivariate Adaptive Regression Spline (MARS)) as well as other alternative methods (*i.e.* RLMD-MARS, and RLMD-SVR). Verified by statistical and visual infographics, the results demonstrate that the proposed RLMD-LSTM hybrid model can generate a satisfactory GSR prediction and can outperform the counterpart's predictive modelling methods. The outstanding performance of the RLMD-LSTM hybrid model ascertains that this newly designed approach has the potential to be applied in real-time energy management and operation systems.

5.2 Published article II

Novel short-term solar radiation hybrid model: Long Short-Term Memory Network integrated with Robust Local Mean Decomposition

Anh Ngoc-Lan Huynh [a, *], Ravinesh C. Deo [a, *], Mumtaz Ali [b], Shahab Abdulla [c] and Nawin Raj [b]

[a] School of Sciences, University of Southern Queensland, Queensland 4350, Australia; ravinesh.deo@usq.edu.au, nawin.raj@usq.edu.au

[b] Deakin-SWU Joint Research Centre on Big Data, School of Information Technology, Deakin University, Burwood VIC 2134, Australia; mumtaz.ali@deakin.edu.au

[c] USQ College, University of Southern Queensland, Queensland 4500, Australia; Shahab.Abdulla@usq.edu.au

* Corresponding Authors: anh.huynh@usq.edu.au (A.N.-L.H.); ravinesh.deo@usq.edu.au (RC Deo)

Abstract

Data-intelligent algorithms tailored for short-term energy forecasts generate meaningful information on future variability of renewable energy, supply, demand and price management, and the harvesting of solar energy in solar concentrators and photovoltaic systems. Traditional forecasting methods find it relatively difficult to obtain reliable prediction of solar energy generation because of the inherent nonlinearities in solar radiation and related atmospheric variables. This paper proposes a new artificial intelligence hybrid model by employing robust version of local mean decomposition (RLMD) and Long Short-term Memory (LSTM) network denoted as RLMD-LSTM. This objective model is trained and evaluated on near real-time, half-hourly ground-based solar radiation (*GSR*) dataset for solar rich, metropolitan sites in Vietnam. The RLMD-based data decomposition algorithm is firstly implemented robustly to decompose *GSR* data to reveal historical behaviour of solar radiation, represented by product functions, whereas LSTM is employed for predictive modelling purposes, with all results benchmarked through classical modelling approaches (*i.e.*, Support Vector Regression SVR, Long Short-term Memory LSTM, Multivariate Adaptive Regression Spline MARS) as well as the other alternative hybrid methods (*i.e.*, RLMD-MARS, and RLMD-SVR). Verified by statistical metrics and visual infographics, the present results demonstrate that the proposed LSTM-RLMD hybrid model can generate satisfactory *GSR* predictions, outperforming several counterpart methods. The outstanding performance of RLMD-LSTM hybrid model ascertains that the newly designed approach can be a potential candidate for real-time energy management, renewable energy integration into a power grid and other decisions to optimise the scheduling and performance of the overall system.

Keywords Short-term solar radiation prediction, Robust Local Mean Decomposition, Deep Learning

Acronyms

PV	Photovoltaic solar	*MSE*	Mean Squared Error
DWT	Discrete wavelet transform	MARS	Multivariate adaptive regression spline
RLMD	Robust Local Mean Decomposition	PACF	Partial autocorrelation function
CNN-LSTM	Convolutional Long Short-term memory	E_{NS}	Nash-Sutcliffe Efficiency
PF	Product Function	*LM*	Legate & McCabe's Index
$< \mathrm{GSR}_{\mathrm{FOR}} >$	Average value of Forecasted value	LMD	Local Mean Decomposition
GSR_{N}	Normalised value	*RMSE*	Root Mean Square Error
GSR_{MAX}	Maximum value	NWP	Numerical weather predictions
GSR_{MIN}	Minimum value	GSR_{ACTUAL}	Actual value
R^2	Coefficient of determination	*RRMSE*	Relative Root Mean Square Error
$< \mathrm{GSR}_{\mathrm{OBS}} >$	Average value of Observed value	*GSR*	Global Solar Radiation
$\mathrm{GSR}_{\mathrm{FOR}}$	Forecasted value	LSTM	Long Short-term Memory
EMD	Empirical model decomposition	SVR	Support Vector Regression
MAE	Mean Absolute Error	$\lvert FE \rvert$	Absolute Forecasted Error
GSR_{OBS}	Observed value		

1.0 Introduction

Vietnam is a developing nation with an annual average gross domestic product of ~7% in which the energy sector accounts for a vital part of its economy [1, 2]. In general, both the primary and the secondary energy demand has experienced a rise of over 5% per year with significant growth coming from coal and oil where an acceptable portion of renewables were largely negligible. Due to growing energy demand, Vietnam has a strong reliance on foreign imported energy (*e.g.*, coal and oil) which, in the current time of economic instability, can put national energy security at severe risk. Moreover, fossil fuels are causing environmental risks and energy crisis that pose a significant threat to our lives

[3], hence, the exploitation of renewable energy resources have attracted public controversy in pursuit of environment protection [4]. In fact, renewable energy has been considered because of its inherent nature that avoids the passage of contaminants (*e.g.*, CO_2) into the natural environment [3] In 2017, the capacity of solar PV power was almost 400 GW, and that for the production volume was ~2% of global power [5]. There has been an increasing attention on United Nations Sustainable Development goals (SDG) [6] recommending for cleaner energy resources globally. This places Vietnam a strong potential advocate of SDGs, particularly, empowering solar energy generation, given its geographic location being close to the equator. The nation is taking enormous advantage towards developing essential projects in both wind and solar energy resources [7]. Theoretically, the maximum volume of energy that Vietnam can generate is estimated to be ~100 GWh year from concentrated solar power and ~1.2 GWh from photovoltaic power [8], respectively. From these statistics, there appears to be a continuous growth of solar energy systems that require appropriate energy management technologies based on latest forecasting systems that can also help power utilities in management of renewable power that is passed into their grid system.

Considering the foresaid, solar energy generation is a promising option to fulfil the United Nation's SDGs [9], however, its principal element, solar irradiance can be heavily depended on the naturally uncontrolled atmospheric conditions. The global solar radiation (*GSR*) is highly intermittent and chaotic in nature [10] that a small variability in solar radiation can possibly lead to a major influence on the security of power generated. In this respect, reliable and versatile short-term solar energy forecasting models are vital for management of power systems that are dependent on incoming solar radiation, and how it is converted into usable electrical power [11].

Forecast models are categorised into three types of data-based systems such as those using satellite data [12], linear regression data [13] and neural network models adopting meteorological parameters to produce a viable relationship with solar radiation [14]. In terms of computational methods, there are two types of models namely the physical and the statistical model. However, literature reports common challenges in these methods such as a shortage of available measured data at solar powered sites to build these models, the time-consuming nature of the data collection of such datasets and the need for enormous computing power in solving the complex nonlinear equations is purely mathematical models are used in solar farms [15]. Therefore, researchers are continuously developing new models that can resolve existing problems associated with solar energy prediction methods.

2.0 Literature review

With increasing in computational techniques, machine learning-based forecast models have been proven to provide good performance over physically-based models [16-18]. Machine learning models

have also outperformed some of the other time series methods [19] that utilize historical patterns to predict solar radiation [20]. However, reported weaknesses in such methods include the occurrence of model bias [21] and possibly, model over-fitting issues [15]. To this end, significant progress has been made resolving the issues by a recent introduction of advanced machine learning methods such as deep learning techniques [22, 23] although the newer methods have not been fully explored in solar energy generation sector.

However, solar radiation phenomenon itself is significantly compounded, which is difficult to be predicted by a simple approach when high precision is required, for example, in real-time energy management connected to a grid. Since a single method can result in huge error [24], a compound approach or more specifically, a hybrid approach is encouraged with the aim of gaining higher precision [25]. In various types of hybrid methods used in solar forecasting, the decomposition-based ensemble learning approach is reported to provide promising results [26]. The guiding principle of this approach is to demodulate original intermittent signal into a more meaningful subset of sub-signals that can unveil the characteristics (*e.g.*, frequencies, jumps, trends and patterns) in a time series data. Each component of the sub-signal is then utilised as an input for a separate predictive model, then, aggregating those from all components to a final forecasted result. Since this approach can show high predictive accuracy, it has been studied extensively. Examples include discrete wavelet transform (DWT) [27], maximum overlap discrete wavelet transform [28], ensemble empirical mode decomposition [29] [30, 31] and its improved versions *e.g.*, complete ensemble empirical mode decomposition with adaptive noise [32]. In addition to these, another potent method in data decomposition that can plausibly to improve the predictive accuracy of a machine learning model is based on the local mean decomposition (LMD) algorithm, particularly, with respect to its application in EEG datasets [33, 34].

Despite their success, these approaches face significant limitations. With LMD as an example, we ascertain that while it can possibly analyse time-frequency representations by decomposing an original subset into a multicomponent subset of amplitude and frequency-modulated signal, this method can be easily influenced by the end effect and mode mixing problems [35]. In order to resolve these issues, the robust local mean decomposition (RLMD) [36], adopted in this paper, was introduced to optimize the moving average algorithm and its sifting process. Hence, the RLMD algorithm can be a promising tool to address issues of model input decomposition [36]. One study has applied RLMD with Long Short Term Memory (LSTM) a deep learning model, for forecasting wind speed [37]. In this study, the authors proposed an RLMD-LSTM model that yielded a greater accuracy compared to the other benchmark model [37]. However, the RLMD method has not yet been explored in solar energy forecasting, especially, the fact that solar radiation data are highly

intermittent and stochastic, and therefore, could enable the RLMD method to improve any standalone LSTM model.

With an aim to address issues of nonstationary properties in short-term solar energy forecasting with computationally efficient data-processing techniques (RLMD), this paper develops models for half hourly horizon. To capture the merits of the overall hybrid model, the study proposed a new model that integrates RLMD and LSTM, to generate the RLMD–LSTM hybrid model where we analyse the model inputs by RLMD which is then absorbed by LSTM predictive model. Four sites in the middle and South of Vietnam, the solar-rich metropolitan areas that earmarked as great potential sites for solar energy, are selected to validate the RLMD–LSTM hybrid model. This paper contributes to a new comprehension of deep learning models, for the first time, that integrates RLMD scheme with LSTM model to predict global solar radiation. The hybrid method is also benchmarked with standalone models namely the LSTM, MARS, SVR and alternative methods such as RLMD-LSTM, RLMD-MARS, RLMD-SVR algorithms. The application of LSTM and their variant algorithms have not been so popular in short-term solar energy forecasting studies in spite of their recent attention [23, 38]. Moreover, it appears to be quite rare to find any study that has used MARS and SVR with the RLMD method.

This paper has six parts. Firstly, Section 1 and 2 present a general background and literature review. Then, Section 3 provides a theoretical background of RLMD algorithm following by an overview of the material and model tunning process. After that, Section 5 discusses experimental results while Section 6 provides future work and limitations. Finally, the ultimate section summarises content of the paper.

3.0 Theoretical Overviews

Only the RLMD algorithm is comprehensively described here since this method is innovative and relatively new in short-term *GSR* forecasting. The conceptual details of LSTM, MARS and SVR are explained elsewhere because they are well-known approaches. The state-of-art literature of an LSTM model, as a deep learning method, is clearly described by Hochreiter [39] and its applications in terms of *GSR* forecasting are mentioned in other works (*e.g.*, [23], [40]). In this study, to compare our LSTM-based model, we also use MARS method proposed by Friedman [41] with some of its applications found elsewhere (*e.g.*, [42]). The SVR method is described in [43] with some of its applications found in [40], [27].

3.1 Robust Local Mean Decomposition

The original version of RLMD, also known as the Local Mean Decomposition (LMD) algorithm relies on the idea of decomposing an original signal into a subset of a less complicated sub-signals

which is an adaptive method to analyse time-frequency datasets (*e.g.*, through signal processing [33]). This decomposes a nonstationary signal into its product functions (PF) with the following content describing the process of LMD for a given signal x_t with $t = 0, 1, ..., .$ Firstly, the decomposition process starts at finding all of the maxima and minima of local extrema in the original signal (x_t) for example, the antecedent solar radiation timeseries, or related atmospheric predictor variables, used to predict the future solar radiation values. Following that, we calculate two serial extrema (*i.e.*, n_i and n_{i+1}) 's amplitude (a_i) and the local mean (m_i)as Eq. (1). Then, the algorithm aims to obtain the remaining sub-series by subtracting the inceptive local mean from the pure signal (2), (3). Next, the envelope function and products are obtained as per Eq. (4), (5) and (6)

$$m_i = \frac{n_i + n_{i+1}}{2};\ a_i = \frac{|n_i + n_{i+1}|}{2} \tag{1}$$

Where n_i with ($i = 1, 2, ..., K$) is the i^{th} extreme point and *K* denotes its total number. After this, a moving average method can form a continuous local mean function $m_{ij}(t)$, where $j = (1, 2, ..., K)$: the number of iterations to produce a normalized signal. Likewise, the relative envelope function $a_{ij}(t)$ is also generated. Noticeably, λ^* is expected to control the performance of the moving averaging method.

$$h_{11}(t) = x(t) - m_{11}(t) \tag{2}$$

$$s_{11}(t) = \frac{h_{11}(t)}{a_{11}(t)} \tag{3}$$

$$a_1(t) = a_{11}(t).a_{12}(t)._{....}.a_{1N}(t) = \prod_{j=1}^{N} a_{1j}(t) \tag{4}$$

$$\lim_{N\to\infty} a_{1N}(t) = 1$$

$$c_1(t) = a_1(t).s_{1N}(t) \tag{5}$$

$$x(t) = \sum_{v=1}^{k+1} c_v(t) \tag{6}$$

Although the LMD algorithm has a promising merit such as avoiding the existence of negative frequencies in an extracted subset of signals [34], the end effects and mode mixing issues are two most common problems. In order to overcome these drawbacks, the proposed RLMD method firstly needs to determine and further optimise some of the parameters. This process considers envelope estimation, boundary condition, and the sifting stopping criterion. Consequently, the new RLMD

algorithm was introduced so that it can simultaneously consider all these issues. More details about the RLMD algorithm can be found in [36]. Regarding the boundary condition problem, the RLMD algorithm generally uses mirror extension techniques to ascertain commensurate points for two ends of the original signal. For the estimation of envelope, the RLMD algorithm can achieve the best subset size based on Eq (7) and (8). Finally, the RLMD method also sifts the stopping criterion by using the function *F* as described in Eq. (9), (10) and (11).

$$\lambda^* = odd(\mu_s + 3\times\delta_s) \tag{7}$$

$$\mu_s = \sum_{i=1}^{K-1} s(i).f(i)\,;\ \delta_s = \sqrt{\sum_{i=1}^{K-1}\left[s(i)-\mu_s\right]^2.f(i)} \tag{8}$$

$$F = RMS\left[z(t)\right] + EK\left[z(t)\right] \tag{9}$$

$$RMS\left[z(t)\right] = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left[z(t)\right]^2} \tag{10}$$

$$EK\left[z(t)\right] = \frac{\frac{1}{n}\sum_{t=1}^{n}[z(t)-\bar{z}]^4}{(\frac{1}{n}\sum_{t=1}^{n}[z(t)-\bar{z}]^2)^2} - 3\,;\ \bar{z} = \frac{1}{n}\sum_{t=1}^{n} z(t) \tag{11}$$

In which:

$s(i)$: the local mean or amplitude signals‘ the step length;

$f(i)$: Product Function;

μ_s: mean of $s(i)$;

δ_s: $s(i)$ ’s standard deviation;

$odd(.)$: the process to attract the closest integer equal to or greater than the input.

$z(t) = a_i(t) - 1$: the zero-baseline envelope signal.

4.0 Materials and method

4.1 Global solar radiation

This paper has obtained and utilized *GSR* data (Wm^{-2}) from World Bank. The *GSR* data were measured simultaneously, 24 hours a day with an interval of one minute in four different regions in Vietnam from July 2017 to May 2019 as shown in Figure 1. In Figure 1, the presented study sites (*i.e.* the stations) comprise of Da-Nang (red), Central Highlands (green), Song Binh (blue) and Tri-An

(cyan). All locations are in the equatorial dry winter (September-March) and a dry season (April-August). To evaluate the versatility of the proposed RLMD-LSTM model, the selected study regions are representative of a diversity of climatic condition. For instance, Da-Nang is located in the middle centre of Vietnam with mountains, coastal plains and tropical climate whereas Central Highlands has many mountains with a cooler climate. Song-Binh and Tri-An are in the South of Vietnam which are more humid with two different seasons in a year [44]. This study utilises a 30-minute interval time series of *GSR*, for near real-time forecasting, albeit, from 6.00 AM to 06:00 PM as the daylight hours providing solar radiation as input to the solar generating system. There were some missing data mainly due to technical faults that were filled with the calendar mean values. The details of *GSR* dataset used to construct the RLMD-LSTM hybrid model for each location are presented in Table 1.

<Figure 1>; <Table 1>

4.2 Predictive Model Development

To fully benchmark the performance of the newly proposed RLMD-LSTM hybrid model in context of *GSR* forecasting in Vietnam, this study has compared its performance with the hybrid model versions namely, the RLMD-MARS and RLMD-SVR model, including their traditional or standalone versions (*i.e.*, LSTM, MARS and SVR).

The latter three models have used the statistically significant lagged *GSR* variables generated from the original *GSR* datasets (*i.e.*, considering historical behaviour of solar radiation) by means of the partial autocorrelation function, *PACF*. Henceforth, the RLMD algorithm has demodulated the original data series into the different sub-series before applying the *PACF* method to generate viable inputs for the first three (*i.e.*, RLMD-LSTM, RLMD-MARS and RLMD-SVR) models. Consequently, this study has developed a suite of six machine learning models used in half-hourly forecasting of solar radiation, with a deep learning (*i.e.*, LSTM) against two traditional machine learning (*i.e.*, MARS & SVR) models to fully evaluate the utility of the objective (*i.e.*, RLMD-LSTM) model.

It is noteworthy that for the case of the RLMD-based hybrid models, the complex time series were firstly dissolved by the RLMD scheme that aimed to decompose *GSR* based on the maximum to minimum frequencies which are technically known as Product Functions [45], [36]. To summary the general idea of this paper Figure 2 presents a flowchart whereas Figure 3(a) illustrates the dataset for all study sites demonstrating the six PFs, while Figure 3(b) illustrates the significant lagged *GSR* produced from the *PACF* function. Beside these, Table 1 also shows the number of PFs obtained for the various study locations.

<Figure 2>; <Figure 3a>; <Figure 3b>

All data were normalised and converted to [0,1] based on Eq. 12.

$$GSR_N = \frac{GSR_{ACTUAL} - GSR_{MIN}}{GSR_{MAX} - GSR_{MIN}} \tag{12}$$

$$GSR_{ACTUAL} = GSR_N (GSR_{MAX} - GSR_{MIN}) + GSR_{MIN} \tag{13}$$

Before building the models, partitioning of data into training and testing subsets was necessary [46]. There is no direct rule for data division, for example, the proportion of training and testing sets used in [38] was 75:25 whereas it was 80:20 in [23]. This paper has therefore used a 80:20 ratio for the training and testing sub-sets, respectively (Table 1).

This paper has adopted MATLAB (Ver 2019) with Intel *i*7, 3.6 GHz processor for the RLMD process. This process of building an RLMD-based model followed four primary steps. Firstly, it involved applying the RLMD algorithm to decompose *GSR* series into its Product Functions, as per Figure 3(a). Secondly, using $PACF$¸ this step reconstructed each subset into statistically significant lagged input series (Figure3b). After that, the RLMD-based models utilising the LSTM, MARS and SVR predictive algorithms were constructed using the sub-series, in order to forecast each of the PFs separately, and finally, the method also obtained the forecasted *GSR* from RLMD-based models by aggregating the forecasted PFs. Figure 3b illustrates this process.

To benchmark RLMD-LSTM hybrid models, two other kinds of machine learning models, namely, the MARS and SVR, were developed with and without RLMD. In the training phase, all predictive models (*i.e.*, LSTM, MARS and SVR) were constructed under a Python programming environment (Ver 3.6) with the Keras library [47]. The details of constructing an LSTM, MARS and the SVR model in Python environment are described in other works (*e.g.*, [40]). For the specific case of a MARS model, we have used the lowest Generalized Cross-Validation to optimise this algorithm following an earlier work [41]. Finally, we applied various performance metrics, as per Eq 14–24, to evaluate the precision of the proposed models. In the training phase, the RLMD-LSTM hybrid model was seen to generate the smallest value of *RMSE* and the highest value of *LM* – which are robust metrics indicating superior performance of this model over the counterpart models. Table 2 shows the performance of all models in the training phase.

<Table 2>

4.3 Model Performance Criteria

In respect to the precision and accuracy assessment of the proposed objective model (*i.e.*, RLMD-LSTM) *vs.* the RLMD-MARS, RLMD-SVR and the standalone machine learning models (*i.e.*,

LSTM, MARS, and SVR) applied in *GSR* forecasting, this study has employed performance metrics in the testing phase with their mathematical formulations as follows[48, 49]

a. Correlation coefficient (*r)*

$$r = \left(\frac{\sum_{i=1}^{n} (GSR^{OBS} - < GSR^{OBS} >)(GSR^{FOR} - < GSR^{FOR} >)}{\sqrt{\sum_{i-1}^{N} (GSR^{OBS} - GSR^{OBS})^2} \sqrt{\sum_{i-1}^{N} (GSR^{FOR} - GSR^{FOR})^2}} \right)^2 \quad (14)$$

b. Mean Absolute Error (*MAE*)

$$MAE = \left(\sum_{N}^{t=1} \left| GSR^{OBS} - GSR^{FOR} \right| \right) / N \quad (15)$$

c. Root mean square error (*RMSE):*

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (GSR^{OBS} - GSR^{FOR})^2} \quad (16)$$

d. Nash-Sutcliffe coefficient (E_{NS} *)*

$$E_{NS} = 1 - \frac{\sum_{i-1}^{N} \left[GSR^{OBS} - GSR^{FOR} \right]^2}{\sum_{i-1}^{N} \left[GSR^{OBS} - GSR^{FOR} \right]^2} \quad (17)$$

, - $-\infty \le E_{NS} \le 1$

e. Willmott's index (*WI*)

$$WI = 1 - \frac{\sum_{i-1}^{N} \left[GSR^{OBS} - GSR^{FOR} \right]^2}{\sum_{i-1}^{N} \left[\left| (GSR^{FOR} - (GSR^{OBS})) \right| + \left| (GSR^{OBS} - (GSR^{FOR})) \right| \right]^2} \quad (18)$$

, $0 \le WI \le 1$

f. Legates-McCabe's (*LM*)

$$LM = 1 - \frac{\sum_{i=1}^{n} \left| GSR^{OBS} - GSR^{FOR} \right|}{\sum_{i=1}^{n} \left| GSR^{OBS} - < GSR^{OBS} > \right|} \quad (19)$$

, - $-\infty \le LM \le 1$

g. Relative root mean square error (RRMSE¸%)

$$RRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(GSR_i^{FOR} - GSR_i^{OBS})^2}}{\frac{1}{N}\sum_{i=1}^{N} GSR_i^{OBS}} \times 100 \quad (21)$$

5.0 Results and Discussion

The results presented here demonstrate the efficacy of the hybrid LSTM model integrated with an RLMD algorithm for half hourly *GSR* forecasting at four Vietnamese cities (*i.e*., Da-Nang, Central Highlands, Song-Binh, Tri-An). The benchmark models, which demonstrate the superiority of RLMD-LSTM hybrid model are the RLMD-MARS and the RLMD-SVR models, together with standalone models such as LSTM, MARS and SVR. The efficacy of our objective model is based on the testing phase where statistical performance metrics (Section 4.3) are adopted to validate the results, presented and argued in the following.

The predictive capacity of RLMD-LSTM hybrid model for all study sites is summarised in Table 3. As can be seen, in comparison to the standalone models, the RLMD-LSTM hybrid model performs the best among all the four study sites with the lower *RMSE, MAE* and the largest magnitude of *WI* and E_{NS} . Hence, it this indicates that the RLMD algorithm, which can reveal the concealed patterns in *GSR* time-series, can possibly addresses the complex data patterns employed as potential inputs (Figure 2).

For instance, for the case of Da-Nang, the *WI*, *RMSE*, and *MAE* values were ~4% larger and ~ -31.5% and 35% lower for the RLMD-LSTM hybrid model compared to the LSTM standalone model. Moreover, the accuracy statistics of the RLMD-LSTM model in terms of E_{NS} =0.9248 compares with the LSTM value of E_{NS} =0.8700 which also indicates a better performance of the objective hybrid model. Likewise, for the other three study sites, in Table 3 we illustrate the relative advantages of the RLMD-MARS and RLMD-SVR hybrid models over the standalone MARS and SVR models, although the overall performance of these do not exceed our RLMD-LSTM hybrid model.

<**Table 3**>

Taking the Da-Nang study site as an example, the results indicate that the RLMD-LSTM model yields a value of (*r= 0.9628, RMSE=2,602* Wm^{-2}, *MAE=1,702* Wm^{-2}, *RRMSE=19%, LM=0.7988)*, which unambiguously seem to outperform the metrics of the RLMD-MARS model with (*r= 0.96048, RMSE=2,773* Wm^{-2}, *MAE=1,958* Wm^{-2}, *RRMSE=20%, LM=0.7685)* and the RLMD-SVR with (*r= 0.9512, RMSE=4,874* Wm^{-2}, *MAE=4,226* Wm^{-2}, *RRMSE=35%, LM=0.5004).* A similar deduction can also be made for the other study sites such as Central Highlands, Song-Binh,

and Tri-An, further emphasising the superior capability of the newly proposed RLMD-LSTM hybrid model.

When evaluated in respect to the *RRMSE* and *LM* values, the more robust metrics providing an objective evaluation of the newly proposed RLMD-LSTM hybrid model [50], we can clearly see the benefits of the RLMD-LSTM hybrid model for all four study sites (Table 3). For example, for the case of Da-Nang, the *RRMSE* value was ~19% with an *LM* value of 0.7988 for the RLMD-LSTM approach compared to a higher value of ~20% and a lower value of ~0.7685 for the case of RLMD-MARS model, ~35% and ~0.5004 for the case of RLMD-SVR model. As it can be seen in the other study sites, the performance of the standalone models (without actually utilising the RLMD-based data decomposition methods) registered a higher value of *RRMSE* and a lower value of *LM* that were largely not satisfactory.

In terms of the predicted and the observed *GSR* values, the RLMD-LSTM hybrid model has yielded the highest correlation coefficient with *r=0.9628, 0.9639, 0.9597, 0.9714* for all four sites following by the relatively lower value for the RLMD-MARS, RLMD-SVR, and the LSTM, MARS, SVR models for Da-Nang (*r=0.93-0.96),* Central Highlands (*r=0.94-0.96),* Song-Binh (*r=0.92-0.95),* Tri-An (*r=0.94-0.96).* Consistent with these results, Figure 4 illustrates a scatterplot. This figure shows correlative relationship between forecasted and observed *GSR* by linear equation, $GSR = m\ GSR + c$ with a coefficient of determination R^2 [14]. Note that the '*m*' is a gradient for 1:1 correlation, R^2 measures the covariance and '*c*' shows an intercept on the *y*-axis that should be trivial for a perfect forecasting model.

Notably, the RLMD-LSTM hybrid model has yielded a superior performance than the counterpart models so that for the case of Da-Nang site (Fig. 3a), the RLMD-LSTM yielded $m = 1.02$, $c = 206$ and $R^2 = 0.9628$ which exceeds the value for RLMD-MARS ($m = 1.04$, $c = 156$ and $R^2 = 0.9604$), and RLMD-SVR ($m = 0.96$, $c = -3316$ and $R^2 = 0.9512$). A similar conclusion can also be made in Figure 3(b–d) for Central Highlands, Song-Binh, Tri-An, respectively. Importantly, R^2 for the case of the RLMD-LSTM hybrid model is also higher than that of the RLMD-MARS and RLMD-SVR models showing the better capability of the LSTM-based deep learning method even with the incorporating of RLMD method with MARS and SVR (non-deep learning) models. Additionally, the standalone models such as LSTM, MARS and SVR also generated poorer results compared to the RLMD-LSTM hybrid model.

<Figure 4>; <Figure 5>

To further investigate the machine learning models that generate smaller errors for half-hourly *GSR* prediction, Figure 5 illustrates a boxplot that has the absolute prediction error in terms of forecasted and observed solar radiation. Like earlier results, the standalone models such as LSTM,

MARS and SVR show errors that are dramatically larger than their hybrid model counterparts at all study locations. This concurs with previous results (*i.e.*, Table 3, Figure 3) that the RLMD-LSTM hybrid model is seen to occupy a magnitude of the forecasted errors compared to their counterpart models. Overall, the RLMD-LSTM hybrid model has attained a greater precision in forecasting of half-hourly solar radiation for the sites in Vietnam, further cementing the advanced modelling capabilities of the proposed RLMD method mainly to improve the deep learning, LSTM-based model.

<Figure 5>

As a complementary visual measure of evaluating the half-hourly forecasting model, we revert to a Taylor diagram, as per Figure 6 [51] that illustrates the relative distance between observed and forecasted values in the testing phase of the comparative models. In this figure, the radial axis presents the *r*-value whereas the polar axis presents the standard deviation. In all cases, it is evident that the RLMD-LSTM model yields the highest *r-value* and has achieved the closest distance between the forecasted and the observed *GSR* values whereas those for the comparative models were further away from the reference (observed) value. It is therefore plausible that the RLMD method integrated with LSTM has great ability to improve the forecasted values in testing phase, by approximately 50%. Given the robustness of the newly proposed RLMD-LSTM model and its capability to outperform the benchmark models, this this study clearly points out the benefits of this data decomposition approach, and its use in reliable global solar radiation forecasting tools for short-term horizons. Not only is the RLMD-LSTM method a likely candidate for global solar radiation prediction but also could offer an effective tool for the monitoring of electricity production, energy supply-demand systems and operation management of solar farms. While we have tested the benefits of RLMD-LSTM hybrid model for the specific case of solar radiation forecasting, this method may also be useful other renewable energy systems such as wind, tidal and wave energy conversion systems [26, 32, 52-55] where model input data also has a significant signature of non-stationary and chaotic behaviour.

6.0 Limitations and Future Work

Despite good success of the newly designed RLMD-LSTM hybrid model in context of half-hourly *GSR* forecasting, ascertained by statistical and visual analysis of results, there remain some degree of limitations that should be addressed in future studies. First of all, in this paper, only the historical *GSR* datasets were used as potential inputs for the RLMD-LSTM model, however, one must also consider the other climatic factors such as temperature, rainfall, humidity and most importantly, the cloud cover data that could also affect the ground signature of *GSR* [56]. Therefore, developing an alternative forecasting model utilising both atmospheric input data and weather variables could be a feasible pathway for future work to include more physical variables that drive solar energy generation

process. This is because models can be affected by sky and weather conditions [57], so the inclusion of numerical weather models such as global forecast system (GFS) as weather forecast model producing 6-hourly forecasts, made up of separate forecasts, or ensemble members can be used to address uncertainty in weather observations, or to initialize weather forecast models, should be integrated to an RLMD-LSTM model. In addition to this, the RLMD-LSTM approach can also be enhanced by the use of an ensemble algorithm for uncertainty testing perhaps, through a Bayesian Model Averaging framework [58] and a bootstrapping method [59].

7.0 Conclusion

This study has innovated an RLMD-LSTM hybrid model for predicting half-hourly *GSR* at four study sites in Vietnam. To fulfil this aim, the solar radiation datasets in Vietnam from 2017 to 2019 at 30-minute intervals were decomposed by an RLMD algorithm into the product functions (PFs) that better represented the patterns in historical *GSR*, capturing the physical influence of atmospheric variables, and then used to simulate future *GSR* values. After generating the PACF of the corresponding PFs, statistically significant lagged inputs were passed into the deep learning, LSTM predictive model to formulate the RLMD-LSTM hybrid model with the overall aim to achieve a better accuracy compared to its standalone forms. As a result, the RLMD technique was found to be a robust version of the LMD method to address nonstationary properties in global solar radiation, mainly by its capability to demodulate the original series into its respective PFs (see Table 1 and Fig. 2a).With the overall purpose of evaluating the predictive capability of the proposed model, this paper for the first time, has also developed two other forms of hybrid models (*i.e.*, RLMD-MARS, RLMD-SVR) and three other forms of standalone models (*i.e.*, LSTM, MARS, and SVR). By assessing the predictive performance through using various infographics with a combination of statistical score metrics (*e.g., r-values, RMSE*, Legates & McCabe's Index) computed in the independent tested datasets, the high performance metrics and low forecasting errors were evident for the case of RLMD-LSTM model, outperforming the counterpart models (see Table 3). Moreover, with a reasonable variability in the performance of all models considering the diverse locations they were tested for, the RLMD-LSTM hybrid model achieved a generally stable forecasting capability. As a result, we conclude that the RLMD-LSTM hybrid model is a considerably superior methodology for short-term solar radiation forecasting. This conclusion stands valid in respect to the advantages of the proposed RLMD-LSTM model over the other models, albeit, also suggesting that the utilisation of seasonal and climate data, cloud cover, and including variables from numerical weather simulation model to better capture the physical atmospheric conditions could a pathway to further adopt RLMD-LSTM model in future studies.

This study has generated a novel methodology for very short-term forecasting horizons (*i.e.*, half-hourly) to be utilized to forecast future *GSR* values. Based on the efficacy of this forecast system, we conclude that the use of the method in solar and other forms of renewable energy managements can provide stakeholders with key information on solar energy variability to help develop strategic plans for energy production, grid operation and the integration of renewables into their power conventional generation systems for better energy security. Moreover, the proposed forecast system based on RLMD-LSTM model can also be applied in other fields such as oceanic wave based electricity generation to assist policymakers in a better and optimal management of freely available energy resources.

Acknowledgements

The solar radiation data was obtained from World Bank solar energy projects for Vietnam. The authors are also grateful to all reviewers in providing meaningful feedback in order to improve this research paper.

References

1. Hung, N.N., N.H. Anh, and N.T. Hai, *Institute of Energy.* ENERGY SECTOR IN VIETNAM AND RENEWABLE ENERGY ROLE, 2018.
2. Sanseverino, E.R., et al., *Review of Potential and Actual Penetration of Solar Power in Vietnam.* Energies, 2020. **13**(10): p. 1-25.
3. Luong, N.D., *A critical review on energy efficiency and conservation policies and programs in Vietnam.* Renewable and Sustainable Energy Reviews, 2015. **52**: p. 623-634.
4. Tan, Z., et al., *Photovoltaic power generation in China: Development potential, benefits of energy conservation and emission reduction.* Journal of Energy Engineering, 2012. **138**(2): p. 73-86.
5. Hoat, D., et al., *Research Overview of New and Renewable Energy in Vietnam and Development Orientation.* Vietnam Academy of Science and Technology: Hanoi, Vietnam, 2007.
6. !!! INVALID CITATION !!! (Luong, 2015, IEA, 2012, Shem et al., 2019).
7. Outlook, I.S.A.E., *International Energy Agency: Bangkok*. 2019, Thailand.
8. Polo, J., et al., *Maps of solar resource and potential in Vietnam.* 2015.
9. Farivar, G. and B. Asaei, *A new approach for solar module temperature estimation using the simple diode model.* IEEE transactions on energy conversion, 2011. **26**(4): p. 1118-1126.
10. Perez, R., et al., *The cost of mitigating short-term PV output variability.* Energy Procedia, 2014. **57**(1000178044 & 1000177911): p. 755-762.
11. Mostafavi, E.S., et al., *A hybrid computational approach to estimate solar global radiation: an empirical evidence from Iran.* Energy, 2013. **49**: p. 204-210.
12. Lu, N., et al., *A simple and efficient algorithm to estimate daily global solar radiation from geostationary satellite data.* Energy, 2011. **36**(5): p. 3179-3188.
13. Benghanem, M. and A. Joraid, *A multiple correlation between different solar parameters in Medina, Saudi Arabia.* Renewable Energy, 2007. **32**(14): p. 2424-2435.
14. Ghimire, S., et al., *Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cities of Queensland Australia.* Journal of cleaner production, 2019. **216**: p. 288-310.
15. Voyant, C., et al., *Machine learning methods for solar radiation forecasting: A review.* Renewable Energy, 2017. **105**: p. 569-582.
16. Alzahrani, A., et al., *Solar irradiance forecasting using deep neural networks.* Procedia Computer Science, 2017. **114**: p. 304-313.
17. Paulescu, E. and R. Blaga, *Regression models for hourly diffuse solar radiation.* Solar Energy, 2016. **125**: p. 111-124.

18. Coulibaly, O. and A. Ouedraogo, *Correlation of global solar radiation of eight synoptic stations in Burkina Faso based on linear and multiple linear regression methods.* Journal of Solar Energy, 2016. **2016**.
19. Bracale, A., et al., *A Bayesian method for short-term probabilistic forecasting of photovoltaic generation in smart grid operation and control.* Energies, 2013. **6**(2): p. 733-747.
20. Ramedani, Z., et al., *Potential of radial basis function based support vector regression for global solar radiation prediction.* Renewable and Sustainable Energy Reviews, 2014. **39**: p. 1005-1011.
21. De Leone, R., M. Pietrini, and A. Giovannelli, *Photovoltaic energy production forecast using support vector regression.* Neural Computing and Applications, 2015. **26**(8): p. 1955-1962.
22. Ghimire, S., et al., *Deep learning neural networks trained with MODIS satellite-derived predictors for long-term global solar radiation prediction.* Energies, 2019. **12**(12): p. 2407.
23. Ghimire, S., et al., *Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms.* Applied Energy, 2019. **253**: p. 113541.
24. Orjuela-Cañón, A.D., J. Hernández, and C.R. Rivero. *Very short term forecasting in global solar irradiance using linear and nonlinear models*. in *2017 IEEE workshop on power electronics and power quality applications (PEPQA)*. 2017. IEEE.
25. Boroojeni, K.G., et al., *A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon.* Electric Power Systems Research, 2017. **142**: p. 58-73.
26. Ali, M., et al., *Near real-time significant wave height forecasting with hybridized multiple linear regression algorithms.* Renewable and Sustainable Energy Reviews, 2020. **132**: p. 110003.
27. Ghimire, S., et al., *Wavelet-based 3-phase hybrid SVR model trained with satellite-derived predictors, particle swarm optimization and maximum overlap discrete wavelet transform for solar radiation prediction.* Renewable and Sustainable Energy Reviews, 2019. **113**: p. 109247.
28. Deo, R.C., X. Wen, and F. Qi, *A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset.* Applied Energy, 2016. **168**: p. 568-593.
29. Ali, M. and R. Prasad, *Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition.* Renewable and Sustainable Energy Reviews, 2019. **104**: p. 281-295.
30. Prasad, R., et al., *Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation.* Applied energy, 2019. **236**: p. 778-792.
31. Prasad, R., et al., *A double decomposition-based modelling approach to forecast weekly solar radiation.* Renewable Energy, 2020. **152**: p. 9-22.
32. Al-Musaylh, M.S., et al., *Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting.* Applied Energy, 2018. **17**: p. 422-439.
33. Smith, J.S., *The local mean decomposition and its application to EEG perception data.* Journal of the Royal Society Interface, 2005. **2**(5): p. 443-454.
34. Cheng, J., Y. Yang, and Y. Yang, *A rotating machinery fault diagnosis method based on local mean decomposition.* Digital Signal Processing, 2012. **22**(2): p. 356-366.
35. Junsheng, C., et al., *Comparison between the methods of local mean decomposition and empirical mode decomposition.* Journal of Vibration and Shock, 2009. **28**(5): p. 13-16.
36. Liu, Z., et al., *Time-frequency representation based on robust local mean decomposition for multicomponent AM-FM signal analysis.* Mechanical Systems and Signal Processing, 2017. **95**: p. 468-487.
37. Jiang, Y., et al., *A novel wind speed prediction method based on robust local mean decomposition, group method of data handling and conditional kernel density estimation.* Energy Conversion and Management, 2019. **200**: p. 112099.
38. Zhou, H., et al., *Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism.* IEEE Access, 2019. **7**: p. 78063-78074.
39. Hochreiter, S. and J. Schmidhuber, *Long short-term memory.* Neural computation, 1997. **9**(8): p. 1735-1780.
40. Huynh, A.N.-L., et al., *Near real-time global solar radiation forecasting at multiple time-step horizons using the long short-term memory network.* Energies, 2020. **13**(14): p. 3517.
41. Friedman, J.H., *Multivariate adaptive regression splines.* The annals of statistics, 1991: p. 1-67.

42. Srivastava, R., A. Tiwari, and V. Giri, *Solar radiation forecasting using MARS, CART, M5, and random forest model: A case study for India.* Heliyon, 2019. **5**(10): p. e02692.
43. Drucker, H., et al. *Support vector regression machines*. in *Advances in neural information processing systems*. 1997.
44. Yusuf, A.A. and H. Francisco, *Climate change vulnerability mapping for Southeast Asia.* 2009.
45. Färe, R., S. Grosskopf, and D. Tyteca, *An activity analysis model of the environmental performance of firms—application to fossil-fuel-fired electric utilities.* Ecological economics, 1996. **18**(2): p. 161-175.
46. Prasad, R., et al., *Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm.* Atmospheric Research, 2017. **197**: p. 42-63.
47. Chollet, F., *et al. 2015. Keras*.
48. Dawson, C.W., R.J. Abrahart, and L.M. See, *HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts.* Environmental Modelling & Software, 2007. **22**(7): p. 1034-1052.
49. Legates, D.R. and G.J. McCabe Jr, *Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation.* Water resources research, 1999. **35**(1): p. 233-241.
50. Mohammadi, K., et al., *A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation.* Energy Conversion and Management, 2015. **92**: p. 162-171.
51. Jolliff, J.K., et al., *Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment.* Journal of Marine Systems, 2009. **76**(1-2): p. 64-82.
52. Deo, R.C., et al., *Multi-layer perceptron hybrid model integrated with the firefly optimizer algorithm for windspeed prediction of target site using a limited set of neighboring reference station data.* Renewable Energy, 2018. **116**: p. 309-323.
53. Al-Musaylh, M.S., et al., *Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia.* Advanced Engineering Informatics, 2018. **35**: p. 1-16.
54. Al-Musaylh, M.S., et al., *Short-term electricity demand forecasting using machine learning methods enriched with ground-based climate and ECMWF Reanalysis atmospheric predictors in southeast Queensland, Australia.* Renewable and Sustainable Energy Reviews, 2019. **113**: p. 109293.
55. Mohanad, S.A.-M., C.D. Ravinesh, and L. Yan, *Particle Swarm Optimized–Support Vector Regression Hybrid Model for Daily Horizon Electricity Demand Forecasting Using Climate Dataset.* E3S Web Conf., 2018. **64**: p. 08001.
56. Kashyap, Y., A. Bansal, and A.K. Sao, *Solar radiation forecasting with multiple parameters neural networks.* Renewable and Sustainable Energy Reviews, 2015. **49**: p. 825-835.
57. Zang, H., et al., *Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations.* Renewable Energy, 2020. **160**: p. 26-41.
58. Sloughter, J.M., T. Gneiting, and A.E. Raftery, *Probabilistic wind speed forecasting using ensembles and Bayesian model averaging.* Journal of the american statistical association, 2010. **105**(489): p. 25-35.
59. Tiwari, M.K. and C. Chatterjee, *A new wavelet–bootstrap–ANN hybrid model for daily discharge forecasting.* Journal of Hydroinformatics, 2011. **13**(3): p. 500-519.

CHAPTER 6: CONCLUSIONS

6.1 Foreword

This chapter presents the conclusions and the limitations of this study. The key contributions and findings are mentioned, as well as various future research suggestions.

6.2 Conclusions

This study, within the context of a Master's Thesis, has conducted an in-depth analysis of solar radiation forecasting model techniques in the context of Vietnam. Several standalone and hybridised models were presented at multiple forecast horizons (*i.e.* 1-minute, 5-minute, 10-minute, 15-minute, 30-minute). The adopted forecasting models derived from machine learning, and deep learning techniques, including data decomposition and model parameter optimisations selected in this work, were: multivariate adaptive regression splines (MARS); support vector regression (SVR); autoregressive integrated moving average (ARIMA); Long Short-Term Memory (LSTM); Deep Recurrent Neural Network (RNN); Multilayer Perceptron (MLP); and, Robust Local Mean Decomposition (RLMD).

The first set of findings is explained in the first paper. In particular, the SVR, ARIMA, LSTM, RNN, MLP with grid search were developed and evaluated in Chapter 3 for short-term global solar radiation forecasting in Bac-Ninh, Vietnam, using ground-measured based data from the World Bank. Moreover, several types of evaluation metrics, visual analysis and Diebold–Mariano statistic tests were employed to assess the performance of forecasting models, from which it was shown that the LSTM model yielded the most accurate results. In short, LSTM was a useful tool to absorb intricate patterns of solar radiation and provided a highly reliable forecasting result.

Using the same source data for Objective 1, the second paper (Chapter 5) developed the LSTM model integrated with RLMD for predicting half-hourly GSR at four study sites in Vietnam. RLMD possibly addresses the non-stationarity problem of solar radiation by decomposing GSR time-series data into the product functions (PFs). After generating the PACF of the corresponding PFs, statistically significant lagged inputs were passed into the deep learning, LSTM predictive model to formulate the RLMD-LSTM hybrid model, with the overall aim of achieving better accuracy compared to its standalone forms.

With the overall purpose of evaluating the predictive capability of the proposed model, this paper, for the first time, has also developed two other forms of hybrid models (*i.e.* RLMD-MARS, RLMD-SVR) and three other forms of standalone models (*i.e.* LSTM, MARS and SVR). By assessing the predictive performance through using various infographics with a combination of statistical score metrics (*e.g.* r-value, RMSE, Legates & McCabe's Index), computed in the independently tested datasets, RLMD-LSTM outperformed the counterparts with the highest performance metrics and low forecasting errors. Moreover, in terms of considering the diverse studied locations, the RLMD-LSTM hybrid model achieved a generally stable forecasting capability. As a result, the RLMD-LSTM hybrid model can be seen as a superior methodology in terms of short-term solar radiation forecasting.

To sum up, this study provides various novel contributions in terms of a predictive model for short-term solar radiation forecasting. According to the results, the proposed models outperformed those relative standalone models. The main contributions of this study are as follows:

1. The first contribution is to explore novel forecasting approaches that have not been previously undertaken in terms of short-term solar radiation forecasting in Vietnam. The study applied LSTM as the main model to be compared with those methods that were previously developed (*i.e.* ARIMA, SVR, DNN, MLP).
2. The second contribution is to develop a hybrid two-phase model utilising the RLMD algorithm to address the non-stationarity and complex patterns of solar radiation, by decomposing them into different sub-sets, whereas LSTM was an outstanding forecasting model when comparing with the counterparts (*i.e.* MARS, SVR).
3. In terms of short-term forecasting, the forecasted error reduced significantly, which proved the capability of the proposed model in this study. In particular, the proposed model integrating a pre-processing technique with a deep learning algorithm can overcome the drawbacks of standalone optimisation algorithms.
4. This study provides robust forecasting models, which are helpful for generators and authorities in Vietnam specifically and worldwide, by addressing the existing challenges in sustainable development and security of solar energy.

6.3 Limitations and Recommendations for future works

Although this study has developed and evaluated many different predictive models for short-term solar radiation forecasting that performed quite accurately, there are some limitations and difficulties experienced during research implementation. This section summarises several challenges in both developments and applications of the proposed techniques, which could potentially form interesting subjects for further investigation to address these limitations in the future. In general, the development of hybrid LSTM-based models includes two separate parts: decomposing time-series data, and fitting the LSTM model. Two considerable limitations relate to each of those parts.

1. The time-series data used in this study: Since the related parameters of global solar radiation in short-term time horizons are rare, only the antecedent global solar radiation datasets were used as potential inputs. However, future study should also take other climatic factors (*e.g.* temperature, rainfall, humidity and cloud cover data) into consideration as these factors might also influence the ground signature of global solar radiation (Kashyap et al., 2015). Thus, proposing other forecasting models, applying the atmospheric input data, as well as weather variables possibly, are feasible pathways for future study. Besides, one can include more physical variables that drive the solar energy generation process. Since sky and weather conditions can affect forecasting model performance (Zang et al., 2020), future study can also consider the inclusion of numerical weather models or ensemble members can be used to address uncertainty in weather observations or to initialise weather forecast models. Furthermore, future work can use a low-cost sky imaging system, together with covering different weather conditions (*e.g.* cloudy, dusty, sunny, foggy or rainy), as a potential data source.

2. Methodologies: Based on results from this study and previous studies (Huynh et al., 2020; Ghimire et al., 2019b), although learning methods are powerful tools which deserve a place in terms of solar energy, there are some limitations caused by their black box characteristics. In particular, the black box properties limit the understanding and verification of the complex relationships between predictor variables that a learning model performs. Thus, an integration of data-driven and numerical weather parameter models has become common. It is also a potential approach to combine learning and physical models for future study. Moreover, uncertainty as a common issue can be enhanced in the future by the use of an

ensemble algorithm for uncertainty testing, perhaps through a Bayesian Model Averaging framework (Sloughter et al., 2010) and a bootstrapping method (Tiwari and Chatterjee, 2011).

3. Despite effective skill of the proposed method in addressing non-stationarity of global solar radiation as well as solving the issue from different significant lagged inputs of each signal (Al-Musaylh et al., 2018b), future study could take the selection of parameters before applying forecasting technique into consideration. The main aim of feature selection is important (*e.g.* to prevent over-fitting and enhance model performance) since many data-driven methods do not cope with a large volume of irrelevant features. For this reason, feature selection has been common in many studies such as (Abedinia et al., 2018; Almaraashi, 2018). In addition, there are common Multiresolution Analysis techniques (Ruiz-Arias et al.) (*e.g.* empirical mode decomposition (EMD)) which can be considered since they extract relevant information of time-frequency domain without any loss of information (Guermoui et al., 2020). Furthermore, LSTM is only one of the deep learning neural networks, and there are other models (*e.g.* Convolutional Neural Network (CNN) or Bayesian Deep Learning) (Ghimire et al., 2019a; Qian et al., 2019) that can be used. For instance, convolutional neural nets can be efficient in terms of spatial averaging using weather variables, which can be seen as a filter to learn among geographical regions. Therefore, the proposed method in this study should be meaningfully compared to other hybridised models formed by the integration of Multiresolution Analysis techniques (Ruiz-Arias et al.) and deep learning techniques for future study.

In summary, based on the efficacy of this forecast system, this thesis provides new contributions in solar and other forms of renewable energy management through predictive models. The reliable and precise forecasting results provide stakeholders with key information on solar energy variability to help develop strategic plans for energy production, grid operation and the integration of renewables into their power conventional generation systems, for better energy security.

LIST OF REFERENCES

Abedinia, O.,Amjady, N. & Ghadimi, N. 2018. Solar energy forecasting based on hybrid neural network and improved metaheuristic algorithm. *Computational Intelligence,* 34**,** 241-260.

Al-Musaylh, M. S.,Deo, R. C.,Adamowski, J. F. & Li, Y. 2018a. Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. *Advanced Engineering Informatics,* 35**,** 1-16.

Al-Musaylh, M. S.,Deo, R. C.,Li, Y. & Adamowski, J. F. 2018b. Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting. *Applied Energy,* 217**,** 422-439.

Alfadda, A.,Adhikari, R.,Kuzlu, M. & Rahman, S. Hour-ahead solar PV power forecasting using SVR based approach. 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2017. IEEE, 1-5.

Almaraashi, M. 2018. Investigating the impact of feature selection on the prediction of solar radiation in different locations in Saudi Arabia. *Applied Soft Computing,* 66**,** 250-263.

Alzahrani, A.,Shamsi, P.,Dagli, C. & Ferdowsi, M. 2017. Solar irradiance forecasting using deep neural networks. *Procedia Computer Science,* 114**,** 304-313.

Anderson, D. & Leach, M. 2004. Harvesting and redistributing renewable energy: on the role of gas and electricity grids to overcome intermittency through the generation and storage of hydrogen. *Energy policy,* 32**,** 1603-1614.

Azimi, R.,Ghayekhloo, M. & Ghofrani, M. 2016. A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting. *Energy Conversion and Management,* 118**,** 331-344.

Benmouiza, K. & Cheknane, A. 2016. Small-scale solar radiation forecasting using ARMA and nonlinear autoregressive neural network models. *Theoretical and Applied Climatology,* 124**,** 945-958.

Box, G. 1994. P & Jenkins, GM & Reinsel, GC. *Time Series Analysis, Forecasting and Control*. Bracale, A.,Caramia, P.,Carpinelli, G.,Di Fazio, A. & Ferruzzi, G. 2013. A Bayesian method for short-term probabilistic forecasting of photovoltaic generation in smart grid operation and control. *Energies,* 6**,** 733-747.

Caballero, R.,Zarzalejo, L. F.,Otero, Á.,Piñuel, L. & Wilbert, S. 2018. Short term cloud nowcasting for a solar power plant based on irradiance historical data. *Journal of Computer Science & Technology,* 18.

Colak, I.,Yesilbudak, M.,Genc, N. & Bayindir, R. Multi-period prediction of solar radiation using ARMA and ARIMA models. 2015 IEEE 14th international conference on machine learning and applications (ICMLA), 2015. IEEE, 1045-1049.

Coulibaly, O. & Ouedraogo, A. 2016. Correlation of global solar radiation of eight synoptic stations in Burkina Faso based on linear and multiple linear regression methods. *Journal of Solar Energy,* 2016.

Dash, P.,Majumder, I.,Nayak, N. & Bisoi, R. 2020. Point and Interval Solar Power Forecasting Using Hybrid Empirical Wavelet Transform and Robust Wavelet Kernel Ridge Regression. *Natural Resources Research***,** 1-29.

De Leone, R.,Pietrini, M. & Giovannelli, A. 2015. Photovoltaic energy production forecast using support vector regression. *Neural Computing and Applications,* 26**,** 1955-1962.

Deo, R. C.,Wen, X. & Qi, F. 2016. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Applied Energy,* 168**,** 568-593.

Díaz–Vico, D.,Torres–Barrán, A.,Omari, A. & Dorronsoro, J. R. 2017. Deep neural networks for wind and solar energy prediction. *Neural Processing Letters,* 46**,** 829-844.

Espinar, B.,Aznarte, J.-L.,Girard, R.,Moussa, A. M. & Kariniotakis, G. Photovoltaic Forecasting: A state of the art. 5th European PV-Hybrid and Mini-Grid Conference, 2010. OTTI-Ostbayerisches Technologie-Transfer-Institut, Pages 250-255-ISBN 978-3-941785-15-1.

Farivar, G. & Asaei, B. 2011. A new approach for solar module temperature estimation using the simple diode model. *IEEE transactions on energy conversion,* 26**,** 1118-1126.

Fentis, A.,Bahatti, L.,Mestari, M. & Chouri, B. Short-term solar power forecasting using Support Vector Regression and feed-forward NN. 2017 15th IEEE International New Circuits and Systems Conference (NEWCAS), 2017. IEEE, 405-408.

Gensler, A.,Henze, J.,Sick, B. & Raabe, N. Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks. 2016 IEEE international conference on systems, man, and cybernetics (SMC), 2016. IEEE, 002858-002865.

Ghimire, S.,Deo, R. C.,Raj, N. & Mi, J. 2019a. Deep learning neural networks trained with MODIS satellite-derived predictors for long-term global solar radiation prediction. *Energies,* 12**,** 2407.

Ghimire, S.,Deo, R. C.,Raj, N. & Mi, J. 2019b. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Applied Energy,* 253**,** 113541.

Ghimire, S.,Deo, R. C.,Raj, N. & Mi, J. 2019c. Wavelet-based 3-phase hybrid SVR model trained with satellite-derived predictors, particle swarm optimization and maximum overlap discrete wavelet transform for solar radiation prediction. *Renewable and Sustainable Energy Reviews,* 113**,** 109247.

Golestaneh, F.,Pinson, P. & Gooi, H. B. 2016. Very short-term nonparametric probabilistic forecasting of renewable energy generation—With application to solar energy. *IEEE Transactions on Power Systems,* 31**,** 3850-3863.

Goodfellow, I.,Bengio, Y. & Courville, A. 2016. *Deep learning*, MIT press.

Guermoui, M.,Melgani, F.,Gairaa, K. & Mekhalfi, M. L. 2020. A comprehensive review of hybrid models for solar radiation forecasting. *Journal of Cleaner Production,* 258**,** 120357.

Hocaoğlu, F. O. 2011. Stochastic approach for daily solar radiation modeling. *Solar Energy,* 85**,** 278-287.

Hochreiter, S. & Schmidhuber, J. 1997. Long short-term memory. *Neural computation,* 9**,** 1735-1780.

Huynh, A. N.-L.,Deo, R. C.,An-Vo, D.-A.,Ali, M.,Raj, N. & Abdulla, S. 2020. Near real-time global solar radiation forecasting at multiple time-step horizons using the long short-term memory network. *Energies,* 13**,** 3517.

IEA 2012. World Energy Outlook 2016 Executive Summary. International Energy Agency.

Jieni, X. & Zhongke, S. 2008. Short-time traffic flow prediction based on chaos time series theory. *Journal of Transportation Systems Engineering and Information Technology,* 8**,** 68-72.

Kaplani, E. & Kaplanis, S. 2012. A stochastic simulation model for reliable PV system sizing providing for solar radiation fluctuations. *Applied Energy,* 97**,** 970-981.

Kashyap, Y.,Bansal, A. & Sao, A. K. 2015. Solar radiation forecasting with multiple parameters neural networks. *Renewable and Sustainable Energy Reviews,* 49**,** 825-835.

Khelifi, R.,Guermoui, M.,Rabehi, A. & Lalmi, D. 2020. Multi-step-ahead forecasting of daily solar radiation components in the Saharan climate. *International Journal of Ambient Energy,* 41**,** 707-715.

Khodayar, M.,Mohammadi, S.,Khodayar, M. E.,Wang, J. & Liu, G. 2019. Convolutional graph autoencoder: a generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting. *IEEE Transactions on Sustainable Energy*.

Längkvist, M.,Karlsson, L. & Loutfi, A. 2014. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters,* 42**,** 11-24.

Lauret, P.,Boland, J. & Ridley, B. 2013. Bayesian statistical analysis applied to solar radiation modelling. *Renewable Energy,* 49**,** 124-127.

Lee, H. & Lee, B.-T. Bayesian deep learning-based confidence-aware solar irradiance forecasting system. 2018 International Conference on Information and Communication Technology Convergence (ICTC), 2018. IEEE, 1233-1238.

Lee, W.,Kim, K.,Park, J.,Kim, J. & Kim, Y. 2018. Forecasting solar power using long-short term memory and convolutional neural networks. *IEEE Access,* 6**,** 73068-73080.

Li, F.-F.,Wang, S.-Y. & Wei, J.-H. 2018. Long term rolling prediction model for solar radiation combining empirical mode decomposition (EMD) and artificial neural network (ANN) techniques. *Journal of Renewable and Sustainable Energy,* 10**,** 013704.

Liu, W.,Wang, Z.,Liu, X.,Zeng, N.,Liu, Y. & Alsaadi, F. E. 2017a. A survey of deep neural network architectures and their applications. *Neurocomputing,* 234**,** 11-26.

Liu, Z.,Jin, Y.,Zuo, M. J. & Feng, Z. 2017b. Time-frequency representation based on robust local mean decomposition for multicomponent AM-FM signal analysis. *Mechanical Systems and Signal Processing,* 95**,** 468-487.

Lorenz, E.,Hurka, J.,Heinemann, D. & Beyer, H. G. 2009. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *IEEE Journal of selected topics in applied earth observations and remote sensing,* 2**,** 2-10.

Luong, N. D. 2015. A critical review on energy efficiency and conservation policies and programs in Vietnam. *Renewable and Sustainable Energy Reviews,* 52**,** 623-634.

Martín, L.,Zarzalejo, L. F.,Polo, J.,Navarro, A.,Marchante, R. & Cony, M. 2010. Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning. *Solar Energy,* 84**,** 1772-1781.

Martinez-Anido, C. B.,Botor, B.,Florita, A. R.,Draxl, C.,Lu, S.,Hamann, H. F. & Hodge, B.-M. 2016. The value of day-ahead solar power forecasting improvement. *Solar Energy,* 129**,** 192-203.

Mccormick, P. & Suehrcke, H. 2018. The effect of intermittent solar radiation on the performance of PV systems. *Solar Energy,* 171**,** 667-674.

Mellit, A. & Pavan, A. M. 2010. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. *Solar Energy,* 84**,** 807-821.

Ming, D. & Ningzhou, X. 2011. A method to forecast short-term output power of photovoltaic generation system based on Markov chain. *Power System Technology,* 35**,** 152-157.

Moreno-Munoz, A.,De La Rosa, J.,Posadillo, R. & Bellido, F. Very short term forecasting of solar radiation. 2008 33rd IEEE Photovoltaic Specialists Conference, 2008. IEEE, 1-5.

Mostafavi, E. S.,Ramiyani, S. S.,Sarvar, R.,Moud, H. I. & Mousavi, S. M. 2013. A hybrid computational approach to estimate solar global radiation: an empirical evidence from Iran. *Energy,* 49**,** 204-210.

Muhammad, A.,Lee, J. M.,Hong, S. W.,Lee, S. J. & Lee, E. H. Deep Learning Application in Power System with a Case Study on Solar Irradiation Forecasting. 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2019. IEEE, 275-279.

Mukherjee, A.,Ain, A. & Dasgupta, P. Solar Irradiance Prediction from Historical Trends Using Deep Neural Networks. 2018 IEEE International Conference on Smart Energy Grid Engineering (SEGE), 2018. IEEE, 356-361.

Nobre, A. M.,Severiano Jr, C. A.,Karthik, S.,Kubis, M.,Zhao, L.,Martins, F. R.,Pereira, E. B.,Rüther, R. & Reindl, T. 2016. PV power conversion and short-term forecasting in a tropical, densely-built environment in Singapore. *Renewable Energy,* 94**,** 496-509.

Nova, J. C.,Cunha, J. B. & De Moura Oliveira, P. Solar irradiation forecast model using time series analysis and sky images. Proceedings of the 5th Conference of the European Federation for Information Technology in Agriculture, Food and Environment, 2005. 1408-1415.

Paulescu, E. & Blaga, R. 2016. Regression models for hourly diffuse solar radiation. *Solar Energy,* 125**,** 111-124.

Paulescu, M. & Paulescu, E. 2019. Short-term forecasting of solar irradiance. *Renewable Energy,* 143**,** 985-994.

Perez, R.,Hoff, T.,Dise, J.,Chalmers, D. & Kivalov, S. 2014. The cost of mitigating short-term PV output variability. *Energy Procedia,* 57**,** 755-762.

Peters, A. W.,Pyda, J.,Menon, G.,Suzuki, E. & Meara, J. G. 2019. The World Bank Group: Innovative financing for health and opportunities for global surgery. *Surgery,* 165**,** 263-272.

Pierro, M.,Bucci, F.,Cornaro, C.,Maggioni, E.,Perotto, A.,Pravettoni, M. & Spada, F. 2015. Model output statistics cascade to improve day ahead solar irradiance forecast. *Solar Energy,* 117**,** 99-113.

Polo, J.,Bernardos, A.,Martínez, S. & Peruchena, C. F. 2015. Maps of solar resource and potential in Vietnam.

Qian, Z.,Pei, Y.,Zareipour, H. & Chen, N. 2019. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. *Applied Energy,* 235**,** 939-953.

Qing, X. & Niu, Y. 2018. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy,* 148**,** 461-468.

Ramedani, Z.,Omid, M.,Keyhani, A.,Shamshirband, S. & Khoshnevisan, B. 2014. Potential of radial basis function based support vector regression for global solar radiation prediction. *Renewable and Sustainable Energy Reviews,* 39**,** 1005-1011.

Rao, K. S. K.,Rani, B. I. & Ilango, G. S. Estimation of daily global solar radiation using temperature, relative humidity and seasons with ANN for Indian stations. 2012 International Conference on Power, Signals, Controls and Computation, 2012. IEEE, 1-6.

Raza, M. Q.,Nadarajah, M. & Ekanayake, C. 2016. On recent advances in PV output power forecast. *Solar Energy,* 136**,** 125-144.

Ren, Y.,Suganthan, P. & Srikanth, N. 2015. Ensemble methods for wind and solar power forecasting—A state-of-the-art review. *Renewable and Sustainable Energy Reviews,* 50**,** 82-91.

Rigler, E.,Baker, D.,Weigel, R.,Vassiliadis, D. & Klimas, A. 2004. Adaptive linear prediction of radiation belt electrons using the Kalman filter. *Space Weather,* 2**,** 1-9.

Ruiz-Arias, J.,Alsamamra, H.,Tovar-Pescador, J. & Pozo-Vázquez, D. 2010. Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. *Energy Conversion and Management,* 51**,** 881-893.

Ryu, A.,Ito, M.,Ishii, H. & Hayashi, Y. Preliminary Analysis of Short-term Solar Irradiance Forecasting by using Total-sky Imager and Convolutional Neural Network. 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia), 2019. IEEE, 627-631.

Sak, H.,Senior, A. & Beaufays, F. 2014. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. *arXiv preprint arXiv:1402.1128*.

Sfetsos, A. & Coonick, A. 2000. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. *Solar Energy,* 68**,** 169-178.

Shem, C.,Simsek, Y.,Hutfilter, U. F. & Urmee, T. 2019. Potentials and opportunities for low carbon energy transition in Vietnam: A policy analysis. *Energy Policy,* 134**,** 110818.

Siddiqui, T. A.,Bharadwaj, S. & Kalyanaraman, S. A deep learning approach to solar-irradiance forecasting in sky-videos. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), 2019. IEEE, 2166-2174.

Sivaneasan, B.,Yu, C. & Goh, K. 2017. Solar forecasting using ANN with fuzzy logic pre-processing. *Energy procedia,* 143**,** 727-732.

Sloughter, J. M.,Gneiting, T. & Raftery, A. E. 2010. Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *Journal of the american statistical association,* 105**,** 25-35.

Srivastava, R.,Tiwari, A. & Giri, V. 2019. Solar radiation forecasting using MARS, CART, M5, and random forest model: A case study for India. *Heliyon,* 5**,** e02692.

Srivastava, S. & Lessmann, S. 2018. A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. *Solar Energy,* 162**,** 232-247.

Steyerberg, E. W.,Vickers, A. J.,Cook, N. R.,Gerds, T.,Gonen, M.,Obuchowski, N.,Pencina, M. J. & Kattan, M. W. 2010. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.),* 21**,** 128.

Sun, Y.,Venugopal, V. & Brandt, A. R. Convolutional neural network for short-term solar panel output prediction. 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)(A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC), 2018. IEEE, 2357-2361.

Tan, Z.,Zhang, H.,Xu, J.,Wang, J.,Yu, C. & Zhang, J. 2012. Photovoltaic power generation in China: Development potential, benefits of energy conservation and emission reduction. *Journal of Energy Engineering,* 138**,** 73-86.

Tiwari, M. K. & Chatterjee, C. 2011. A new wavelet–bootstrap–ANN hybrid model for daily discharge forecasting. *Journal of Hydroinformatics,* 13**,** 500-519.

Vaz, A.,Elsinga, B.,Van Sark, W. & Brito, M. 2016. An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht, the Netherlands. *Renewable Energy,* 85**,** 631-641.

Voyant, C.,Notton, G.,Kalogirou, S.,Nivet, M.-L.,Paoli, C.,Motte, F. & Fouilloy, A. 2017. Machine learning methods for solar radiation forecasting: A review. *Renewable Energy,* 105**,** 569-582.

Wang, F.,Yu, Y.,Zhang, Z.,Li, J.,Zhen, Z. & Li, K. 2018a. Wavelet decomposition and convolutional LSTM networks based improved deep learning model for solar irradiance forecasting. *Applied Sciences,* 8**,** 1286.

Wang, K.,Qi, X. & Liu, H. 2019. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Applied Energy,* 251**,** 113315.

Wang, S. & Jiang, J. 2015. Learning natural language inference with LSTM. *arXiv preprint arXiv:1512.08849*.

Wang, W.-C.,Chau, K.-W.,Xu, D.-M. & Chen, X.-Y. 2015. Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resources Management,* 29**,** 2655-2675.

Wang, Y.,Shen, Y.,Mao, S.,Chen, X. & Zou, H. 2018b. LASSO and LSTM integrated temporal model for short-term solar intensity forecasting. *IEEE Internet of Things Journal,* 6**,** 2933-2944.

Wang, Y. & Wu, L. 2016. On practical challenges of decomposition-based hybrid forecasting algorithms for wind speed and solar irradiation. *Energy,* 112**,** 208-220.

Wang, Z.,Tian, C.,Zhu, Q. & Huang, M. 2018c. Hourly solar radiation forecasting using a volterra-least squares support vector machine model combined with signal decomposition. *Energies,* 11**,** 68.

Wenbin, H.,Ben, H. & Changzhi, Y. 2002. Building thermal process analysis with grey system method. *Building and Environment,* 37**,** 599-605.

Werbos, P. 1974. Beyond regression:" new tools for prediction and analysis in the behavioral sciences. *Ph. D. dissertation, Harvard University*.

Wittmann, M.,Breitkreuz, H.,Schroedter-Homscheidt, M. & Eck, M. 2008. Case studies on the use of solar irradiance forecast for optimized operation strategies of solar thermal power plants. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 1**,** 18-27.

Yang, H.-T.,Huang, C.-M.,Huang, Y.-C. & Pai, Y.-S. 2014. A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output. *IEEE transactions on sustainable energy,* 5**,** 917-926.

Yona, A.,Senjyu, T.,Funabashi, T.,Mandal, P. & Kim, C.-H. 2013. Decision technique of solar radiation prediction applying recurrent neural network for short-term ahead power output of photovoltaic system. *smart grid and renewable energy,* 4**,** 32.

Yu, Y.,Cao, J. & Zhu, J. 2019. An LSTM Short-Term Solar Irradiance Forecasting Under Complicated Weather Conditions. *IEEE Access,* 7**,** 145651-145666.

Zamora, R. J.,Dutton, E. G.,Trainer, M.,Mckeen, S. A.,Wilczak, J. M. & Hou, Y.-T. 2005. The accuracy of solar irradiance calculations used in mesoscale numerical weather prediction. *Monthly weather review,* 133**,** 783-792.

Zang, H.,Cheng, L.,Ding, T.,Cheung, K. W.,Liang, Z.,Wei, Z. & Sun, G. 2018. Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network. *IET Generation, Transmission & Distribution,* 12**,** 4557-4567.

Zang, H.,Liu, L.,Sun, L.,Cheng, L.,Wei, Z. & Sun, G. 2020. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. *Renewable Energy,* 160**,** 26-41.

Zhang, J.,Zhao, L.,Deng, S.,Xu, W. & Zhang, Y. 2017. A critical review of the models used to estimate solar radiation. *Renewable and Sustainable Energy Reviews,* 70**,** 314-329.

Zhou, H.,Zhang, Y.,Yang, L.,Liu, Q.,Yan, K. & Du, Y. 2019. Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism. *IEEE Access,* 7**,** 78063-78074.

Zhu, G.,Zhang, L.,Shen, P. & Song, J. 2017. Multimodal gesture recognition using 3-D convolution and convolutional LSTM. *Ieee Access,* 5**,** 4517-4524.