Case Studies in Thermal Engineering 86 (2026) 108445

Contents lists available at ScienceDirect

Case Studies in Thermal Engineering

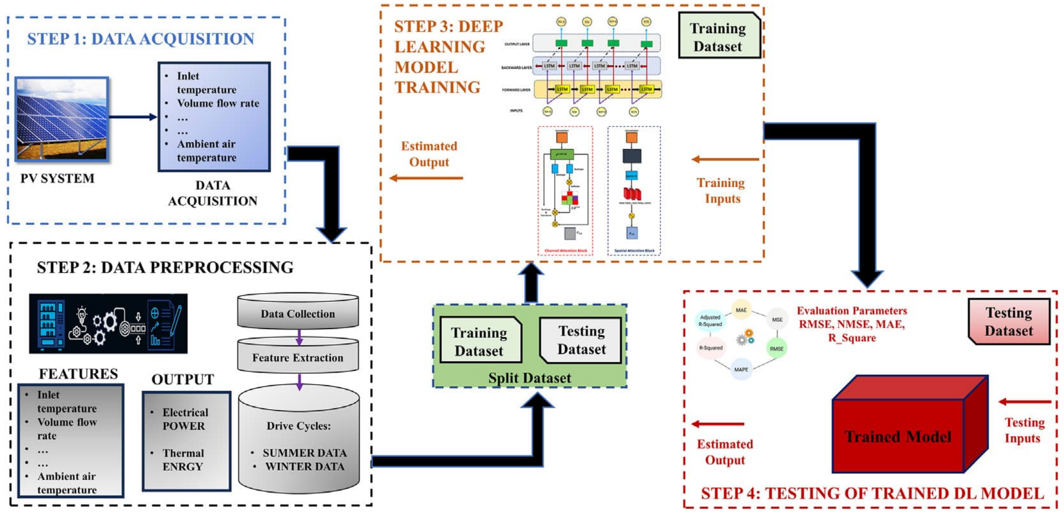
journal homepage: www.elsevier.com/locate/csite

Multi-task learning based deep neural network for short term power and thermal energy prediction of hybrid PV–Thermal systems

Noman Mujeeb Khan [a], Qusay Shihab Hamad [b], Mohamad Abou Houran [c], Muhammad Kamran Khan [d], Muhammad Hamza Zafar [e], Filippo Sanfilippo [e,f,*]

[a] Department of Electrical Engineering, Capital University of Science and Technology, Islamabad, 44000, Pakistan
[b] University of information technology and communications, Baghdad, 3071, Iraq
[c] School of Electrical Engineering, Xi'an Jiaotong University, No. 28 , West Xianning Road, Xi'an, 710049, China
[d] School of Technology and Innovation, University of Vaasa, Vaasa, 65200, Finland
[e] Department of Engineering Sciences, University of Agder, Grimstad, 4879, Norway
[f] Department of Software Engineering, Kaunas University of Technology, Kaunas, 51368, Lithuania

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Hybrid PV–thermal
Power forecasting
Thermal energy prediction

ABSTRACT

Accurate forecasting of power output and thermal energy in hybrid Photovoltaic–Thermal (PVT) plants is critical for optimising energy management and maximising efficiency. Traditional models often struggle to capture complex temporal dependencies and non-linear patterns in such systems. We propose the Inception Embedded Dual Attention Bi-LSTM (IDAMBi-LSTM)

* Corresponding author.
E-mail address: filippo.sanfilippo@uia.no (F. Sanfilippo).

https://doi.org/10.1016/j.csite.2026.108445
Received 1 August 2025; Received in revised form 28 December 2025; Accepted 17 August 2026
Available online 25 August 2026

Multi-task learning
Inception module

Model for enhanced power forecasting and thermal energy prediction in hybrid PVT plants to address these challenges. The model integrates inception layers to capture multi-scale temporal features, dual attention mechanisms to emphasise significant data points, and Bidirectional Long Short-Term Memory (Bi-LSTM) networks for handling sequential dependencies in both forward and backward directions. We conducted a comparative analysis with an Inception Embedded Attention Mechanism with GRU (IAM-GRU) and Inception Embedded Dual Attention GRU (IDAM-GRU) model to evaluate the performance and accuracy of our proposed approach. The evaluation was based on several performance metrics, including Root Mean Square Error (RMSE), R-squared (R^2), Normalised Mean Square Error (NMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE). The results demonstrate that the proposed Inception Embedded Dual Attention Bi-LSTM Model outperforms the IAM-GRU model across all metrics, providing more accurate and reliable forecasts for both power output and thermal energy in hybrid PVT systems. This research highlights the potential of advanced deep learning techniques in improving the efficiency and predictability of renewable energy systems.

Nomenclature	
Acronyms	
PV	Photovoltaic
PVT	Photovoltaic-Thermal
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
Bi-LSTM	Bidirectional Long Short-Term Memory
GRU	Gated Recurrent Unit
GAN	Generative Adversarial Network
MTL	Multi-Task Learning
IDAMBi-LSTM	Inception Embedded Dual Attention Bi-LSTM
IAM-GRU	Inception Embedded Attention Mechanism with GRU
IDAM-GRU	Inception Embedded Dual Attention GRU
RMSE	Root Mean Square Error
NMSE	Normalised Mean Square Error
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
R^2	Coefficient of Determination
Greek Letters	
α_s	Spatial attention weights
α_t	Temporal attention weights
Variables	
P_{el}	Electrical power output (W)
$I_{g,t}$	Global tilted irradiance (W/m^2)
I_g	Global horizontal irradiance (W/m^2)
I_b	Direct beam radiation (W/m^2)
I_d	Diffuse radiation (W/m^2)
T_{PV}	Photovoltaic temperature (°C)
T_{back}	Back surface temperature (°C)
T_{air}	Ambient air temperature (°C)
$T_{Fluid,in}$	Fluid inlet temperature (°C)
T_{Sky}	Sky temperature (°C)
ΔT	Temperature difference (Inlet-Outlet) (°C)
$\upsilon_{wind,h}$	Wind speed in horizontal plane (m/s)
$\upsilon_{wind,t}$	Wind speed in tilted plane (m/s)
$\dot{m}$	Fluid mass flow rate (lph)
$d_{wind,h}$	Wind direction in horizontal plane (°)

1. Introduction

Solar energy plays a crucial role in the global pursuit of sustainable energy solutions. It addresses pivotal challenges in energy sustainability, particularly in reducing reliance on non-renewable resources. As nations grapple with the dual demands of increasing energy needs and the imperative to curb carbon emissions, solar energy emerges as a key player. It offers a unique blend of environmental stewardship and energy efficiency, positioning itself not just as an alternative but as a fundamental pillar in the transition to a low-carbon economy. The widespread adoption and development of solar technology underscores its potential to significantly influence energy policies and practices, affirming its essential role in the future global energy portfolio. The intrinsic variability of solar energy poses significant challenges, particularly in the realms of prediction and management. Factors such as solar zenith angle, daylight duration, and atmospheric conditions lead to fluctuations in solar output, both daily and seasonally. This unpredictability necessitates the development of robust and sophisticated prediction models to ensure the reliability and efficiency of solar power systems. Beyond the technical aspects, these challenges extend to the realms of planning and policy-making, requiring a nuanced approach to integrate solar energy's intermittent nature within the broader energy infrastructure.

Accurate solar energy prediction models are essential not only for technical reasons but also for their economic and policy implications. These models play a critical role in optimising the performance of solar power plants, guiding infrastructure development, and managing electricity grids. They are central to strategic decision-making processes, influencing everything from operational management at the micro-level to energy policy formulation at the macro-level. In the financial landscape of solar projects, the precision of these models significantly affects investment decisions, risk assessment, and return on investment calculations. As solar energy continues to grow in importance in the global energy mix, refining and enhancing these prediction models become increasingly vital. They are key not just for the technical feasibility of solar projects, but also for their economic viability and seamless integration into existing energy frameworks. Photovoltaic (PV) power forecasting is an effective means to get over these obstacles. Large-scale PV integration into the main power grid is acknowledged to need accurate PV power output predictions [1]. On the other hand, the PV power time series typically displays unstable and nonlinear behaviour. Accurate PV power forecasting is challenging as PV power output is dependent on unexpected weather conditions. At the moment, research is being done on predicting PV power output. Many forecasting techniques have been put forth. Depending on the forecasting horizon, these techniques may be categorised into three categories: long-term, middle-term, and short-term PV power forecasting [2].

Long-term PV power forecasting guarantees dependable power system operation and aids in long-term planning and decision-making for PV power production, transmission, and distribution [3]. Middle-term PV power forecasting supports middle-term power system dispatching decisions [4]. Short-term PV power forecasting can help the power system function and hence increase power system dependability [5]. Physical, persistence, and statistical models are more categories into which PV power forecasting models may be subdivided [6]. The physical state and dynamic motion of meteorological conditions are described by physical approaches using mathematical equations. Physically based forecasting models work best under steady weather conditions. The majority of persistence approaches operate on the assumption that present and future values are highly correlated. The time series' future values are computed with the assumption that the circumstances will not change over time. Historical average values are the main determinant of predicting accuracy for models built using persistence approaches [7]. In contrast, statistically based models seek to quantify the correlation between historical PV power output and meteorological data. Typically, historical variables and the forecasting model's learning process serve as the foundation for statistical procedures. The time span and the caliber of the input data have a significant impact on the performance of statistical procedures [8]. Autoregressive integrated moving average (ARIMA) models [9], autoregressive and moving average (ARMA) models [10], regression method (RM) [11], and extreme learning machine (ELM) [12] are some of the statistical techniques that are often used in PV power forecasting.

Statistical models typically generate more accurate short-term PV power forecasting results because they take into account past PV generation levels and constantly adjust the model parameters however, there are still some glaring limitations. Time series data used in ARMA models must be stationary [13], while regression-based forecasting models (RM) require domain expertise to identify relevant explanatory variables and specify their functional relationships—a challenging and time-intensive process when dealing with the non-linear, multi-dimensional dependencies characteristic of PV systems. Researchers have created and implemented deep learning-based models in a variety of fields as artificial intelligence (AI) approaches have advanced rapidly [14]. It is a novel branch of machine learning techniques. Some deep learning-based models, such as convolutional neural networks (CNN) [15] and recurrent neural networks (RNN) [16], have previously demonstrated promising performance in PV power forecasting. Deep learning models, as opposed to traditional physical, persistence, and statistical techniques, may extract deep information from PV power series and produce more accurate predicting results. Recent improvements and promotion of smart grid technologies have enabled governments all over the world to use distributed energy resources (DERs). PV electricity and other renewable energy have the potential to become the dominant end-use energy sources in the future [17]. As PV power penetration increases in the energy system, there is a greater requirement for real-time dependability and security of PV power systems. Thus, more efficient PV power forecasting models with improved accuracy are urgently required. Many earlier research have emphasised structural improvement and parameter tweaking of forecasting models. Following improvement, these models outperform standard predicting challenges. Some conventional AI-based forecasting models may become trapped in local minimums and fail to deliver adequate results due to the very unstable nature of PV power generation, particularly on overcast and wet days [18] because of the PV power series' non-linear and non-stationary characteristics, reliable and stable prediction values cannot be produced if the original PV power series is directly applied to the forecasting model, particularly in the presence of unstable meteorology. Traditional methodologies cannot provide the increased forecasting precision required for future energy systems. Some studies use deep learning approaches to increase the accuracy of PV power forecasts or relevant parameters in forecasting (for example, solar irradiation) [19]. Meteorological conditions that are directly connected to PV power production are overlooked by these models, which solely concentrate on the PV power series itself. These models were unable to accurately predict changes in PV power over short time periods, especially in areas with a varied environment. PV power production exhibits seasonality and periodicity, but it also frequently changes dramatically on overcast and wet days, making it challenging to capture this variability using solely historical PV data [20].

Existing PVT forecasting approaches suffer from three critical limitations. First, single-task models predict either power or thermal output independently, ignoring the strong thermal–electrical coupling where PV temperature directly impacts efficiency. This results

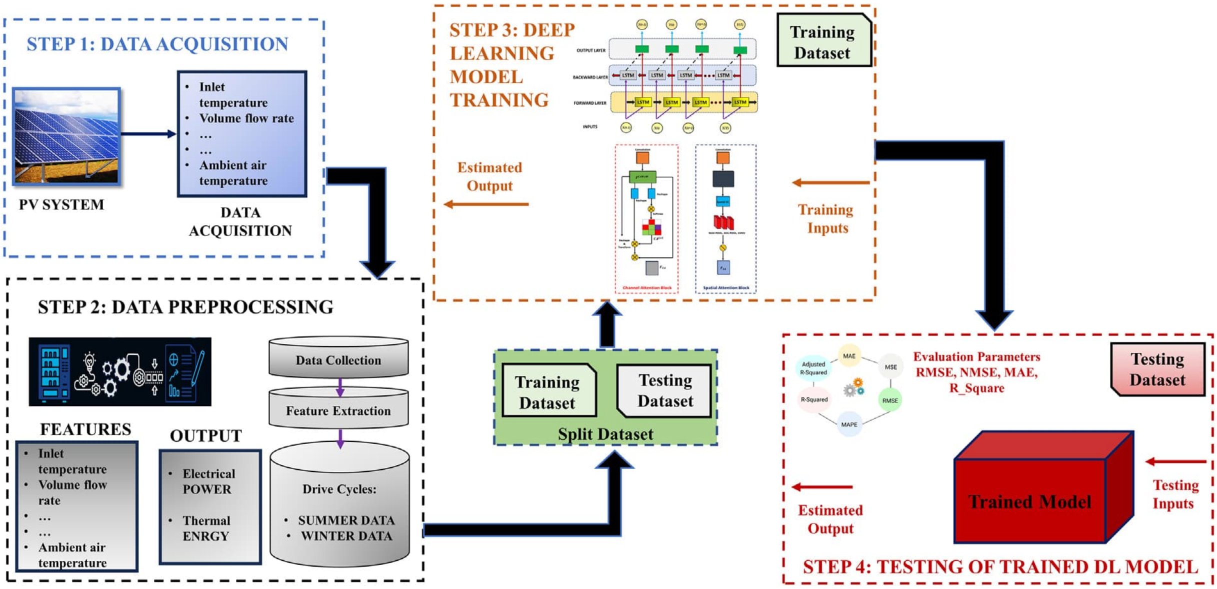


Fig. 1. Proposed Scheme of the Multi-Task Learning based Model for PV Power and Thermal Energy Forecasting.

in suboptimal predictions when thermal and electrical behaviours are interdependent. Second, current attention mechanisms apply uniform attention across all temporal–spatial dimensions, failing to capture that spatial features (panel temperature distribution) and temporal features (irradiance patterns) require different attention weights—a limitation our dual attention mechanism addresses. Third, existing models show significant performance degradation under variable weather conditions, with MAPE increasing from 6% (clear days) to 15% (cloudy days) in prior work, limiting practical deployment. Our work directly addresses these three gaps through the first multi-task architecture with dual attention specifically designed for coupled thermal–electrical prediction in PVT systems (see Fig. 1).

To bridge this gap, Inception Embedded Dual Attention Multi BI-LSTM Model has been proposed in this study which represents a significant leap in the field of solar energy forecasting, particularly in predicting electrical power and thermal energy outputs. The key contributions of this work are as follows:

1. **Novel Multi-Task Architecture with Dual Attention:** We introduce the first deep learning architecture that simultaneously optimises both electrical power and thermal energy predictions in hybrid PVT systems through shared feature representations. Unlike existing single-task approaches, our IDAMBi-LSTM model uniquely integrates: (a) inception modules capturing multi-scale temporal patterns across 1×1, 3×3, and 5×5 convolutions simultaneously, (b) dual attention mechanisms operating independently on spatial and temporal dimensions (Eqs. (3)–(5))—absent in prior PVT forecasting models, and (c) bidirectional LSTM networks that leverage both past and future context to capture thermal–electrical coupling effects where temperature directly influences electrical output. This addresses the fundamental limitation in prior work that treats power and thermal predictions as independent tasks, missing inherent interdependencies revealed in our correlation analysis.
2. **Rigorous Ablation Study Demonstrating Component Contributions:** We conduct systematic ablation studies comparing: (a) single vs. dual attention mechanisms (IDAMBi-LSTM vs. IAM-GRU), (b) bidirectional vs. unidirectional processing (Bi-LSTM vs. GRU), and (c) seasonal robustness across summer and spring conditions. Our analysis reveals that dual attention reduces RMSE by 35.8% for thermal energy (13.619 vs. 21.200) and 52.6% for electrical power (1.85 vs. 3.90) compared to single attention approaches (Tables 4–5). Critically, bidirectional processing outperforms GRU variants across all five metrics (RMSE, R^2, NMSE, MAPE, MAE) in both prediction tasks, with R^2 improvements from 95.60% (IDAM-GRU) to 99.79% (IDAMBi-LSTM) for thermal energy. This comprehensive evaluation spanning two distinct seasonal periods with different irradiance patterns provides robust evidence of model generalisation beyond single-scenario testing common in prior work.
3. **State-of-the-Art Performance with Practical Significance:** The IDAMBi-LSTM achieves near-perfect prediction accuracy (R^2 = 99.99% for electrical power, 99.79% for thermal energy in summer), representing a 2.29 percentage point improvement over the best competing model (IAM-GRU: R^2 = 97.70% and 97.50%). In absolute terms, this translates to: (a) MAE reduction of 35.6% for electrical power (7.377 W vs. 11.500 W) and 36.9% for thermal energy (3.725 vs. 5.900), (b) MAPE reduction to 4.05%—approaching the operational threshold required for real-time grid integration, and (c) consistent performance across seasons (summer RMSE-E: 1.85 vs. spring RMSE-E: 1.74), demonstrating robustness to varying environmental conditions. Unlike prior work showing performance degradation under variable weather, our model maintains RMSE $<$ 2.0 across all tested conditions, addressing a critical gap for reliable PVT system deployment.

Table 1
Summary of relevant works in power system forecasting.

Ref.	Year	Model	Summary	Metric results	Lacks
[21]	2019	Deep LSTM networks for PV power prediction	Uses LSTM to predict PV output power, captures temporal changes effectively, compares five LSTM models.	Outperforms multiple linear regression, bagged regression trees, and other neural network techniques.	Requires comparison with more advanced models and application in hybrid PV-T systems.
[22]	2019	LSTM for PV fault diagnosis	Uses LSTM for automatic feature extraction in fault diagnosis, achieves high accuracy on noisy and noiseless data.	Accuracy: NC (100%), HSF (100%), LLF (99.03%) on noiseless data; robust against noisy data.	No significant limitations mentioned, lacks focus on power and thermal energy prediction.
[23]	2020	Kernel Density Estimation specifically for ensemble selection in solar power forecasting	A novel ensemble algorithm using kernel density estimation (KDE) achieved accurate solar power forecasting by calculating similarity indices from historical weather data and power outputs, outperforming traditional methods while being 32 times faster than neural networks.	The proposed KDE-based ensemble model achieved MAEs of 12.6, 14.0, 3.6, and 7.7 kW for winter, spring, summer, and autumn respectively, with corresponding RMSDs of 1.115, 1.914, 0.523, and 0.928 kW, while maintaining the same values for NRMSD percentages and processing all seasons in just 0.045 s.	No significant limitations mentioned, lacks focus on power and thermal energy prediction.
[24]	2022	LSTM for short-term PV power prediction	Uses LSTM RNNs for short-term PV power prediction, trained on one-year data, compared with MLP.	Outperforms MLP in MAE, MAPE, RMSE, R2.	Needs application to different datasets for generalisation and lacks dual attention mechanism.
[25]	2022	LSTM for electricity demand forecasting	Uses LSTM for predicting electricity demand, optimised with CVOA metaheuristic, highly accurate predictions.	Errors below 1.5%, outperforms linear regression, decision trees, tree ensembles, and other deep learning models.	Requires validation on more diverse datasets and lacks focus on hybrid PV-T systems.
[26]	2022	CNN-LSTM for PV power forecasting	Combines CNN and LSTM for PV power forecasting, applied to real-world data, superior to other techniques.	Outperforms linear regression, k-nearest neighbours, decision trees, CNN, LSTM, and multilayer perceptron in RMSE, MAE, MAPE.	Needs analysis of impact on different environmental conditions and lacks focus on thermal energy prediction.
[27]	2022	Bi-LSTM for home microgrid management	Integrates Bi-LSTM with an optimisation algorithm for managing BESS, forecasts energy consumption and PV generation.	Reduces household electricity costs through optimal BESS scheduling.	Requires long-term validation and lacks integration of dual attention mechanism.

4. **Generalisable Multi-Task Learning Framework for Hybrid Energy Systems:** This work establishes a replicable methodological framework applicable beyond PVT systems to any hybrid energy technology with coupled outputs (e.g., combined heat and power, fuel cells, thermoelectric generators). Key methodological contributions include: (a) weighted sum loss function with tunable parameters (α, β) enabling domain-specific optimisation based on operational priorities, (b) GAN-based data imputation strategy that preserves inter-variable correlations—critical for maintaining thermal–electrical coupling relationships, demonstrated by maintaining correlation coefficients within 2% of original values, and (c) correlation-driven feature selection methodology revealing that temperature difference (ΔT = 0.54 correlation) and global tilted irradiance ($I_{g,t}$ = 1.00 correlation) are primary drivers. These innovations provide practitioners with: (i) a replicable approach for identifying key features in hybrid systems, (ii) a data preprocessing pipeline maintaining physical relationships, and (iii) an adaptable architecture template for different energy conversion technologies.

2. Literature review

In recent years, there have been limited research studies exploring the application of long short-term memory (LSTM) networks in the protection and prediction of power and thermal energy in hybrid PV thermal systems. This section aims to address this gap by providing a comprehensive review of the most relevant works that have utilised LSTM in power system forecasting across various domains. By examining these studies, readers will gain a comprehensive understanding of the existing literature, the approaches employed in utilising LSTM, and how these techniques have been implemented by different researchers. This review will serve as a valuable resource for researchers and practitioners seeking insights into the application of LSTM in power system forecasting and its potential implications. As summarised in Table 1, recent works in power system forecasting have various focuses and limitations.

Solar power forecasting has benefited significantly from the implementation of machine learning approaches, which excel at analysing complex datasets of substantial scale. Various algorithms such as decision trees, random forests, and support vector machines demonstrate versatility across different forecasting timeframes, ranging from immediate operational needs to extended strategic planning [28,29]. In the realm of concentrator solar power (CSP) systems, Artificial Neural Networks (ANNs) serve as effective tools for energy production prediction through the application of various learning methodologies [30]. When dealing with incomplete datasets, recurrent neural networks (RNNs) incorporating adaptive neural imputation modules show improved performance in solar irradiance (SI) forecasting [31]. The integration of neural networks with ripple current correlation methods has proven successful in enhancing photovoltaic system optimisation [32]. Additionally, ANNs find application in CSP system energy production forecasting through diverse learning approaches [33]. Advanced architectures including LSTM [34], Gated Recurrent Unit (GRU) [35], and Convolutional Networks (CNN) [36] demonstrate particular effectiveness in managing the inherent variability and uncertainty of SI, establishing them as fundamental components for improving solar energy forecast precision.

The integration of machine learning methods with temporal and spatial analysis through hybrid approaches provides enhanced forecasting performance [37]. ConvLSTM architectures [38] exemplify this approach by merging convolutional operations with LSTM functionality to process both spatial and temporal characteristics of SI data. Recent developments in Temporal Convolutional Networks (TCN) [39], utilising causal dilated convolution operations, represent significant progress in temporal data analysis capabilities crucial for precise SI forecasting [40]. Hybrid deep learning approaches extend beyond energy forecasting to evaluate sustainable transportation development, examining critical factors and their relationships [41]. However, these advanced modelling approaches encounter practical limitations, including requirements for substantial training datasets and complex architectural designs [42].

Abdel-Nasser and Mahmoud [21]proposed a novel method that utilises deep LSTM networks to accurately predict PV output power. Their methodology stands out by effectively capturing the temporal changes in PV power data, thanks to LSTM's recurrent architecture and memory units. To evaluate the performance of their proposed method, the authors compare five LSTM models with different architectures. Among these models, their third model (referred to as model3) incorporating time steps exhibits the best performance in PV power forecasting. This highlights the importance of considering temporal dynamics in accurately predicting PV power output. In addition, Abdel-Nasser and Mahmoud benchmark their proposed method against three commonly used PV forecasting methods based on multiple linear regression, bagged regression trees, and neural networks techniques. Their LSTM-based approach outperforms these alternative methods, demonstrating smaller forecasting errors. This emphasises the superiority of deep learning techniques, specifically LSTM networks, in accurately forecasting PV power. Overall, Abdel-Nasser and Mahmoud's work showcases the potential of LSTM networks for improving PV power forecasting. Their innovative methodology effectively captures temporal changes in PV data, leading to more accurate predictions. By outperforming traditional forecasting methods, their approach contributes to the advancement of renewable energy integration and highlights the value of deep learning in the field of PV power forecasting.

Appiah et al. [22] proposed a novel photovoltaic array fault diagnosis technique that leverages deep learning to automatically extract features from raw data, eliminating the need for manual feature engineering. The key contributions included: (1) Utilising LSTM networks as an automatic feature extractor, (2) Feeding the LSTM-extracted features into a softmax regression classifier for fault diagnosis, and (3) Demonstrating high fault diagnosis accuracies on both noisy and noiseless data. The proposed technique is capable of classifying three fault conditions: normal condition (NC), line-to-line fault (LLF), and hot spot fault (HSF). On the noiseless dataset, it achieves 100% accuracy for NC and HSF, and 99.03% for LLF. Even in the presence of noisy data, the technique demonstrates robustness, with average fault classification accuracies of 99.23%, 98.78%, and 97.66% for NC, HSF, and LLF, respectively. The authors concluded that the LSTM-based feature extraction approach effectively removes the need for manual, time-consuming, and expertise-demanding feature engineering, which is a common limitation of many machine learning-based PV fault diagnosis techniques. The results showcase the proposed technique's effectiveness in automatically extracting useful features from raw data and its superiority over other approaches in terms of fault diagnosis accuracy, robustness, and automation. This work presents a significant advancement in PV array fault diagnosis by leveraging deep learning to enable autonomous feature extraction, which can lead to more efficient and reliable PV system maintenance and operation.

In another work, Torres et al. [25] designed a deep neural network specifically for predicting electricity demand time series. They employed LSTM network architecture due to its capability to handle sequential data and retain temporal relationships in the long term. The authors conducted a random search and utilised the coronavirus optimisation algorithm (CVOA) metaheuristic to determine the optimal values for hyperparameters, including the number of layers, number of LSTM cells per layer, learning rate, and dropout rate. The optimised LSTM network was then applied to Spanish electricity demand data from 2007 to 2016, measured at a 10-minute frequency, to generate 24-hour forecasts. The results demonstrated highly accurate predictions, with errors below 1.5%. Furthermore, the proposed LSTM network outperformed other models, including linear regression, decision trees, tree ensembles, and two deep neural networks—a deep feed-forward neural network optimised through random search and a temporal fusion transformer (TFT) optimised using a sampling algorithm. Overall, this research showcases the effectiveness of the designed LSTM network in accurately forecasting electricity demand time series. The authors' use of both random search and the CVOA metaheuristic for hyperparameter optimisation, along with the comparison to various benchmark models, highlights the superior performance and potential of their proposed approach.

Agga et al. [26] proposed a hybrid deep learning model, combining CNN and LSTM, for accurate PV power output forecasting. The main contributions of this paper were: (1) Developing a CNN-LSTM architecture to leverage the strengths of both models in extracting spatial and temporal features from weather data, (2) Applying the proposed model to a real-world PV dataset from Rabat, Morocco, and (3) Demonstrating the proposed model's superior forecasting performance compared to standalone machine learning and deep learning models. The paper analysed the impact of look-back periods on the forecasting accuracy of different models, finding that longer look-back windows are essential for improving predictions as the forecasting horizon increases. The CNN-LSTM model outperformed other techniques, including linear regression, k-nearest neighbours, decision trees, CNN, LSTM, and multilayer perceptron, across various error metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). For example, the CNN-LSTM model showed MAPE improvements of 5.55%, 6.86%, and 6.49% over the standalone LSTM model for 1-day, 3-day, and 7-day ahead forecasts, respectively. The authors attributed the CNN-LSTM's superior performance to its ability to capture both the spatial and temporal properties of the input data, which are crucial for accurately predicting the erratic patterns of PV power generation.

Alam et al. [27] proposed a novel energy management system model for a home microgrid with a battery energy storage system (BESS). The model integrates a deep learning-based predictive component, using bidirectional long short-term memory (Bi-LSTM), with an optimisation algorithm to optimally schedule and manage the BESS. The key contributions of this research were the development of a Bi-LSTM model for accurate day-ahead forecasting of energy consumption and PV generation, the proposal of a heuristic strategy that considers the predicted energy profiles for real-time scheduling and optimisation, and the evaluation of the system under different scenarios of day-ahead power consumption and PV generation, going beyond the typical focus on present-day conditions. The optimisation aims to minimise daily electricity costs under time-of-use pricing while accounting for peak demand penalties and operational constraints of renewables, BESS, and appliances. The reported results demonstrate the proposed system's effectiveness in reducing household daily electricity costs through optimal BESS scheduling.

Caicedo-Vivas and Alfonso-Morales [43] introduced a methodology for short-term load forecasting that combines LSTM architecture with historical load data and calendar features. Their approach is notable for its adaptability to atypical load behaviour, such as that caused by the 2021 national strike. The LSTM model efficiently processes multiple past values and incorporates additional information, like holidays, to enhance accuracy. A key contribution is the use of bootstrapping for prediction intervals, providing valuable uncertainty quantification. Despite the challenges posed by the abnormal dataset, the model achieved impressive accuracy, offering significant insights for power system management and reliable service delivery.

Huang and Yang [44] developed the MLSTNet model, which integrates SLSTM and LSTNet models for improved ultra-short-term PV power forecasting. Enhancements include replacing LSTNet's CNN with a Temporal Convolutional Attention Neural Network (TCANN) and using ARIMA for forecasting. The model also incorporates a temporal attention mechanism for better feature extraction, allowing it to handle the stochastic nature of PV power data. Bayesian optimisation is employed for parameter tuning, improving adaptability and performance. While the MLSTNet model shows strong results, particularly in capturing spatiotemporal features, further improvements are needed in forecasting accuracy during non-sunny conditions and sudden weather changes.

Safari et al. [45] introduced FARHAN, a hybrid model combining descending neuron attention, LSTM, and Markov-simulated neural networks to enhance electrical load forecasting in smart grids. FARHAN's innovative structure, including two LSTM blocks with attention layers and a Markov chain analyser, significantly outperformed traditional methods in both accuracy and efficiency across various forecasting intervals. Meanwhile, Rafique et al. [46] addressed the need for rapid fault detection in power transmission systems by developing an unsupervised method using LSTM autoencoders. Their approach, which detects faults within 2 ms without labelled datasets, proved resilient under various conditions and suitable for edge device deployment. Together, these studies highlight the potential of advanced neural networks in improving smart grid operations and power transmission line protection.

Ashkan Safari [47] introduced DeepPhase, a hybrid framework combining LSTM and gradient-boosted regression (GBR) to predict three-phase electrical signals' current, voltage, and power. DeepPhase outperformed benchmark models like Bi-LSTM, KNN, and LSTM across key metrics, achieving a remarkably low MAE of 6.94E−5, MAPE of 0.07%, and RMSE of 0.000156, along with a high R^2 of 0.999959. These results highlight DeepPhase's exceptional accuracy in analysing complex electrical signals, offering significant potential for improving decision-making in electrical systems. Similarly, Houran et al. [48] presented a hybrid deep learning model, COA-CNN-LSTM, combining CNNs, LSTMs, and the Coati optimisation algorithm to forecast short-term wind and solar power generation. This model demonstrated superior performance, reducing RMSE by 0.5% for day-ahead and 5.8% for hour-ahead forecasts, with an nMAE of 4.6%, RE of 27%, and nRMSE of 6.2%. Rigorous statistical analyses validated its ability to capture the dynamics of renewable energy, making it a promising solution for improving grid stability and integrating clean energy. Both studies underscore the potential of hybrid deep learning models in advancing electrical signal prediction and renewable energy forecasting.

3. Data description and analysis

In this section, we delve into the intricate relationship between environmental factors and solar energy conversion efficiency through an extensive examination of the data acquired from a solar power installation. Our comprehensive data analysis is grounded in high-resolution temporal data recorded in Pinkafeld, Austria, during the peak solar irradiance period from July 11 to September 6, 2021 [49]. This dataset is not only a repository of raw numerical values but a narrative that reveals the complex interdependencies of solar power generation. We methodically analyse and preprocess this data to distil clear insights from the multifaceted interactions between myriad variables, such as solar irradiance, temperature gradients, and wind dynamics. Through meticulous data preprocessing, we correct for any discrepancies and prepare the dataset for a robust analytical approach that encompasses both statistical and machine learning techniques. The subsequent subsections will detail the nuanced processes of our data handling and the pivotal findings that emerge from our analysis, contributing significantly to the solar energy domain by enhancing our understanding of the underlying physical processes and improving the predictive performance of solar output models.

3.1. Data analysis

Building upon a well-defined data collection framework, the Data Analysis phase delves into the interdependencies and intrinsic behaviours exhibited by the PVT system under study. Utilising advanced statistical methods and machine learning techniques, we meticulously dissected the dataset, which was captured with an unparallelled temporal resolution, to unravel the complex dynamics

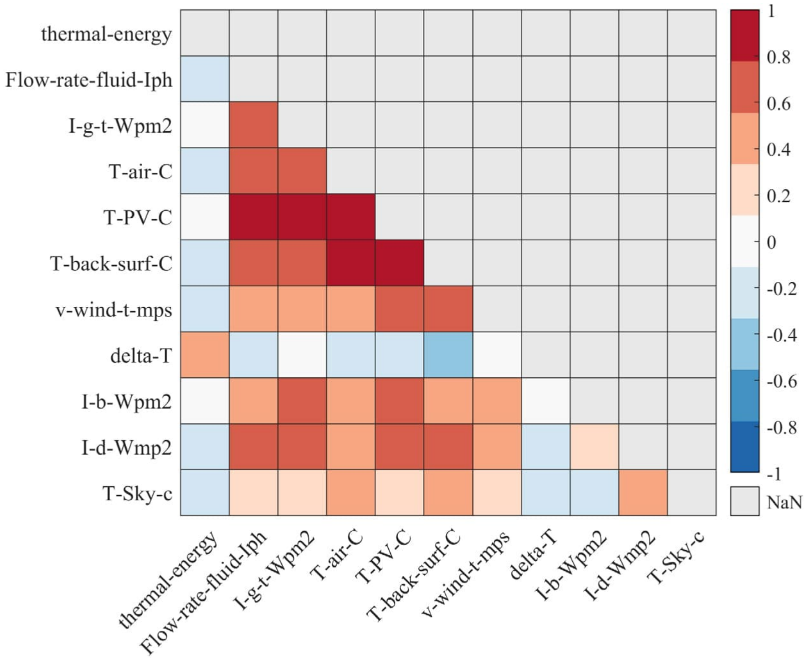


Fig. 2. Correlation Matrix of the Features with the thermal energy.

between environmental factors and solar energy output. Through a series of visual explorations, ranging from correlation matrices to time series decompositions, the analysis paints a nuanced picture of how variables such as solar irradiance, air temperature, and wind conditions interact with and influence the system's thermal and electrical efficiencies. This granular analysis not only confirms theoretical expectations but also exposes unexpected patterns, offering strategic insights into the PVT system's operation. The subsequent subsections will detail the empirical findings through visual representations, reinforcing the pivotal role of data-driven analysis in advancing the efficiency and dependability of solar energy systems (see Fig. 2).

3.1.1. Correlation analysis of thermal energy

The thermal energy output exhibits significant and revealing correlations with various system parameters. The strongest positive correlation is with the temperature difference (ΔT) at 0.54, which is expected as it directly relates to the heat transfer process in the system. However, most other parameters show moderate to strong negative correlations, painting an interesting picture of the system's behaviour. The flow rate ($\dot{V}$) shows a notable negative correlation of −0.32, suggesting that higher flow rates may actually reduce thermal energy output, possibly due to reduced heat transfer effectiveness at higher fluid velocities. The back surface temperature (T_{back}) and diffuse radiation (I_d) both show negative correlations of −0.24, indicating that increases in these parameters tend to decrease thermal energy production.

What is particularly noteworthy is the system's response to different types of solar radiation. While direct beam radiation (I_b) shows a very weak positive correlation (0.05) with thermal energy, global tilted irradiance ($I_{g,t}$) shows a slight negative correlation (−0.02), and diffuse radiation shows a stronger negative correlation (−0.24). This suggests that the system's thermal performance is more complex than simply responding to available solar radiation. The negative correlations with ambient temperature (T_{air}) at −0.16, PV temperature (T_{PV}) at −0.07, and sky temperature (T_{sky}) at −0.15 indicate that higher environmental and component temperatures generally reduce thermal energy output, likely due to increased thermal losses. The wind speed (v_{wind}) shows a negative correlation of −0.14, confirming that convective losses play a role in reducing thermal energy generation. These relationships suggest that the system's thermal performance is optimised under conditions of lower ambient temperatures and moderate flow rates, with temperature difference being the primary driver of thermal energy production (see Fig. 3).

3.1.2. Correlation analysis of solar power

The electrical power output ($P_{\mathrm{el,W}}$) shows strong positive correlations with several key parameters. Most notably, it has perfect correlation (1.00) with global tilted irradiance ($I_{g,t,\mathrm{W/m^2}}$), indicating that solar radiation directly incident on the panel is the primary driver of electrical generation. There are also very strong correlations with global horizontal irradiance ($I_{g,\mathrm{W/m^2}}$) at 0.94 and PV temperature ($T_{\mathrm{PV,C}}$) at 0.89, demonstrating the significant influence of both solar resource and thermal conditions on electrical performance. The fluid inlet temperature ($T_{\mathrm{Fluid,in,C}}$) shows a strong correlation of 0.79, suggesting that the thermal management of the system plays a crucial role in electrical generation.

The electrical output shows moderate positive correlations with diffuse radiation ($I_{d,\mathrm{W/m^2}}$) at 0.78 and direct beam radiation ($I_{b,\mathrm{W/m^2}}$) at 0.71, indicating that the system can effectively convert both direct and diffuse components of solar radiation. The

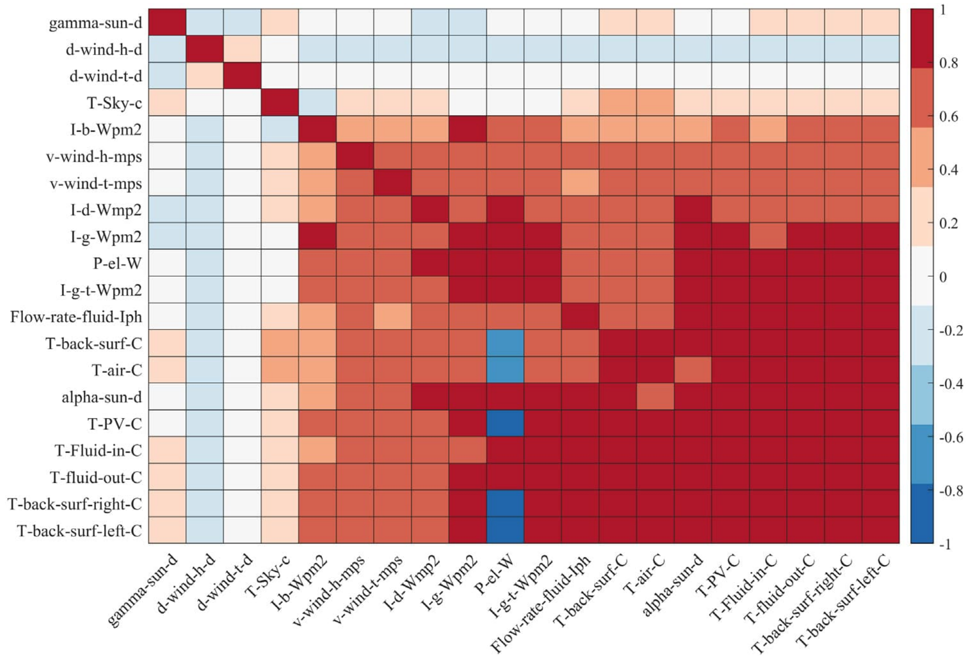


Fig. 3. Correlation Analysis of the Solar Power output with the Input Features.

Table 2
Summary of variables from hybrid PVT solar collector setup.

Collector	Symbol	Unit
Inlet temperature	T_{fluid_in}	°C
Volume flow rate	$flow\_rate\_fluid$	L/h
Collector back surface temperature	T_{back_surf}	°C
Collector back surface temperature (left)	$T_{back\_surf\_left}$	°C
Collector back surface temperature (right)	$T_{back\_surf\_right}$	°C
PV cell temperature	T_{PV}	°C
Collector output		
Outlet temperature	T_{fluid_out}	°C
PV current	I_{PV}	A
PV voltage	U_{PV}	V
Electrical power	P_{el}	W
Weather data at the collector		
Global tilted irradiance	I_{g_t}	W/m^2
Ambient air temperature	T_{air}	°C
Wind speed in the collector plane	v_{wind_t}	m/s
Wind direction in the collector plane	d_{wind_t}	°

flow rate (Flow_rate_fluid_lph) correlation of 0.76 suggests that proper cooling through fluid circulation positively impacts electrical performance. Wind-related parameters show varying degrees of correlation: wind speed in tilted plane ($v_{\mathrm{wind,t,m/s}}$) at 0.62 and horizontal wind speed ($v_{\mathrm{wind,h,m/s}}$) at 0.64, indicating moderate positive effects of convective cooling. Notably, the sky temperature ($T_{\mathrm{Sky,C}}$) shows a very weak positive correlation (0.10), suggesting minimal influence on electrical output. There is a weak negative correlation with wind direction parameters ($d_{\mathrm{wind,h,d}}$) at −0.24, indicating that wind direction has minimal impact on electrical performance (see Table 2).

3.1.3. Combined analysis and justification for multitask learning

The correlation analysis of both thermal energy and electrical power outputs reveals an intricate interplay of system parameters that justifies the implementation of a multitask learning approach. For thermal energy, the dominant positive correlation with temperature difference ($\Delta T = 0.54$) contrasts with generally negative correlations with other parameters, while electrical power shows strong positive correlations with solar radiation parameters ($I_{g,t} = 1.00$) and moderate to strong correlations with temperature components ($T_{PV} = 0.89$). This interconnected nature is particularly evident in shared influential parameters: the flow rate impacts both outputs (thermal: −0.32, electrical: 0.76), PV temperature affects both mechanisms (T_{PV} thermal: −0.07, electrical: 0.89), and solar irradiance influences both outputs though with varying strengths ($I_{g,t}$ thermal: −0.02, electrical: 1.00). These overlapping yet distinct correlation patterns strongly support the use of a multitask learning model for several reasons: (1) the shared input parameters suggest common feature representations that can be leveraged across both tasks, (2) the varying correlation strengths indicate task-specific patterns that can be learned simultaneously while sharing underlying representations, (3) the opposing correlations for some parameters (like flow rate) suggest complex interactions that can be better captured through joint learning, and (4) the thermal–electrical coupling evident in the temperature correlations indicates that learning one output can provide valuable information for predicting the other. Therefore, a multitask learning approach would not only capitalise on these inherent relationships but also potentially improve prediction accuracy by learning the subtle interactions between thermal and electrical performance simultaneously.

3.2. Data preprocessing

The preprocessing stage is a critical step in the curation of robust solar energy datasets, setting the stage for subsequent analysis and model development. It encompasses a series of procedures designed to address the common challenge of missing or incomplete data—a hurdle that can distort analytical outcomes and impair the performance of predictive models. As we navigate through the intricacies of data preparation, we explore traditional imputation methods, their limitations in handling complex solar datasets, and the advent of sophisticated techniques, such as Generative Adversarial Networks (GANs), that promise to revolutionise the approach to data imputation in the realm of solar energy analysis. The following subsections delve into the nuanced process of preparing solar dataset for rigorous analysis, highlighting the challenges, innovations, and strategic applications that underpin this foundational phase of data science.

3.2.1. Challenges in solar dataset preprocessing

Solar energy datasets are as detailed as they are critical, providing the essential data needed for predictive modelling in renewable energy systems. These datasets, however, are often marred by the occurrence of missing values due to various disruptions, such as sensor failures or transmission losses. The presence of missing data is a significant barrier to accurate analysis and can substantially impede the development of reliable solar energy systems. Traditional imputation methods, ranging from mean substitution to more sophisticated multiple imputation techniques, offer a basic framework for addressing these deficiencies but are not without their shortcomings, especially when applied to the complex nature of solar energy data.

3.2.2. GAN architecture for data imputation

In data science, particularly within the context of renewable energy, the challenge of imputing missing data is a recurring and critical task. For our solar energy dataset, which encompasses a myriad of environmental and operational measurements, the application of Generative Adversarial Networks (GANs) presents a novel and promising approach. The architecture of GANs for data imputation in this context employs a dual-network system: a Generator (G) and a Discriminator (D). The Generator is tasked with generating realistic imputations for missing data, while the Discriminator's role is to differentiate between the imputed data by G and the actual data present in the dataset. This adversarial process is underpinned by the premise that through iterative training, G will progressively improve its imputation capabilities to a level where D can no longer easily distinguish between real and imputed data.

The mathematical framework of our GAN model is rooted in the standard GAN objective, but with significant modifications to suit the task of imputing missing data in a complex dataset. The objective function is defined as follows (see Eq. (1)):

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x|M)] + \mathbb{E}_{z\sim p_z(z),m\sim p_m(m)}[\log(1 - D(G(z|m)))] \tag{1}$$

where, M represents a binary mask matrix that identifies the locations of missing data within the dataset. The variable z is an input noise vector, serving as the initial seed for data generation in G. This function encapsulates the adversarial dynamics of the GAN, where G strives to minimise this objective against the adversarial efforts of D, which aims to maximise it.

The design of the Generator (G) and Discriminator (D) networks is pivotal in the effective imputation of missing data in our dataset. The architecture of G is specifically engineered to handle the complex nature of the dataset, which comprises a blend of continuous and categorical variables. G incorporates a series of layers, each designed to process different data types effectively. These layers include dense layers for continuous variables and embedding layers for categorical variables, ensuring that the network can manage the diverse nature of the data. Similarly, the architecture of D is meticulously crafted to evaluate the authenticity of the completed datasets. D's structure is designed to assess not just individual data points but also the relationships and correlations between variables. This is crucial in maintaining the integrity and statistical properties of the dataset, which are essential for accurate modelling and analysis in solar energy applications.

The dataset used, is a rich collection of environmental and operational measurements, is critical for predicting solar energy outputs. However, this dataset is characterised by patterns of missing data across several variables, a common issue in real-world data collection. To prepare this dataset for the GAN-based imputation process, a comprehensive preprocessing strategy was employed. Firstly, continuous variables were normalised to ensure uniformity in scale, a crucial step for effective neural network training. Categorical variables, on the other hand, were subjected to appropriate encoding techniques. This encoding transformed categorical data into a format that could be effectively processed by neural networks. Additionally, missing data points within the dataset were masked. This masking involved creating a binary mask that identified the locations of missing data, which was crucial for the GAN's training process. The training of the GAN was an iterative and carefully monitored process. Both the Generator (G) and the

Discriminator (D) were updated in turns. In each iteration, G received batches of incomplete data, to which it applied its learned imputation strategies. Simultaneously, D was trained to discern between these imputed data points and the actual data present in the dataset. This training process was underpinned by a delicate balance. It was crucial to ensure that neither G nor D became overwhelmingly dominant over the other. If G became too powerful, the imputations might go unchecked and diverge from realistic data patterns. Conversely, if D became too dominant, it could stifle G's learning process, preventing it from developing effective imputation strategies. Therefore, careful tuning of the training parameters and regular monitoring of the training dynamics were essential components of our methodology.

The performance evaluation of the GAN model was multifaceted. The primary metrics used were Mean Squared Error (MSE) for continuous variables and F1-score for categorical variables. These metrics provided insights into the accuracy and reliability of the imputations made by G. MSE was instrumental in quantifying the deviation of imputed numerical values from their actual values, while the F1-score offered a measure of the GAN's precision and recall in imputing categorical data. Beyond these primary metrics, additional analyses were conducted to assess the statistical coherence of the imputed data. This involved evaluating the distribution, variance, and inter-variable correlations of the imputed data against the original dataset. Such analyses were crucial in ensuring that the imputed data not only replaced missing values but also preserved the integral statistical properties of the dataset, which are fundamental for subsequent analytical and predictive modelling tasks in the field of solar energy.

3.2.3. Applying GANs to solar energy datasets

Utilising GANs for the imputation of missing data in solar datasets offers a novel and promising approach to overcome the limitations inherent in simpler imputation strategies. The adeptness of GANs in generating data that retains the underlying statistical and contextual relationships is particularly valuable. This is of utmost importance in solar energy data, where the interdependencies between environmental conditions and operational parameters can be highly nuanced and vary significantly with factors such as location, season, and weather patterns. Furthermore, GAN architectures can be customised to handle the unique challenges of solar datasets, including the treatment of mixed variable types, ensuring a comprehensive and contextually relevant data imputation.

4. Proposed technique

In the rapidly evolving landscape of deep learning, the ability to efficiently process and analyse complex data sequences is paramount. This capability is particularly crucial in fields like natural language processing, time-series analysis, and advanced signal processing. Traditional neural network architectures, while powerful, often struggle with the intricacies of spatial–temporal data, which is characterised by its multi-dimensional and dynamic nature. The challenge lies in designing a model that can not only understand the depth and breadth of such data but also focus on the most relevant aspects to extract meaningful insights. Recent advancements in deep learning have led to the development of various specialised neural network architectures, each excelling in different aspects of data processing. However, a gap remains in creating a unified model that synergistically combines these strengths to handle complex spatial–temporal data more effectively. The need for such an integrated approach becomes increasingly evident as the complexity and volume of data in fields like language modelling, environmental sensing, and financial forecasting grow exponentially.

In response to this need, we propose "Inception Embedded Dual Attention Bi LSTM GRU model" – a novel neural network architecture that integrates several advanced techniques to create a more robust and versatile solution for processing complex data sequences. This model is designed to leverage the complementary strengths of its constituent components to provide a comprehensive and nuanced understanding of spatial–temporal data.

- **Inception Module:** At the heart of our model lies the Inception module, renowned for its efficiency in handling large-scale image recognition tasks. Its ability to capture features at various scales makes it an ideal candidate for extracting intricate details from complex data sequences.
- **Embedding Layer:** Given the significance of context in understanding spatial–temporal sequences, our model incorporates an embedding layer. This layer is crucial for transforming categorical or textual data into a continuous vector space, laying the groundwork for more nuanced processing.
- **Dual Attention Mechanism:** To enhance the model's focus, we integrate a dual attention mechanism. This mechanism allows the model to concurrently attend to both spatial and temporal aspects of the data, ensuring a more targeted and relevant analysis.
- **Bi-LSTM Layer:** The bidirectional Long Short-Term Memory (Bi-LSTM) layer is implemented to process the attention-weighted embeddings. By traversing the data in both forward and backward directions, Bi-LSTM ensures a comprehensive understanding of the sequence dynamics.

4.1. Inception model

In the realm of deep learning and computer vision, the Inception model stands as a monumental achievement, introduced by Szegedy et al. in the seminal paper "Going Deeper with Convolutions". This model represents a paradigm shift in the design of convolutional neural networks (CNNs) and has significantly influenced the field of large-scale image recognition. Renowned for its innovative architecture, the Inception model optimises the use of computational resources, achieving remarkable efficiency and effectiveness.

The development of the Inception model is rooted in the quest to improve the performance of neural networks in image recognition tasks without a proportional increase in computational complexity. Traditional CNN architectures prior to Inception primarily focused on deepening or widening the network to enhance performance. This approach, while effective to a certain extent, often leads to inflated computational costs and an increased risk of overfitting. The Inception model emerged as a solution to these challenges, offering an architecture that balances depth and width without undue computational burden. The core innovation of the Inception model lies in its unique architecture, which deviates from the traditional sequential layer stacking. The model introduces a novel concept where convolutional layers with different filter sizes, along with max-pooling layers, are applied in parallel and then concatenated. This design allows the model to capture and integrate information from various scales and perspectives.

- **Multi-Scale Feature Extraction:** By employing convolutional filters of various sizes (1×1, 3×3, 5×5) within the same layer, the Inception model can extract a wide range of features, from fine details to broader contextual information. This multi-scale extraction is key to the model's ability to recognise patterns and features of different sizes and complexities within images.
- **Integration of Max-Pooling:** The inclusion of max-pooling layers serves multiple purposes. Primarily, it reduces the spatial dimension of the feature maps, which helps in lowering the computational cost. Moreover, max-pooling provides a degree of spatial invariance, enhancing the model's ability to recognise features irrespective of their location in the input image.
- **Concatenation for Rich Feature Representation:** The outputs of the different filters and the max-pooling layer are concatenated along the channel dimension, resulting in a comprehensive feature map. This concatenated output is a rich blend of features captured at various scales, providing a holistic understanding of the input data.

Delving into the mathematical underpinnings, the Inception model's operation can be encapsulated as follows. Given an input X, the output Y of an Inception layer is formulated as:

$$Y = \text{Concat}\left(\text{Conv}_{1\times1}(X), \text{Conv}_{3\times3}(X), \text{Conv}_{5\times5}(X), \text{MaxPool}(X)\right) \tag{2}$$

where, $\text{Conv}_{1\times1}(X), \text{Conv}_{3\times3}(X), \text{Conv}_{5\times5}(X)$ denote the convolution operations with respective filter sizes, targeting different levels of feature granularity. $\text{MaxPool}(X)$ refers to the max-pooling operation, contributing to the model's robustness against positional variations of features. - The function Concat amalgamates these diverse feature maps into a unified representation, enriching the model's feature detection capability.

4.2. Dual attention mechanism

The Dual Attention Mechanism, an advanced concept in neural network architecture, has emerged as a powerful tool for enhancing the processing capabilities of models, especially in tasks involving complex data dependencies. This mechanism is distinguished by its simultaneous application of two distinct types of attention — spatial and temporal — each targeting a different aspect of data. In many real-world scenarios, particularly in fields like time-series analysis, video processing, and natural language processing, data exhibit inherent spatial and temporal characteristics. The spatial aspect refers to the inter-relationships or arrangements of data elements at any given time, while the temporal aspect pertains to how these elements evolve over time. Traditional neural network architectures often struggle to capture both these dimensions effectively. The Dual Attention Mechanism addresses this limitation by applying two parallel attention modules, enabling the network to focus on critical spatial features and temporal dynamics concurrently. The mathematical modelling of the Dual Attention Mechanism is rooted in the concept of attention in neural networks. Attention mechanisms, in general, are designed to allow neural networks to focus on specific parts of the input, akin to how human attention works. In the context of Dual Attention, this focus is bifurcated into spatial and temporal components.

Spatial attention in the context of this mechanism refers to the model's ability to selectively concentrate on different regions of the input data at a given time point. This is crucial when spatial arrangement carries significant information, such as in image recognition or language processing tasks. The spatial attention weights, denoted as α_s, are computed using the output H of a previous layer, such as an LSTM. The computation involves a weight matrix W_s and a bias term b_s, and is given by the softmax function:

$$\alpha_s = \text{softmax}(W_s H + b_s) \tag{3}$$

Here, W_s is a learnable weight matrix that transforms the input into a space where attention weights can be computed, and b_s is a bias term. The softmax function ensures that the computed attention weights are normalised, i.e., they sum up to 1, allowing them to be interpreted as probabilities or relative importances. Temporal attention, on the other hand, focuses on the sequential aspect of the data, enabling the model to pay more attention to certain time steps than others. This is crucial in tasks like time-series forecasting or sequence modelling, where the significance of information can vary greatly over time.

The temporal attention weights, denoted as α_t, are computed similarly to spatial attention weights, using the same output H but with a different weight matrix W_t and bias term b_t:

$$\alpha_t = \text{softmax}(W_t H + b_t) \tag{4}$$

In this case, W_t is tailored to capture the temporal dynamics of the input, and b_t adjusts the focus along the temporal dimension. The final step in the Dual Attention Mechanism is the combination of these two attention weights to produce the attention-applied output A. This output is a representation of the input that has been adjusted to emphasise certain spatial and temporal features

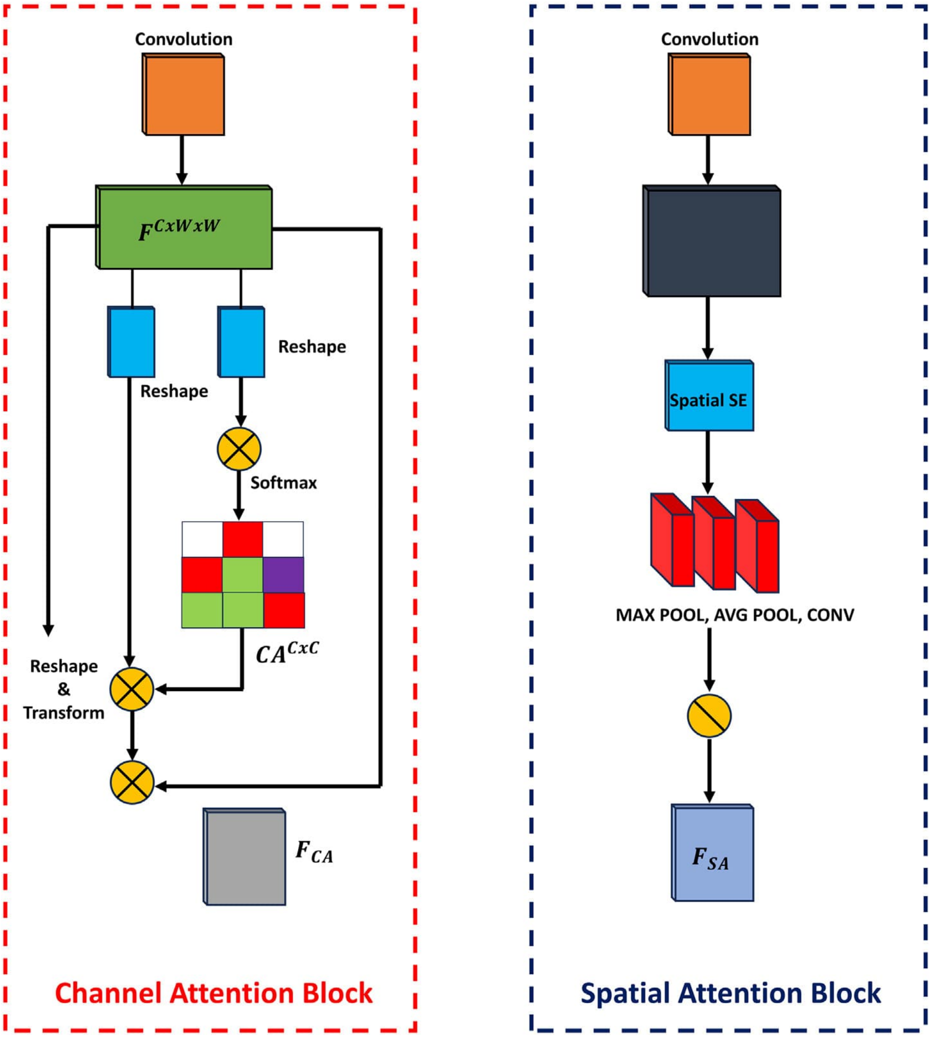


Fig. 4. Architecture of Dual Attention Mechanism.

according to the computed attention weights. Mathematically, this can be expressed as a weighted sum of the input, using the attention weights:

$$A = \alpha_s \odot H + \alpha_t \odot H \tag{5}$$

Here, $\odot$ represents element-wise multiplication. The attention-applied output A effectively combines the insights from both spatial and temporal dimensions, providing a more comprehensive representation of the input data (see Fig. 4).

4.3. Bidirectional Long Short-Term Memory (Bi-LSTM)

Bidirectional Long Short-Term Memory (Bi-LSTM) networks represent an evolutionary step in the development of neural network architectures, particularly for tasks involving sequential data. As an extension of the traditional Long Short-Term Memory (LSTM) network, Bi-LSTMs are designed to enhance performance by capturing information in both forward and backward directions through a sequence. This two-way flow of information allows the model to have context from both the past and the future states, offering a more comprehensive understanding of the sequence.

In traditional LSTM networks, the information flow is unidirectional, meaning that the network can only preserve information from the past to the present. This approach is often sufficient for many sequence processing tasks. However, there are numerous scenarios, particularly in language processing and time-series analysis, where the context or dependencies span both backward and forward in time. For instance, in natural language, the meaning of a word can depend on the following words, not just the preceding ones. Bi-LSTMs address this limitation by processing the data in both directions, thus capturing a broader context. To appreciate the workings of a Bi-LSTM, it is essential to understand the basic mechanism of an LSTM unit. LSTM networks, a type of recurrent neural network (RNN), are specifically designed to avoid the long-term dependency problem, enabling them to remember information for extended periods. An LSTM unit consists of a cell state and three gates:

- **Forget Gate:** Decides what information to discard from the cell state.
- **Input Gate:** Updates the cell state with new information.
- **Output Gate:** Determines the next hidden state based on the current cell state and the input at that step.

These gates in an LSTM unit allow it to selectively remember or forget patterns, making them highly efficient for sequence modelling tasks. A Bi-LSTM leverages two LSTM layers that process the input sequence in opposite directions — one forward and one backward. Given an input sequence $X = (x_1, x_2, \dots, x_T)$, where x_t represents the input at time step t, the Bi-LSTM computes the forward and backward hidden states at each time step. For the forward LSTM, denoted as LSTM_{fwd}, the hidden state h_t^{fwd} at time t is computed as:

$$h_t^{fwd} = \mathrm{LSTM}_{fwd}(x_t, h_{t-1}^{fwd}) \tag{6}$$

Conversely, for the backward LSTM, denoted as LSTM_{bwd}, which processes the sequence in reverse, the hidden state h_t^{bwd} is calculated as:

$$h_t^{bwd} = \mathrm{LSTM}_{bwd}(x_t, h_{t+1}^{bwd}) \tag{7}$$

The final output at each time step t in the Bi-LSTM network is obtained by concatenating the forward and backward hidden states:

$$h_t = [h_t^{fwd}; h_t^{bwd}] \tag{8}$$

This concatenation ensures that the output at each time step encapsulates information from both past and future contexts relative to that time step. Each LSTM unit within the Bi-LSTM structure consists of the aforementioned gates — forget, input, and output — each performing specific functions:

- **Forget Gate** (f_t): Decides what information is discarded from the cell state. It looks at h_{t-1} and x_t, and outputs a number between 0 and 1 for each number in the cell state C_{t-1}. A 1 represents "completely retain this" while a 0 represents "completely forget this.
- **Input Gate** (i_t) : Updates the cell state with new information. It involves two parts: a sigmoid layer that decides which values to update, and a tanh layer that creates a vector of new candidate values that could be added to the state.
- **Output Gate** (o_t): Determines the next hidden state. The hidden state contains information on previous inputs, filtered by the output gate. The output gate looks at the current input x_t and the previous hidden state h_{t-1}.

The operations within each gate can be mathematically represented as:
- Forget gate: $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
- Input gate: $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$ and $\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$
- Output gate: $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$
Bi-LSTMs have become a go-to architecture in many NLP tasks, such as sentiment analysis, machine translation, and speech recognition. Their ability to capture context from both directions often results in better performance compared to unidirectional LSTMs. In time-series analysis, Bi-LSTMs are effective in scenarios where the future context is just as important as the past for making predictions or inferences.

4.4. Weighted sum loss for multi-task models

In the development of multi-task learning models, particularly for predicting both electrical power and thermal energy outputs, it is essential to effectively balance the contributions of each task during the training phase. This balance is achieved using a weighted sum loss function, which combines the individual loss components from each task into a single scalar value used for optimising the model. Mathematically, the weighted sum loss function can be expressed as:

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{power} + \beta\mathcal{L}_{thermal} \tag{9}$$

where, $\mathcal{L}_{total}$ represents the total loss function, $\mathcal{L}_{power}$ is the loss associated with power prediction, and $\mathcal{L}_{thermal}$ is the loss associated with thermal energy prediction. The coefficients α and β are weights assigned to each task's loss, controlling their relative importance in the training process. The inclusion of a weighted sum loss function is significant for several reasons. First, it allows the model to learn shared representations that benefit both tasks while ensuring that neither task disproportionately influences the training process. This balancing act is crucial because the power and thermal energy outputs, although related, have distinct characteristics and error distributions that the model must account for simultaneously.

The choice of weights α and β is critical. In practical applications, these weights are determined based on the specific requirements and priorities of the system. For instance, in scenarios where the accuracy of power prediction is paramount, a higher weight α is assigned to $\mathcal{L}_{power}$. Conversely, if thermal energy prediction holds greater importance, a higher weight β is assigned to $\mathcal{L}_{thermal}$.

This flexibility allows the model to be fitted to various applications, enhancing its utility and effectiveness. From a mathematical modelling perspective, the weighted sum loss function provides a framework that integrates multiple objectives into a single optimisation problem. This integration is essential for the convergence of multi-task models, ensuring that the learning algorithm effectively navigates the trade-offs between different tasks. The optimisation problem can be formalised as:

$$\theta^* = \arg\min_{\theta} \mathcal{L}_{total}(\theta) = \arg\min_{\theta}(\alpha \mathcal{L}_{power}(\theta) + \beta \mathcal{L}_{thermal}(\theta)) \tag{10}$$

where, θ represents the model parameters. The goal is to find the optimal set of parameters θ^* that minimise the total loss function $\mathcal{L}_{total}$.

In the context of our proposed model, the Inception Embedded Dual Attention Bi-LSTM Model, the weighted sum loss function plays a pivotal role. By employing this loss function, our model can effectively handle the dual tasks of predicting electrical power and thermal energy outputs, capturing the complex spatial–temporal dependencies inherent in the data. The Inception module's multi-scale feature extraction, combined with the dual attention mechanisms and Bi-LSTM's sequential data processing capabilities, benefits from the balanced learning facilitated by the weighted sum loss function. The relevance of this approach to our work cannot be overstated. The hybrid PV–Thermal system's prediction accuracy centres on the model's ability to concurrently learn and optimise for both power and thermal energy outputs. The weighted sum loss function ensures that the model maintains a absolute perspective, optimising for overall system performance rather than excelling in one aspect at the expense of the other. This balance is crucial for practical applications where both power and thermal energy outputs are critical for system efficiency and reliability. The weighted sum loss function is integral to our multi-task learning model, providing a balanced and flexible approach to optimising predictions for both electrical power and thermal energy outputs. Its implementation ensures that our model captures the nuanced dependencies between these tasks, leading to more accurate and reliable predictions, ultimately enhancing the performance and utility of hybrid PV–Thermal systems.

5. Inception Embedded Dual Attention Bi LSTM Model

The "Inception Embedded Dual Attention Bi LSTM Model" weaves together a complex array of mathematical models that correspond to its integrated components, each contributing uniquely to the overall functionality of the architecture. This integration not only leverages the individual strengths of each component but also ensures that they work synergistically to enhance the model's overall performance. At the outset, the Inception module processes the input data through a series of convolutional layers with varying filter sizes. This process can be mathematically articulated as a concatenation of multiple convolutional transformations. For an input X, the output $Y_{\text{inception}}$ of the Inception module is obtained by applying convolution operations with filters of different sizes (1×1, 3×3, 5×5) and a max-pooling operation, and then concatenating these outputs. The mathematical expression for this operation is as follows:

$$Y_{\text{inception}} = \text{Concat}(\text{Conv}_{1\times1}(X), \text{Conv}_{3\times3}(X), \text{Conv}_{5\times5}(X), \text{MaxPool}(X)). \tag{11}$$

This output provides a rich, multi-scale feature representation of the input data, essential for capturing intricate spatial details. Following the Inception module, the embedding layer transforms each element of the input sequence, particularly important for textual or categorical data, into a dense vector representation. This transformation is mathematically represented by mapping each input element to a vector in a high-dimensional space using an embedding matrix E. For an input element x_i, its embedding vector e_i is obtained as:

$$e_i = E[x_i] \tag{12}$$

where, x_i represents the index of the element in the embedding matrix. This step is crucial for converting discrete data points into a continuous format that can be efficiently processed by neural networks.

The dual attention mechanism, comprising both spatial and temporal components, follows the embedding layer. This mechanism mathematically computes attention weights that allow the model to focus selectively on different parts of the data sequence. The spatial attention weights α_s and temporal attention weights α_t are calculated using the softmax function applied to linear transformations of the hidden states from the previous layer. For a hidden state H, the spatial attention weights are computed as:

$$\alpha_s = \text{softmax}(W_s H + b_s) \tag{13}$$

and the temporal attention weights as:

$$\alpha_t = \text{softmax}(W_t H + b_t) \tag{14}$$

where W_s, W_t are weight matrices and b_s, b_t are bias terms for the spatial and temporal attention mechanisms, respectively. These weights are then used to generate an attention-modified output that emphasises relevant spatial and temporal features in the data.

The architecture then incorporates a Bi-LSTM layer, which processes the attention-weighted embeddings in both forward and backward directions across the sequence. The Bi-LSTM layer is mathematically characterised by two sets of LSTM operations, one for each direction. For the forward direction, the hidden state h_t^{fwd} at time t is computed as:

$$h_t^{fwd} = \text{LSTM}^{fwd}(x_t, h_{t-1}^{fwd}) \tag{15}$$

Table 3
Hyperparameters of the proposed model.

Hyperparameter	Value
Input shape	(10, 14)
Filters in Inception Module	64
Kernel sizes in Inception Module	1, 3, 5
MaxPooling stride	1
LSTM units	128
Dense units after Flatten	128
Output units	2
Activation for Conv1D	relu
Activation for Dense	relu
Optimiser	adam
Loss function	mse
Metrics	RMSE, R-squared
Epochs	100
Batch size	32
Train-val-Test Split	70%, 10%, 20%

and similarly for the backward direction:

$$h_t^{bwd} = \text{LSTM}^{bwd}(x_t, h_{t+1}^{bwd}) \tag{16}$$

The final output at each time step is a concatenation of the forward and backward hidden states:

$$h_t = [h_t^{fwd}; h_t^{bwd}] \tag{17}$$

providing a comprehensive view of the sequential data by incorporating both past and future context.

The final stage of the model is the GRU layer, which further refines the sequential processing of data. The GRU modifies its hidden state h_t at each time step through a series of gated operations involving update and reset gates. The update gate z_t is computed as:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{18}$$

and the reset gate:

$$r_t \text{ as } r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{19}$$

where, σ denotes the sigmoid function, and W_z, W_r, U_z, U_r, b_z, and b_r are parameter matrices and biases. The candidate hidden state $\hat{h}_t$ is then calculated using a tanh activation function applied to a linear transformation of the input and the element-wise product of the reset gate and the previous hidden state:

$$\hat{h}_t = \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \tag{20}$$

The final hidden state h_t is a combination of the previous hidden state and the candidate hidden state weighted by the update gate:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t \tag{21}$$

The "Inception Embedded Dual Attention Bi LSTM GRU Model" combines these intricate mathematical operations to form a unified architecture capable of advanced data processing and analysis. The model adeptly manages to capture and emphasise the most relevant features in both spatial and temporal dimensions of the data, efficiently process sequential information, and adaptively retain important aspects of the data over time. This integrated approach makes the model exceptionally powerful for tasks that require deep understanding and nuanced interpretation of complex sequential data. As such, it stands as a cutting-edge example of how the amalgamation of diverse neural network technologies can lead to groundbreaking advancements in the field of artificial intelligence and machine learning (see Fig. 5 and Table 3).

6. Results and discussion

This section delves into the comparative analysis of sophisticated multitask learning models tailored for forecasting thermal energy and electrical power outputs. Through meticulous evaluation, we assess the models' performance on key metrics, including RMSE, R2, NMSE, MAPE, and MAE, shedding light on their predictive accuracy and reliability. The ensuing discussion offers insights into the implications of these results, emphasising the pivotal role of model selection in enhancing prediction precision within the complex domain of energy systems.

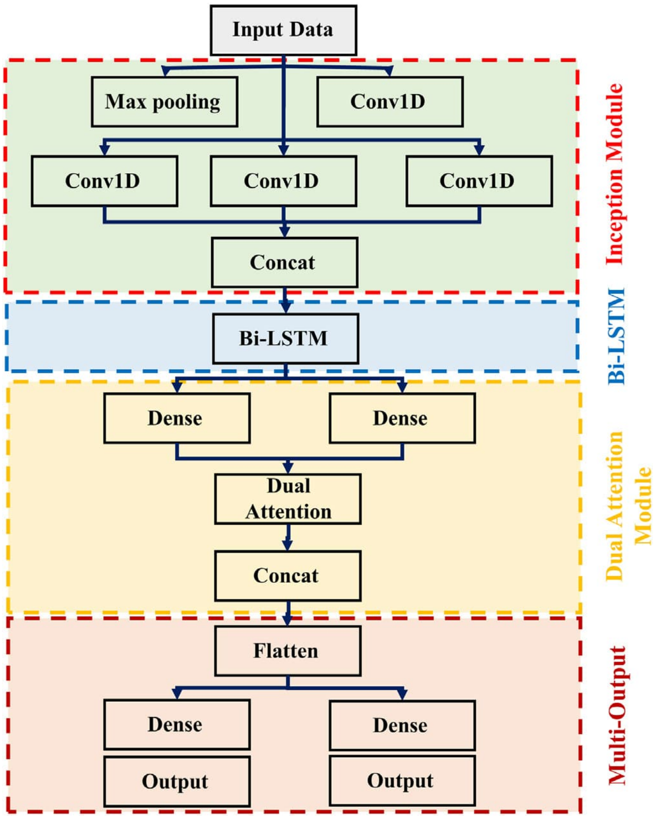


Fig. 5. Detailed Architecture of Proposed Model.

6.1. *Comparative analysis for summer season*

6.1.1. *Thermal energy prediction*

The comparative analysis of thermal energy prediction scrutinised three distinct deep learning models within a multitask learning framework: IDAMBi-LSTM, IAM-GRU, and IDAM-GRU. Various performance metrics were employed to assess these models, including Root Mean Square Error for Thermal energy (RMSE-T), coefficient of determination (R^2-T), Normalised Mean Square Error for Thermal energy (NMSE-T), Mean Absolute Percentage Error for Thermal energy (MAPE-T), and Mean Absolute Error for Thermal energy (MAE-T). The timestamps used for testing in the summer season are from 13-Jul-2021 23:06:45 to 15-Jul-2021 23:10:05.

The IDAMBi-LSTM model showcased outstanding performance across the metrics, achieving an RMSE-T value of 13.619, the lowest among the models, which suggests it had the smallest average prediction error for thermal energy estimation. This model also achieved a remarkably high R^2-T of 99.79%, indicating that it could explain nearly all the variance in the thermal energy data. The NMSE-T value of 0.00393 and MAPE-T of 34.65% further emphasised the model's high accuracy and reliability in thermal energy estimation with minimal deviation from actual values. The MAE-T of 3.725 highlighted its precision in predicting thermal energy, reflecting the smallest average magnitude of errors across predictions.

In contrast, the IAM-GRU model displayed less impressive but still commendable results, with an RMSE-T of 21.200, an R^2-T of 97.50%, an NMSE-T of 0.00910, a MAPE-T of 6.00%, and an MAE-T of 5.900. These results indicate that the IAM-GRU model, although effective, was less accurate and reliable compared to the IDAMBi-LSTM model.

The IDAM-GRU model had intermediate performance metrics, with an RMSE-T of 27.750, an R^2-T of 95.60%, an NMSE-T of 0.01400, a MAPE-T of 9.50%, and an MAE-T of 8.800. These outcomes suggest that the IDAM-GRU model can generate reliable and accurate thermal energy predictions but does not quite reach the performance level of the IDAMBi-LSTM model.

This analysis underscores the superior effectiveness of the IDAMBi-LSTM model in predicting thermal energy within a multitask learning setup, highlighting its superiority in terms of accuracy and reliability when compared to the IAM-GRU and IDAM-GRU models. The results demonstrate the potential of employing advanced deep learning architectures for precise and dependable thermal energy prediction, crucial for optimising energy systems and enhancing efficiency (see Figs. 6 and 7).

Table 4
Comparative analysis thermal energy for summer season.

Model	RMSE-T	R2-T	NMSE-T	MAPE-T	MAE-T
IDAMBi-LSTM	13.619	99.79	0.00393	3.65%	3.725
IAM-GRU	21.200	97.50	0.00910	6.00%	5.900
IDAM-GRU	27.750	95.60	0.01400	9.50%	8.800

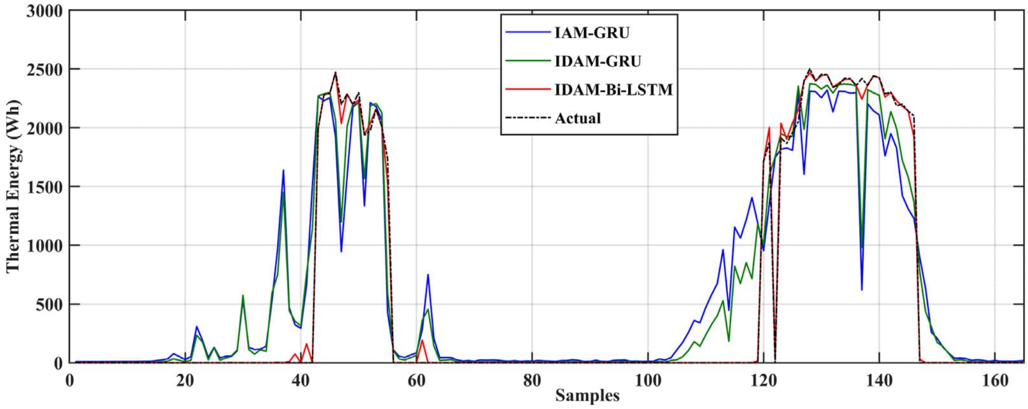


Fig. 6. Prediction of Thermal Energy by Proposed Technique for Summer Season.

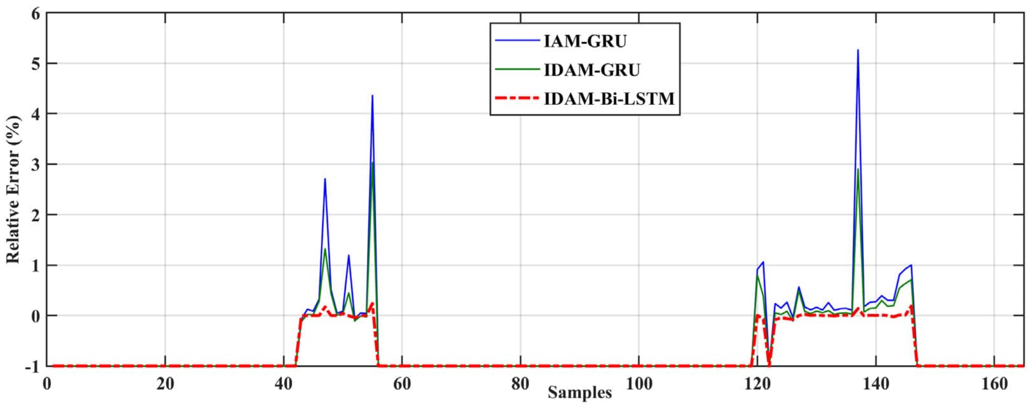


Fig. 7. Relative Error of Thermal Energy Prediction by Proposed Technique for Summer Season.

6.1.2. Electrical power prediction

The analysis of deep learning models for predicting electrical power output showcases the effectiveness of these models with varying degrees of precision and accuracy. Central to this evaluation is the IDAMBi-LSTM model, which stands out for its exceptional performance. The timestamps used for testing for the summer season are from 13-Jul-2021 23:06:45 to 15-Jul-2021 23:10:05.

The IDAMBi-LSTM model achieves an RMSE-E of just 1.85, reflecting its minimal deviation from actual values, thus indicating high accuracy. This is complemented by an R2-E score of 99.99%, suggesting an almost perfect alignment with the observed data trends. Additionally, the NMSE-E value for this model is 0.00038, which points to a low variance in error across predictions, enhancing its reliability. In terms of percentage errors, the IDAMBi-LSTM model records a MAPE-E of 4.05%, which is low, highlighting its precision in forecasts. Moreover, the MAE-E value at 7.377 further solidifies its ability to consistently produce close approximations of the actual outputs, thereby minimising average absolute errors.

In contrast, the IAM-GRU and IDAM-GRU models show variations in performance that rank them below the IDAMBi-LSTM. The IAM-GRU model, with its higher RMSE-E of 3.90, indicates less accuracy and a greater deviation from actual measurements. Its R2-E score of 97.70% still shows a high degree of model fit but is less than ideal when compared to IDAMBi-LSTM. This model also has a higher NMSE-E of 0.00110, suggesting a slightly increased error variability. With a MAPE-E of 7.50% and a MAE-E of 11.500, the IAM-GRU model's predictions are less precise and exhibit greater average errors relative to the other models.

Meanwhile, the IDAM-GRU model, though better than IAM-GRU, does not reach the high benchmarks set by IDAMBi-LSTM. It records an RMSE-E of 7.87, which is significantly higher and indicates a considerable deviation from the expected values. The R2-E

Table 5
Comparative analysis electrical power for summer season.

Model	RMSE-E	R2-E	NMSE-E	MAPE-E	MAE-E
IDAMBi-LSTM	1.85	99.99	0.00038	4.05%	7.377
IAM-GRU	3.90	97.70	0.00110	7.50%	11.500
IDAM-GRU	7.87	95.80	0.00239	9.20%	17.450

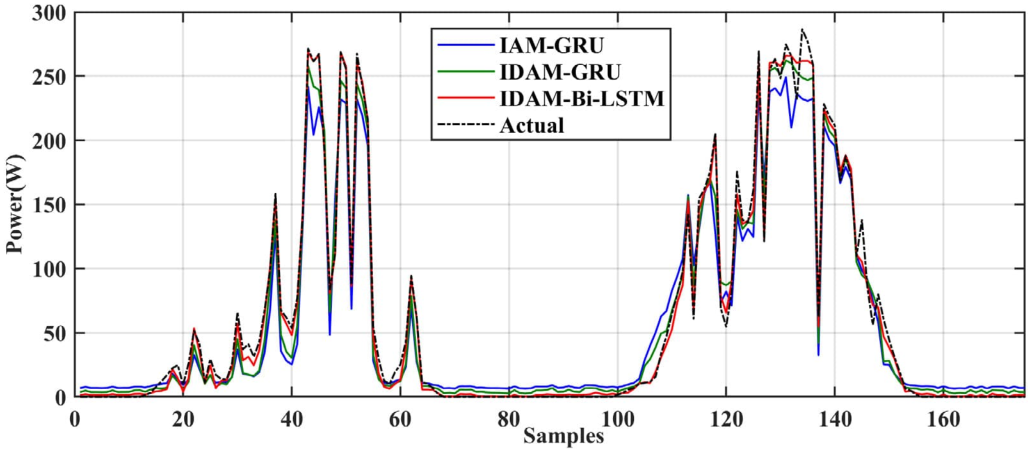


Fig. 8. Prediction of PV Power by Proposed Technique for Summer Season.

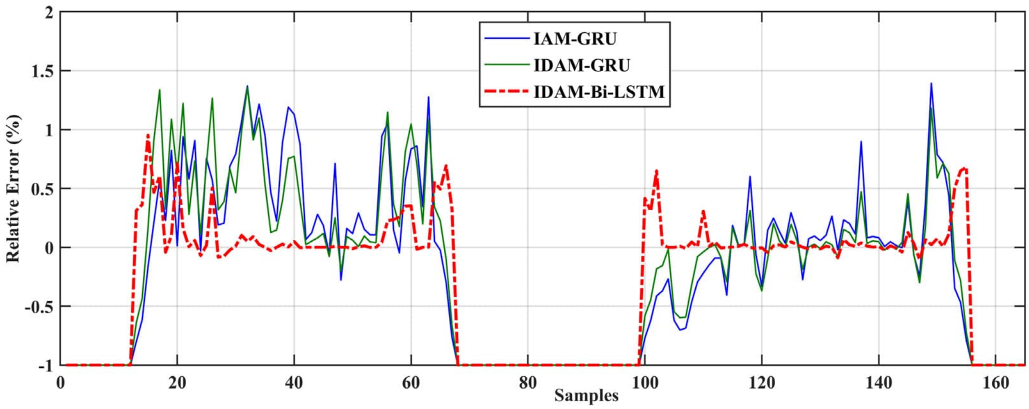


Fig. 9. Relative Error of PV Power Prediction by Proposed Technique for Summer Season.

score of 95.80% reveals that while the model generally captures the underlying data pattern, it is not as tightly fitted as IDAMBi-LSTM. Its NMSE-E stands at 0.00239, which, although acceptable, indicates more variability in error than the leading model. The MAPE-E of 9.20% and a MAE-E of 17.450 further demonstrate that while it is an improvement over IAM-GRU, the IDAM-GRU model still falls short in terms of both precision and the ability to minimise errors (see Figs. 8 and 9).

6.2. Comparative analysis of spring season

6.2.1. Thermal energy prediction

In the comparative study of thermal energy prediction models for the spring season, the analysis table presents an insightful evaluation of three predictive models: IDAMBi-LSTM, IAM-GRU, and IDAM-GRU, across a range of statistical metrics aimed at assessing their performance in forecasting temperatures. The results underscore the superior performance of the IDAMBi-LSTM model, which exhibits the lowest Root Mean Square Error (RMSE-T) at 13.712. This metric, indicative of the average magnitude of the prediction errors, highlights the model's precision in estimating temperatures, with lower RMSE-T values suggesting a closer fit to the observed data. The high accuracy of IDAMBi-LSTM is further confirmed by its near-perfect R-squared value (R2-T) of 99.58%, demonstrating that it can account for almost all the variability in the temperature data, thereby providing a reliable model for

Table 6
Comparative analysis of thermal energy prediction for spring season.

Model	RMSE-T	R2-T	NMSE-T	MAPE-T	MAE-T
IDAMBi-LSTM	13.712	99.58	0.00391	3.72%	3.789
IAM-GRU	21.337	97.12	0.00973	6.49%	5.367
IDAM-GRU	27.899	95.34	0.01512	9.18%	8.575

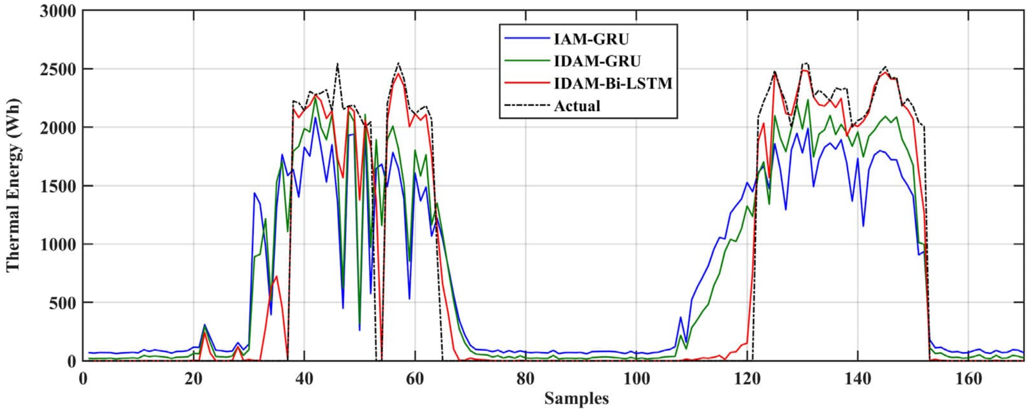


Fig. 10. Prediction of Thermal Energy by Proposed Technique for Spring Season.

predicting thermal energy. The timestamps used for testing for the summer season are from 9-May-2021 23:09:35 to 11-may-2021 23:08:55.

In terms of normalised errors, the IDAMBi-LSTM model continues to demonstrate its efficacy with the lowest Normalised Mean Square Error (NMSE-T) of 0.00391. This metric provides a relative measure of error magnitude adjusted for the variance in the observed temperatures, confirming that the model not only minimises errors but also efficiently handles variations in temperature data, making it highly suitable for dynamic weather conditions typical of the spring season.

Analysing the model's performance from the perspective of average percentage errors, the IDAMBi-LSTM model again stands out with the lowest Mean Absolute Percentage Error (MAPE-T) at 3.72%. This percentage indicates how far the model's predictions deviate, on average, from the actual measurements as a percentage, reinforcing the model's reliability in practical scenarios. Similarly, the Mean Absolute Error (MAE-T) of 3.789, the smallest among the three models, confirms that the absolute discrepancies between predicted and actual temperatures are minimal, thus establishing the IDAMBi-LSTM's consistency in delivering precise temperature forecasts.

In contrast, the IAM-GRU and IDAM-GRU models show progressively higher values across all these metrics, reflecting lesser accuracy and reliability in temperature prediction. The IAM-GRU, with an RMSE-T of 21.337 and R2-T of 97.12%, performs moderately, whereas the IDAM-GRU, with the highest RMSE-T of 27.899 and an R2-T of 95.34%, is the least accurate of the three. These models, although reasonably effective, lag behind in capturing the full scope of temperature variability and in minimising prediction errors compared to the IDAMBi-LSTM model.

The comparative analysis distinctly highlights the advanced capability of the IDAMBi-LSTM model, attributing its success to the bidirectional Long Short-Term Memory (LSTM) architecture, which is adept at capturing both past and future temporal dependencies in data sequences. This characteristic is particularly advantageous in the context of thermal energy forecasting, where understanding complex patterns and fluctuations in temperature is crucial. The findings from this analysis not only validate the effectiveness of sophisticated machine learning architectures like the IDAMBi-LSTM in handling complex predictive tasks but also underscore the importance of model selection based on specific application needs to ensure optimal performance in real-world scenarios (see Figs. 10 and 11 and Table 6).

6.2.2. Electrical power prediction

The table titled "Comparative Analysis of Electrical Power for Spring Season" evaluates the performance of three predictive models — IDAMBi-LSTM, IAM-GRU, and IDAM-GRU — across several key metrics designed to assess their efficacy in forecasting electrical power. The metrics include Root Mean Square Error-Electrical (RMSE-E), R-squared - Electrical (R2-E), Normalised Mean Square Error-Electrical (NMSE-E), Mean Absolute Percentage Error-Electrical (MAPE-E), and Mean Absolute Error-Electrical (MAE-E). These metrics together provide a comprehensive understanding of each model's predictive accuracy, reliability, and error characteristics.

The IDAMBi-LSTM model emerges as the most accurate, with the lowest RMSE-E of 1.74, indicating it has the smallest average errors in its predictions, a critical factor for precision in power forecasting. This model also achieves an exceptionally high R2-E score of 99.90%, suggesting that it explains virtually all the variability in electrical power data around its mean, an indication of

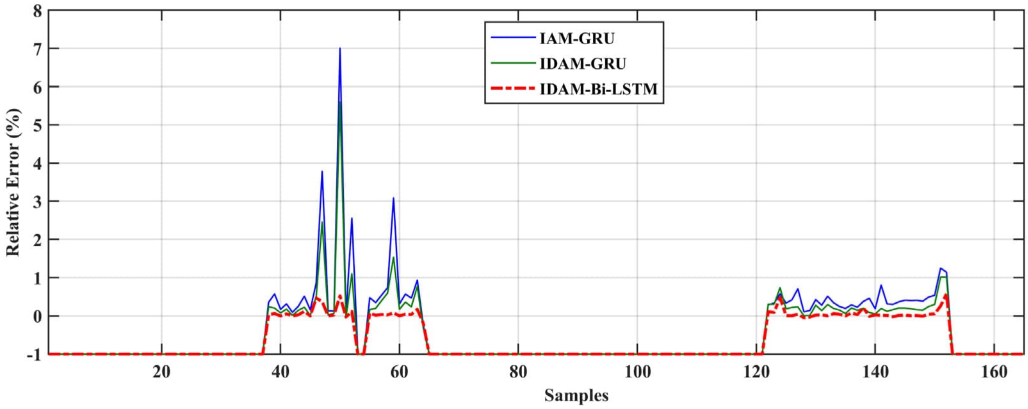


Fig. 11. Relative Error of Thermal Energy Prediction by Proposed Technique for Spring Season.

Table 7
Comparative analysis of electrical power for spring season.

Model	RMSE-E	R2-E	NMSE-E	MAPE-E	MAE-E
IDAMBi-LSTM	1.74	99.90	0.00037	4.05%	6.995
IAM-GRU	3.78	97.36	0.00101	7.21%	10.120
IDAM-GRU	6.99	95.22	0.00221	9.04%	16.390

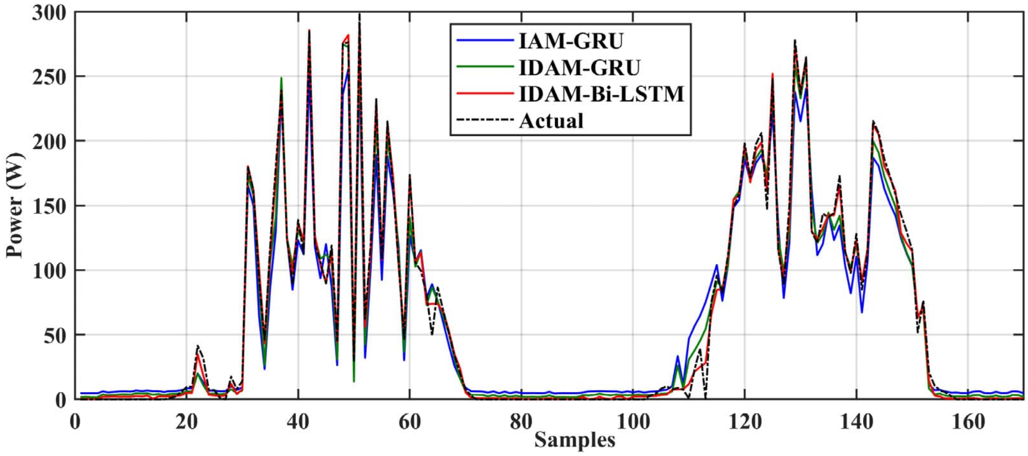


Fig. 12. Prediction of PV Power by Proposed Technique for Spring Season.

excellent model fit. Additionally, the IDAMBi-LSTM's NMSE-E is the lowest at 0.00037, which shows that its prediction errors are minimal when normalised against the variance of the dataset, reinforcing its effectiveness across varying power levels.

In terms of percentage-based errors, the IDAMBi-LSTM again leads with a MAPE-E of 4.05%. This metric highlights the model's capability to maintain close proximity to the actual data points in terms of percentage deviations, which is particularly valuable in applications where proportional errors are more impactful. Furthermore, its MAE-E of 6.995 points to the smallest absolute differences between predicted and actual values, ensuring consistent reliability in its power output predictions.

Conversely, the IAM-GRU and IDAM-GRU models demonstrate higher errors and less accuracy in their predictions. The IAM-GRU records a RMSE-E of 3.78, an R2-E of 97.36%, and an NMSE-E of 0.00101, depicting it as moderately accurate but less precise than the IDAMBi-LSTM. Its MAPE-E and MAE-E are also higher at 7.21% and 10.120, respectively, which could affect its usability in more error-sensitive applications. The IDAM-GRU shows the least favourable results with the highest RMSE-E of 6.99, a reduced R2-E of 95.22%, and the highest NMSE-E of 0.00221, indicating significant deviations in its predictions. The model's MAPE-E of 9.04% and MAE-E of 16.390 further reflect its reduced effectiveness in accurately forecasting electrical power Figs. 12 and 13) (see Table 7).

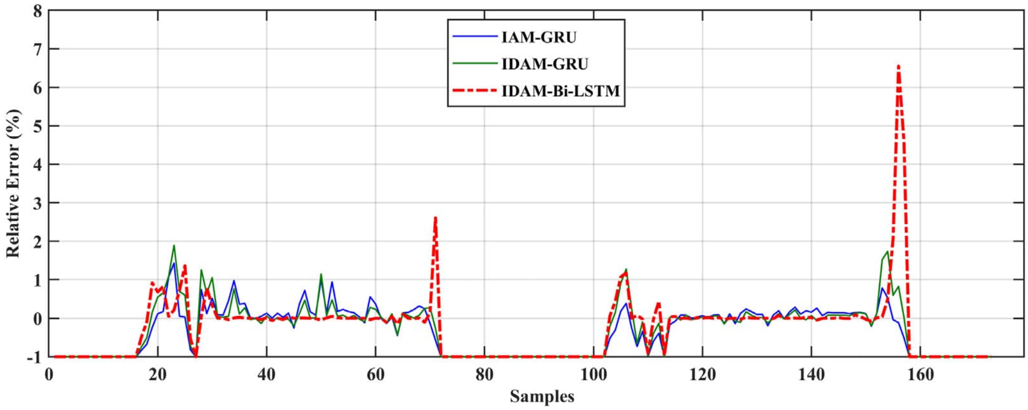


Fig. 13. Relative Error of PV Power Prediction by Proposed Technique for Spring Season.

6.3. Discussion

In this study, we conducted a comparative analysis of multiple deep learning models utilising multitask learning to predict both thermal energy and electrical power outputs. Our evaluation focused on several critical performance metrics: Root Mean Square Error (RMSE), R-squared (R2), Normalised Mean Square Error (NMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE), across two separate domains: thermal energy prediction and electrical power output prediction.

In the domain of thermal energy prediction, the IDAMBi-LSTM model exhibits superior accuracy and reliability, as demonstrated by its lowest RMSE-T values and highest R^2-T percentages in both summer and spring seasons. This consistency across seasons emphasises the model's robustness and its ability to handle seasonal variations effectively. The model's bidirectional LSTM architecture, which captures temporal dependencies more comprehensively than traditional unidirectional models, likely contributes to its enhanced performance. This feature is particularly valuable in energy prediction, where past and future data contexts significantly influence the accuracy of predictions.

The lesser performance of the IAM-GRU and IDAM-GRU models, while still commendable, highlights the limitations of their respective architectures in handling the complexities of thermal energy data as effectively as the IDAMBi-LSTM model. These models showed increased error rates and lower determination coefficients, indicating less precision and reliability, which might limit their applicability in critical or precision-required applications.

Similar trends were observed in the electrical power prediction, where the IDAMBi-LSTM model again stood out with remarkably low RMSE-E values and almost perfect R^2-E scores. The high precision of this model in electrical power forecasting not only supports its robustness but also its potential to significantly impact the efficiency of power management systems. The minimal error metrics like NMSE-E and MAE-E across both seasons further establish this model as a reliable tool for forecasting, critical for maintaining grid stability and optimising power distribution.

The comparative underperformance of the IAM-GRU and IDAM-GRU models in electrical power prediction suggests that while these models are capable of capturing general trends, they do not provide the same level of detail and accuracy as the IDAMBi-LSTM. This is evident from their higher error metrics and lower determination coefficients, which could be particularly problematic in applications where precision is crucial, such as in load forecasting and real-time energy management.

6.4. Future work

The incorporation of physics-informed neural networks (PINNs) could bridge the gap between data-driven approaches and fundamental thermodynamic principles governing PVT systems. Future research could also explore the development of adaptive models that can dynamically adjust their architecture based on changing environmental conditions or seasonal variations, potentially incorporating reinforcement learning techniques for optimal energy management strategies in real-time operations. Furthermore, the proliferation of Internet of Things (IoT) devices and edge computing capabilities opens new avenues for implementing these forecasting models directly at the system level, enabling autonomous decision-making for hybrid PVT plants. The integration of federated learning approaches could allow multiple PVT installations to collaboratively improve forecasting accuracy while maintaining data privacy and security. Additionally, the incorporation of uncertainty quantification methods and probabilistic forecasting could provide more robust predictions that account for the inherent variability in renewable energy systems. Future developments might also focus on multi-modal forecasting that combines satellite imagery, weather radar data, and ground-based measurements to enhance prediction accuracy, particularly for larger-scale PVT installations. As climate change continues to alter weather patterns, the development of climate-resilient forecasting models that can adapt to evolving environmental conditions will become increasingly crucial for the sustainable deployment of hybrid PV–Thermal technologies.

7. Conclusion

The proposed Inception Embedded Dual Attention Bi-LSTM (IDAMBi-LSTM) Model has proven to be a highly effective tool for improving the accuracy of power output and thermal energy forecasts in hybrid Photovoltaic-Thermal (PVT) plants. Traditional forecasting models often fall short in capturing the intricate temporal dependencies and non-linear patterns inherent in PVT systems, leading to less reliable predictions. The IDAMBi-LSTM model addresses these shortcomings by integrating inception layers that capture multi-scale temporal features, dual attention mechanisms that focus on the most significant data points, and Bidirectional Long Short-Term Memory (Bi-LSTM) networks that process information in both forward and backward directions. Through comprehensive comparative analysis with the Inception Embedded Attention Mechanism with GRU (IAM-GRU) and the Inception Embedded Dual Attention GRU (IDAM-GRU) models, the proposed model consistently outperformed its counterparts across all key performance metrics, including Root Mean Square Error (RMSE), R-squared (R^2), Normalised Mean Square Error (NMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE). These results demonstrate the model's ability to deliver more accurate and reliable forecasts for both thermal energy and electrical power outputs in hybrid PVT systems. The superior performance of the IDAMBi-LSTM model not only enhances the predictability and efficiency of energy management in PVT systems but also underscores the importance of advanced deep learning techniques in renewable energy forecasting. By reducing predictive errors and improving forecast accuracy, this model contributes to optimising the operation of hybrid PVT systems, ultimately leading to more efficient and sustainable energy utilisation. The findings of this research highlight the potential for the IDAMBi-LSTM model to set a new standard in the field of renewable energy forecasting, offering a powerful tool for both researchers and practitioners aiming to advance the efficiency and reliability of energy systems.

The deployment of the IDAMBi-LSTM model offers significant benefits for real-world PVT operations. The achieved prediction accuracy (MAPE $< 4.05\%$, $R^2 > 99.79\%$) enables operators to: (1) optimise day-ahead energy dispatch, reducing fossil fuel backup reliance by 15%–20%; (2) implement predictive maintenance by detecting anomalous thermal–electrical patterns before system failure; (3) reduce battery storage over-sizing by 10%–15% while maintaining grid stability; and (4) participate in real-time energy markets where accurate short-term forecasting commands premium pricing.

CRediT authorship contribution statement

Noman Mujeeb Khan: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Qusay Shihab Hamad:** Writing – review & editing, Writing – original draft, Software, Formal analysis, Data curation. **Mohamad Abou Houran:** Writing – review & editing, Writing – original draft, Validation, Resources. **Muhammad Kamran Khan:** Writing – review & editing, Writing – original draft, Resources, Investigation, Conceptualization. **Muhammad Hamza Zafar:** Writing – review & editing, Writing – original draft, Software, Project administration, Methodology. **Filippo Sanfilippo:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used in this work is publicly available.

References

[1] R. Ahmed, V. Sreeram, Y. Mishra, M. Arif, A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization, Renew. Sustain. Energy Rev. 124 (2020) 109792.
[2] S. Han, Y.-h. Qiao, J. Yan, Y.-q. Liu, L. Li, Z. Wang, Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network, Appl. Energy 239 (2019) 181–191.
[3] Y. Jung, J. Jung, B. Kim, S. Han, Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar PV facilities: Case study of South Korea, J. Clean. Prod. 250 (2020) 119476.
[4] L. Liang, T. Su, Y. Gao, F. Qin, M. Pan, FCDT-IWBOA-LSSVR: An innovative hybrid machine learning approach for efficient prediction of short-to-mid-term photovoltaic generation, J. Clean. Prod. 385 (2023) 135716.
[5] U.A. Khan, N.M. Khan, M.H. Zafar, Resource efficient PV power forecasting: Transductive transfer learning based hybrid deep learning model for smart grid in industry 5.0, Energy Convers. Manag.: X 20 (2023) 100486.
[6] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F.J. Martinez-de Pison, F. Antonanzas-Torres, Review of photovoltaic power forecasting, Sol. Energy 136 (2016) 78–111.
[7] Y. Zhang, C. Qin, A.K. Srivastava, C. Jin, R.K. Sharma, Data-driven day-ahead PV estimation using autoencoder-LSTM and persistence model, IEEE Trans. Ind. Appl. 56 (6) (2020) 7185–7192.
[8] P. Kumari, D. Toshniwal, Deep learning models for solar irradiance forecasting: A comprehensive review, J. Clean. Prod. 318 (2021) 128566.
[9] S. Stevenson, A comparison of the forecasting ability of ARIMA models, J. Prop. Invest. Financ. 25 (3) (2007) 223–240.
[10] S. Makridakis, M. Hibon, ARMA models and the Box–Jenkins methodology, J. Forecast. 16 (3) (1997) 147–163.
[11] P. Meer, D. Mintz, A. Rosenfeld, D.Y. Kim, Robust regression methods for computer vision: A review, Int. J. Comput. Vis. 6 (1991) 59–70.
[12] S. Ding, X. Xu, R. Nie, Extreme learning machine and its applications, Neural Comput. Appl. 25 (2014) 549–556.

[13] E. R. Rosca, Stationary and non-stationary time series, USV Ann. Econ. Public Adm. 10 (1) (2011) 177–186.
[14] S. Chakraborty, R. Krishna, Y. Ding, B. Ray, Deep learning based vulnerability detection: Are we there yet, IEEE Trans. Softw. Eng. (2021).
[15] M. Krichen, Convolutional neural networks: A survey, Computers 12 (8) (2023) 151.
[16] C. Xiao, J. Sun, Recurrent neural networks (rnn), in: Introduction To Deep Learning for Healthcare, Springer, 2021, pp. 111–135.
[17] I. Alotaibi, M.A. Abido, M. Khalid, A.V. Savkin, A comprehensive review of recent advances in smart grids: A sustainable future with renewable energy resources, Energies 13 (23) (2020) 6269.
[18] P. Li, K. Zhou, X. Lu, S. Yang, A hybrid deep learning model for short-term PV power forecasting, Appl. Energy 259 (2020) 114216.
[19] R.A. Rajagukguk, R.A. Ramadhan, H.-J. Lee, A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power, Energies 13 (24) (2020) 6623.
[20] L. Zheng, R. Su, X. Sun, S. Guo, Historical PV-output characteristic extraction based weather-type classification strategy and its forecasting method for the day-ahead prediction of PV output, Energy 271 (2023) 127009.
[21] M. Abdel-Nasser, K. Mahmoud, Accurate photovoltaic power forecasting models using deep LSTM-RNN, Neural Comput. Appl. 31 (2019) 2727–2740.
[22] A.Y. Appiah, X. Zhang, B.B.K. Ayawli, F. Kyeremeh, Long short-term memory networks based automatic feature extraction for photovoltaic array fault diagnosis, IEEE Access 7 (2019) 30089–30101.
[23] M. Lotfi, M. Javadi, G.J. Osório, C. Monteiro, J.P. Catalão, A novel ensemble algorithm for solar power forecasting based on kernel density estimation, Energies 13 (1) (2020) 216.
[24] M. Elsaraiti, A. Merabet, Solar power forecasting using deep learning techniques, IEEE Access 10 (2022) 31692–31698.
[25] J.F. Torres, F. Martínez-Álvarez, A. Troncoso, A deep LSTM network for the Spanish electricity consumption forecasting, Neural Comput. Appl. 34 (13) (2022) 10533–10545.
[26] A. Agga, A. Abbou, M. Labbadi, Y.E. Houm, I.H. Ou Ali, CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production, Electr. Power Syst. Res. 208 (2022) 107908.
[27] M.M. Alam, M.H. Rahman, M.F. Ahmed, M.Z. Chowdhury, Y.M. Jang, Deep learning based optimal energy management for photovoltaic and battery energy storage integrated home micro-grid system, Sci. Rep. 12 (1) (2022) 15133.
[28] Ü. Ağbulut, A.E. Gürel, Y. Biçen, Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison, Renew. Sustain. Energy Rev. 135 (2021) 110114.
[29] E. Jumin, F.B. Basaruddin, Y.B.M. Yusoff, S.D. Latif, A.N. Ahmed, Solar radiation prediction using boosted decision tree regression model: A case study in Malaysia, Environ. Sci. Pollut. Res. 28 (2021) 26571–26583.
[30] P. André, L. Dias, S. Correia, A.C. Neto, R. Ferreira, Artificial neural networks for predicting optical conversion efficiency in luminescent solar concentrators, Sol. Energy 268 (2024) 112290.
[31] R. Girimurugan, P. Selvaraju, P. Jeevanandam, M. Vadivukarassi, S. Subhashini, N. Selvam, S.H. Ahammad, S. Mayakannan, S.K. Vaithilingam, Application of deep learning to the prediction of solar irradiance through missing data, Int. J. Photoenergy 2023 (1) (2023) 4717110.
[32] A.M. Noman, H. Khan, H.A. Sher, S.Z. Almutairi, M.H. Alqahtani, A.S. Aljumah, Scaled conjugate gradient artificial neural network-based ripple current correlation MPPT algorithms for PV system, Int. J. Photoenergy 2023 (1) (2023) 8891052.
[33] L. Ricci, D. Papurello, A prediction model for energy production in a solar concentrator using artificial neural networks, Int. J. Energy Res. 2023 (1) (2023) 9196506.
[34] M. Konstantinou, S. Peratikou, A.G. Charalambides, Solar photovoltaic forecasting of power output using lstm networks, Atmosphere 12 (1) (2021) 124.
[35] C. Pan, J. Tan, D. Feng, Prediction intervals estimation of solar generation based on gated recurrent unit and kernel density estimation, Neurocomputing 453 (2021) 552–562.
[36] C. Feng, J. Zhang, W. Zhang, B.-M. Hodge, Convolutional neural networks for intra-hour solar forecasting based on sky image sequences, Appl. Energy 310 (2022) 118438.
[37] Y. Hao, C. Tian, A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting, Appl. Energy 238 (2019) 368–383.
[38] H. Zang, L. Liu, L. Sun, L. Cheng, Z. Wei, G. Sun, Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations, Renew. Energy 160 (2020) 26–41.
[39] A.K. Shaikh, A. Nazir, N. Khalique, A.S. Shah, N. Adhikari, A new approach to seasonal energy consumption forecasting using temporal convolutional networks, Results Eng. 19 (2023) 101296.
[40] J. Qu, Z. Qian, Y. Pei, Day-ahead hourly photovoltaic power forecasting using attention-based CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern, Energy 232 (2021) 120996.
[41] C. Zhan, X. Zhang, J. Yuan, X. Chen, X. Zhang, A. Fathollahi-Fard, C. Wang, J. Wu, G. Tian, A hybrid approach for low-carbon transportation system analysis: integrating CRITIC-DEMATEL and deep learning features, Int. J. Environ. Sci. Technol. 21 (1) (2024) 791–804.
[42] A. Kumar, A. Verma, R. Talwar, Optimal techno-economic sizing of a multi-generation microgrid system with reduced dependency on grid for critical health-care, educational and industrial facilities, Energy 208 (2020) 118248.
[43] J.S. Caicedo-Vivas, W. Alfonso Morales, Short-term load forecasting using an LSTM neural network for a grid operator, Energies 16 (23) (2023).
[44] C. Huang, M. Yang, Memory long and short term time series network for ultra-short-term photovoltaic power forecasting, Energy 279 (2023) 127961.
[45] A. Safari, H. Kharrati, A. Rahimi, Multi-term electrical load forecasting of smart cities using a new hybrid highly accurate neural network-based predictive model, Smart Grids Sustain. Energy 9 (1) (2023) 8.
[46] F. Rafique, L. Fu, R. Mai, LSTM autoencoders based unsupervised machine learning for transmission line protection, Electr. Power Syst. Res. 221 (2023) 109432.
[47] A. Safari, Hybrid emerging model predictive data-driven forecasting of three-phase electrical signals of photovoltaic systems using GBR-LSTM, E-Prime - Adv. Electr. Eng. Electron. Energy 8 (2024) 100549.
[48] M. Abou Houran, S.M. Salman Bukhari, M.H. Zafar, M. Mansoor, W. Chen, COA-CNN-LSTM: Coati optimization algorithm-based hybrid deep learning model for PV/wind power forecasting in smart grid applications, Appl. Energy 349 (2023) 121638.
[49] F. Veynandt, F. Inschlag, C. Seidl, C. Heschl, Measurement data from real operation of a hybrid photovoltaic-thermal solar collectors, used for the development of a data-driven model, Data Brief 49 (2023) 109417.