Accurate and Efficient Short-Term Solar Power Forecasting Across Climatic Regions

MSc Research Project

MSc Data Analytics - C

NIHAL RAJ REDDY DANDA

Student ID: x23317612

School of Computing

National College of Ireland

Supervisor: Harshani Nagahamulla

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: …NIHAL RAJ REDDY DANDA……………………..……………………………

Student ID: …X23317612……..………………………………………………………………

Programm : …DATA ANALYTICS……………………..………. **Year:** ……2023………..

Module: …Research Practicum…………………………………………………………..……

Supervisor: …Harshani Nagahamulla……………………………………………………. ………

Submission Due Date: ….15 september 2025……………………………………………………………….

Project Title: ….Accurate and Efficient Short-Term Solar Power Forecasting Across Climatic Regions………………………………………………………………….

Word Count: …7302……………… **Page Count**………22……………………………………….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: …nihalreddy………………..

Date: ..10/08/2025………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only

Signature:	
Date:	
Penalty Applied (if applicable):	

Accurate and Efficient Short-Term Solar Power Forecasting Across Climatic Regions

Nihal Raj Reddy Danda

x23317612

Abstract

Accurate short-term solar power forecasting is essential for ensuring grid stability, optimizing renewable energy integration, and enabling intelligent energy management across diverse geographic and climatic regions. While traditional and deep learning-based models have improved prediction accuracy, they often struggle with balancing computational efficiency and generalizability. This study proposes a hybrid Gated Recurrent Unit (GRU) model enhanced with an attention mechanism to address these limitations. The GRU component captures sequential dependencies in solar power data with reduced computational overhead, while the attention layer selectively emphasizes critical time steps, improving accuracy and interpretability. Experimental evaluations across multiple countries India, Brazil, Norway, and Australia demonstrate the model's superior performance, achieving the lowest MAE and RMSE compared to standalone ANN and GRU models. The results validate the GRU-attention model's ability to offer high accuracy, lightweight architecture, and adaptability across regions with varied climatic conditions, making it an effective and scalable tool for real-time solar forecasting in modern energy systems.

1 Introduction

1.1 Background and Problem Context

The world energy system is now undergoing a radical transformation because of the increasing climate change concerns, growing energy needs and the acute need to reduce the reliance on fossil fuels. With its environmental sustainability, scalability in modules, and the rapidly decreasing cost of implementation, solar photovoltaic (PV) technology has become one of the most popular and promising solutions to other renewable sources of energy. Solar energy systems are being increasingly used by countries throughout the economic spectrum, including developed countries with well-developed grid infrastructure and developing regions that need energy equity, as part of their transition to a low-carbon future (Zhang et al., 2021; Hanif & Mi, 2024).

Although solar energy has a very high potential, it poses a lot of operational problems mainly due to its very nature of being variable and being subject to the environmental factors like solar irradiance, temperature and weather conditions. Such fluctuations may lead to instability in the supply of energy, particularly in power grids where renewable sources are heavily penetrated. This has led to the inability to accurately predict, in the short term, Minutes to hours ahead (minutes to a few hours ahead) is turning into a critical necessity. It allows the grid operators to make better decisions in load balancing, dispatch scheduling, battery storage control, and demand response optimization. Precise forecasting is also critical to smart microgrids, charging systems of electric vehicles, and decentralized energy trading platforms (Wang et al., 2023). The traditional solar forecasting was based on physical and statistical models, including Numerical Weather Prediction (NWP), autoregressive approaches and regression-based approaches. Although helpful, these models tend to lack in the representation of the nonlinear, dynamic, and chaotic trends of solar energy production (Park et al., 2023). Over the last few years, the ability to forecast has been enhanced greatly with the introduction of machine learning (ML) and deep learning (DL) methods, which learn complex patterns in the historical data. Support Vector Regression (SVR), Random Forest (RF), and Gradient Boosting Machines (GBM) are some of the algorithms that have demonstrated better accuracy when using historical data on power generation and weather features. Nevertheless, these models continue to have limitations in sequential learning and tend to have difficulties in temporal dependencies.

In response to this, scientists have resorted to Recurrent Neural Networks (RNNs) namely Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks that are able to model time-series data as well as capture temporal dependencies. LSTM models have been very successful in the solar power prediction owing to their long-term memory, and capacity to model nonlinear relationships. Nevertheless, LSTMs are computationally demanding and tend to overfit especially when trained on noisy or location-specific data.

GRUs, a simplified version of LSTMs, have been found to have comparable predictive performance with a lower computing burden. GRUs have fewer parameters and converge quicker, which is especially appealing to applications in real-time forecasting and edge deployment settings where computation resources are constrained. However, the RNN-based models, be it LSTM or GRU, take the input sequentially and equally across all time steps and assign equal weight to each input. This shortcoming may cause the model to fail to isolate and highlight most pertinent temporal characteristics.

To counter this, time-series modeling has been given attention mechanisms. Attention mechanisms were originally created in the context of Natural Language Processing (NLP) to enable neural networks to dynamically attend to the most significant components of the input sequence when predicting. In solar forecasting terms, this implies that the model is able to pay attention to particular time instances that are more influential in power output. More recent studies have effectively combined attention layers with LSTM or GRU to achieve better accuracy and interpretability of forecasts, especially in climate-sensitive tasks.

In the meantime, transformer-based models, where the whole model is based on attention mechanisms and do not use recurrence, have shown state-of-the-art results in many time-series tasks, such as solar forecasting. Nonetheless, these models have serious trade-offs: they are computationally costly, need a lot of data, and are hard to implement in edge devices or lowresource settings. They are also less interpretable and harder to tune to specific geographies because they are complex.

With such a landscape, there is a necessity of a model that balances the accuracy, computational efficiency, and generalizability. This paper suggests a hybrid model, which combines Gated Recurrent Units (GRU) and a lightweight attention mechanism to meet these goals. The GRU component makes use of temporal dynamics in solar generation, and the attention layer increases model focus on the most important time steps, thus driving the improvement in forecasting accuracy without adding much computational overhead. Moreover, this study will test the model in various geographically and climatically different areas such as countries like Brazil, India, Norway, and Australia to show the strength of the model and its applicability in international contexts.

1.2 Motivation:

The increasing need to accommodate real-time operations in various solar power settings is why there is an increasing demand in scalable and efficient forecasting models. Although transformer architectures have increased the predictive accuracy bar, their parameter sizes, training complexity, and resource-intensive nature do not make them suitable in most practical implementations.

Conversely, other models such as LSTM and CNN-LSTM are more effective and common in time-series prediction but prone to vanishing gradients and short-term memory. Their context windows are also of fixed length, which is unable to capture the fast changing solar dynamics, particularly in areas with high climatic variability. Additionally, the current methods tend to be overfitted to particular geographies, thus lack portability and generalizability to various climatic regions.

This research can overcome these challenges by suggesting a GRU-attention hybrid model as a lightweight and flexible solution. GRUs are computationally efficient relative to LSTMs, and still have the capability to learn sequential dependencies. The model is able to dynamically attend to the most relevant time steps by incorporating a simple attention mechanism, which increases both interpretability and performance in different weather conditions. This architecture offers the best tradeoff-it allows increasing the accuracy of the predictions without sacrificing efficiency or portability.

1.3 Research Question

> ***How can a Gated Recurrent Unit (GRU) model integrate with an attention mechanism outperform existing models in terms of accuracy, computational efficiency, and generalizability across multiple regions and diverse climatic conditions for shortterm solar power forecasting?***

Justification

This study is motivated by the urgent need for practical and adaptive forecasting tools that are not only accurate but also lightweight and deployable across a range of operational environments. The majority of high-performance models reviewed in the literature (Zhang et al., 2021; Hanif & Mi, 2024) are limited either by large memory footprints or by their reliance on location-specific data and infrastructure.

This paper fills these gaps by suggesting the GRU-attention model, which integrates the effectiveness of GRUs with a narrowed attention mechanism to increase adaptability. The architecture can be deployed in microgrids, smart inverters, or even mobile forecasting devices unlike transformer based models, which necessitate a lot of computational resources and training data. Also, the fact that the model was validated in a number of countries with different solar profiles offers a good basis to evaluate its generalization capacity.

Finally, the study adds a scalable, interpretable, and resource-efficient method to solar power forecasting, which can support smarter energy systems that are more resilient to environmental uncertainty and variability.

2 Literature Review

2.1 Machine Learning-Based Approaches for Solar Forecasting

Machine learning (ML) methods have long been known to find hidden patterns and relationships in solar energy data and, therefore, can be applied to short-term and medium-term forecasting. The most common methods that have been applied due to their stability and simplicity include random forest (RF), Gradient Boosting, and Support Vector Machines (SVM).

Mohamed et al. (2022) suggested a dynamic ML model to predict solar energy in microgrid systems. Their feature engineering helped them to determine the important environmental and operational predictors and make the model sensitive to the current changes in data. Similarly, Rangelov et al. (2023) have proposed a hybrid framework that combines RF, Deep Neural Networks (DNNs) and Long Short-Term Memory (LSTM) models and confirmed that ensemble methods can outperform single-model baselines since they capture nonlinearity and time dependencies.

Wang et al. (2023) followed this pattern and integrated sky images and historical weather records in a hybrid ensemble model and demonstrated that the accuracy of short-term irradiance forecasts was increased substantially. Besides, Mitrentsis and Lens (2022) proposed a probabilistic forecasting approach that is interpretable and based on Natural Gradient Boosting, which allows not only point forecasting but also quantifying uncertainty, which is essential when incorporating solar power into the grid.

Nevertheless, ML approaches usually need a lot of domain expertise to feature engineer manually and are limited in their temporal learning ability, restricting their application in modeling long-range dependencies or changing weather patterns (Sudharshan et al., 2022). This shortcoming has prompted the effort to transition to solutions based on deep learning.

2.2 Deep Learning and Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) and their variants have become a superior alternative because deep learning (DL) methods, in general, can automatically learn hierarchical features and long-term dependencies on sequential data. LSTMs in particular have been useful in the prediction of solar power.

Jailani et al. (2023) confirmed the robustness of LSTM-based models to model the temporal dynamics of solar irradiance, and they outperformed conventional ML models on a number of performance measures. Building on this, Lim et al. (2022) proposed a CNN-LSTM hybrid model that learns spatial features of environmental input first and subsequently models their temporal dynamics, consistently leading to a marked improvement in multi-step forecasting settings.

In order to overcome limitations like vanishing gradients, and fixed-size context windows, recent studies have investigated attention mechanisms. Park et al. (2023) introduced an RNNattention model that dynamically assigned weights to the time windows that matter, thus increasing the accuracy of predictions and their interpretability. Similarly, Rai et al. (2023) suggested a Differential Attention Network that is able to respond to fast variations in solar radiation and make the models more responsive to local weather anomalies.

Although they are effective, RNN-based models tend to require a lot of computing resources and time to train. These disadvantages restrict their use in resource-limited systems, like ongrid edge devices or real-time systems.

2.3 Transformer Models and Hybrid Attention-Based Architectures

Transformers, first popularized in natural language processing, are gaining use in solar forecasting because they can model long-range dependencies without being limited to sequential processing. Transformers have the ability to capture global interactions among timeseries inputs better than RNNs by leveraging multi-head self-attention.

Hanif and Mi (2024) showed that transformer-based models were more robust and more accurate in forecasting than traditional RNNs and LSTMs. The parallel processing aspect of their architecture lowered the latency of inference on a batch processing, but it was at the expense of significant memory consumption. Zhang et al. (2021) further expanded this paradigm by suggesting a GPT-inspired ultra-short-term forecasting model, which has the ability to make minute-level PV power predictions. Nevertheless, its scalability was constrained by its complexity of training and inference latency.

Hybrid architecture has come up to overcome these problems. Kiasari and Aly (2024) compared RNNs, XGBoost, LSTM, KNN, and RF on lagged steps and cross-validation conditions and found transformer variants to be the best performers. However, they also raised issues of computational cost, interpretability of models, and sensitivity of data that makes these models less practical in deployment where computing infrastructure is limited.

These results show the necessity of lightweight, attention-augmented models capable of achieving the accuracy of transformers but with low computational cost and interpretability.

2.4 Research Gap and Motivation for GRU with Attention

Although both ML and DL models have been used to make progress in solar forecasting, one of the challenges that still exists is attaining a trade-off between the accuracy of the prediction, interpretability, and efficiency. Ensemble ML models and CNN-LSTM hybrids have certain advantages and frequently fail to handle long-range dependencies or are highly featureengineered. Although transformer-based models perform well, they consume a lot of resources and are highly susceptible to overfitting on small or noisy data.

This opens a significant research problem: building lightweight models with the advantages of deep learning but a minimal resource requirement. Gated Recurrent Units (GRUs) provide a simplified version of LSTMs with less number of parameters and training time hence they are applicable in real-time and edge implementations (Saxena et al., 2024; Ledmaoui et al., 2024). GRUs can dynamically attend to salient time steps, and with the addition of attention mechanisms, this enables both better accuracy and model transparency.

Also, the majority of the previous investigations use data collected in one place or climate region, which brings up the issue of generalization. This paper fills this gap by testing a GRUattention hybrid model on a variety of geographically distributed datasets, with different climates and irradiation patterns. This multi-regional assessment does not only challenge the scalability of the model but also its resilience to environmental variability a key requirement to deploy globally in smart grids and renewable energy forecasting systems.

2.5 Comparative Analysis of Literature on Solar Power Forecasting

Table 1. Comparison of Existing Research

Author(s) & Year	**Methodology**	**Model(s) Used**	**Pros**	**Cons**
Zhang et al. (2021)	Ultra-short-term forecasting with GPT pretraining & finetuning	Generative Pretrained Transformer (GPT)	Addresses data scarcity; lightweight finetuning	Large model size; reliance on synthetic/virtual data
Mohamed et al. (2022)	Feature engineering and machine learning for microgrid forecasting	XGBoost with feature selection	Reduces input dimensions; improves reliability; suitable for microgrids	Less effective for ultra-shortterm; depends on meteorological inputs
Wang et al. (2023)	Hybrid ensemble learning using historical + sky images	XGBoost with novel loss function	Robust to extreme values; combines temporal and spatial features effectively	Requires highquality sky images; parameter tuning complexity

Park et al. (2023)	Real-time solar estimation with RNN	RNN model	Improves accuracy; reduces need for metering infrastructure	Needs sample site data; model complexity and training cost
Jailani et al. (2023)	Review of standalone and	LSTM and hybrid models	Captures temporal dependencies;	Hybrids need longer training;
	hybrid LSTM models		hybrids improve accuracy	potential overfitting risk
Rangelov et al. (2023)	AI-based shortterm forecasting with readily available weather data	Random Forest, DNN, LSTM	Uses broadly accessible weather data; cost-efficient	Limited to short-term horizons; longer LSTM training times
Hanif & Mi (2024)	Review of transformerbased forecasting	Transformer (single, hybrid)	High accuracy; efficient parallel computing; handles long sequences	High computational cost; needs large diverse datasets
Saxena et al. (2024)	Deep learning load and renewable forecasting with adaptive control	LSTM with enhanced control schemes	Enhances grid reliability; captures longterm dependencies	Complex controller design; needs grid integration

2.6 Discussion and Key Insights

Table 1 gives an overview of a wide range of new methods to predict solar power, demonstrating the shift in the methods of solar power prediction, moving away from standard machine learning models, such as XGBoost and Random Forest, to deep learning models, such as LSTM, GRU, and Transformer architectures. Zhang et al. (2021) used pre-training and finetuning of GPT-based models to overcome the problem of data insufficiency, so they are flexible and effective in low-data settings. Their practicality to real-time applications or lowresource settings is, however, hampered by the vast amount of synthetic data they use and the size of the model. In the same way, Mohamed et al. (2022) applied XGBoost with feature selection to enhance forecasting in microgrids. This simplified the feature complexity and enhanced the reliability but the approach was limited by the quality and availability of meteorological data inputs.

The trend of ensemble and hybrid methods can be traced in Wang et al. (2023), who integrated historical information and sky images to improve the accuracy of forecasts via a new loss function in XGBoost. The method is suitable to capture spatial and temporal patterns, but it is restricted by the need of high quality imagery and computational tuning activities. In the meantime, Park et al. (2023) used RNNs to do real-time solar estimation without the requirement of large metering infrastructure. This is a correct approach but requires sitespecific data and has a high training overhead, which makes deployment and scalability more challenging in diverse regions.

Recent works still focus on deep learning, as evidenced in Jailani et al. (2023) and Rangelov et al. (2023), who investigated standalone and hybrid LSTM models. These architectures are efficient in learning the temporal dependencies, particularly on the sequential data such as solar output. However, they tend to take more time to train and they are prone to overfitting in case the complexity of the model is not proportionate to the data volume. The fact that Rangelov et al. use more accessible weather data in models such as LSTM and Random Forest would encourage cost-effective performance but would limit performance to short-term horizons. It is scalable because of the use of publicly available data, which is a strength, but it will be less accurate in more dynamic or longer forecasting windows.

Transformer-based approaches are becoming more popular recently. Hanif & Mi (2024) pointed out the benefits of transformers in their ability to process long sequences and perform parallel computation. Such models are computationally costly and have a high potential of accuracy, but need to be trained on large, diverse datasets in order to generalize. This was solved by Saxena et al. (2024) who introduced adaptive control into the LSTM-based models to enhance responsiveness to grids and long-term predictions. Nonetheless, the complexity of the model and the difficulty of deployment of such control systems is augmented. Collectively, the literature suggests an apparent need to develop a model that achieves a trade-off between accuracy, efficiency, scalability, and interpretability a gap that GRU with attention mechanisms well fills, since they provide a trade-off between complexity and performance in different solar settings.

3 Methodology

The approach taken in the research is aimed at systematically designing, deploying, and testing a deep learning-based solar forecasting model that is both accurate, interpretable, and computationally efficient. It is based on the CRISP-DM (Cross-Industry Standard Process for Data Mining), a framework that organizes the workflow into several stages beginning with data acquisition and preprocessing all the way to exploratory data analysis, model development, performance evaluation, and validation. Such a methodology will not only allow ensuring methodological rigor but also practical relevance, especially when evaluating the flexibility of the model in various climatic regions. The essence of such methodology lies in the design and implementation of a GRU-based neural network augmented with an attention mechanism, which is appropriate in managing a time-series data with time-dependencies. The next subsections describe the architecture design, data processing pipeline, and evaluation strategy that were used to attain the research objectives.

3.1 GRU-Attention Architecture and Design Rationale

This paper attempts to use a GRU-attention model to address the weaknesses that are usually witnessed in both the conventional statistical models and the complicated deep learning models that have been applied in short-term solar power prediction. Although approaches like LSTM and Transformer models have demonstrated good predictive performance, they are associated with significant computational cost, overfitting and low deployment in real-time or resourcelimited settings. GRUs (Gated Recurrent Units), a less complex version of LSTMs, have the same accuracy but fewer parameters. But, as with any recurrent model, GRUs do not distinguish between more and less relevant temporal patterns, and thus are limited in their capacity to do so, a problem that is particularly troublesome with highly dynamic and weathersensitive data.

In order to overcome this drawback, the GRU component of this paper is coupled with a lightweight attention mechanism. Once the temporal dependencies have been learned by the GRU based on the input sequence, the attention layer is used which applies learned weights to each hidden state so that the model can selectively attend to important time steps that are more influential to the forecasting target. Such dynamic re-weighting does not only increase the interpretability of the model (i.e. indicating which time intervals contributed most to the prediction) but also increases the precision of the forecast, especially in climates with highly variable solar conditions like monsoons in India or seasonal extremes in Norway. The attention mechanism enables the model to perform better than uniformly sequential baselines with no additional architectural complexity, by guiding computation to the information that is relevant. The last GRU-attention architecture therefore offers a very interesting trade-off between computational costs, prediction performance, and generalizability. In contrast to transformerbased models, which need large data and extensive training, this model is appropriate to be deployed in the real world at the edge, in microgrids, solar farms, and inverterintegrated systems. Moreover, its flexibility was confirmed by experiments in several climatic regions such as Brazil, India, Australia, and Norway. This validates the model in terms of scalability and resilience, which proves its usefulness in real-life solar forecasting applications where low latency and high accuracy are essential.

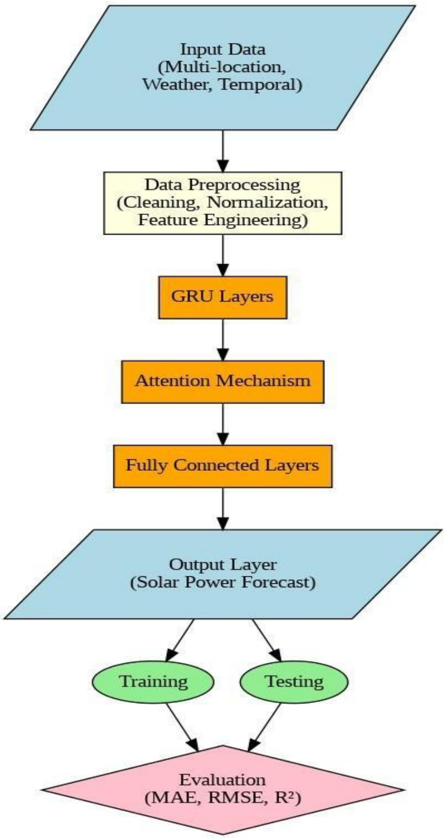


Figure 2 .Flow Chart of Proposed Work

3.2 Sequence of Steps Followed (CRISP-DM Methodology)

This paper presents a systematic approach to the problem of short-term solar power forecasting that seeks to build a computationally efficient and interpretable deep learning model. The data collection is followed by the data preprocessing, exploratory data analysis, model implementation, and evaluation. The data is gathered in the form of a CSV file and contains hourly measurements of solar power at different geographical locations with attributes that include DC/AC power, temperatures, solar irradiation, energy yield, time and location metadata.

The preprocessing was carried out by transforming the column of date into the datetime type to obtain features such as day, month, and year so that the model could be able to capture the seasonal patterns. Columns that were not relevant or redundant like the column DateTime were removed and the column Country was kept to allow cross regional analysis. The IQR method was used to remove outliers in numerical features like power and temperature to decrease the skew and increase the robustness of the model. There was little missing data and this did not need to be imputed.

The analysis was then followed by an exploratory data analysis (EDA) step, in which univariate and multivariate visualizations were used to determine the pattern over time, weather, and location. Histograms, scatter plots, and time-series graphs were used to identify the important relationships especially between the solar power output and environmental variables. The insights informed the choice of the relevant features to be modeled. All numerical characteristics were normalized by Min-Max Scaling to guarantee consistency of scale. The modeling stage started with the application of classical machine learning models, such as linear regression, decision trees, SVR, and random forests. These gave baseline performance, but were constrained in temporal dependencies. Then, architectures of deep learning such as LSTM and GRU were studied. The primary emphasis was put on a GRU model with an attention mechanism, selected due to its capacity to strike the right balance between the accuracy of forecasts and computational cost. The attention layer allowed the model to attend to relevant temporal inputs, increasing interpretability without resource costs of transformer models. Standard regression-based measures of model accuracy and stability including RMSE, MAE, and MAPE were used to compare models. One of the strengths of the paper is its cross-regional validation approach, which applied the GRU-attention model to climates and geographies that differ in their weather and geography. As opposed to the majority of previous studies that have been limited to single-site data, the method guarantees greater generalizability and demonstrates the practical applicability of the model in the deployment of smart grids, solar farms, and microgrid systems that can be used in different environmental conditions.

4 Implementation

4.1 About dataset

The data used in this research comprises of the hourly data on the solar power generation and includes the environmental and time related factors of the data, with each record being labeled by the country Norway in this case. Notable characteristics are direct current (DC) and alternating current (AC) power, which is initially near zero, probably indicating early morning hours when solar power is naturally low. Other environmental variables are ambient temperature and module temperature that can be used to determine the effect of weather conditions on panel efficiency. The module temperature is normally a few degrees above ambient temperature because the solar panels absorb heat.

Irradiation, a critical feature representing the solar energy received per unit area, also registers low or zero values in the early records, aligning with the observed lack of power generation. Temporal attributes such as day, month, hour, and year enable time-series modeling and seasonal trend analysis. The dataset further includes daily and total energy yield measurements. While the daily yield remains constant at the beginning of each day, the total yield increases over time, reflecting the cumulative energy production offering a reliable metric to track overall system performance throughout the day.

3.2 DATA PREPROCESSING

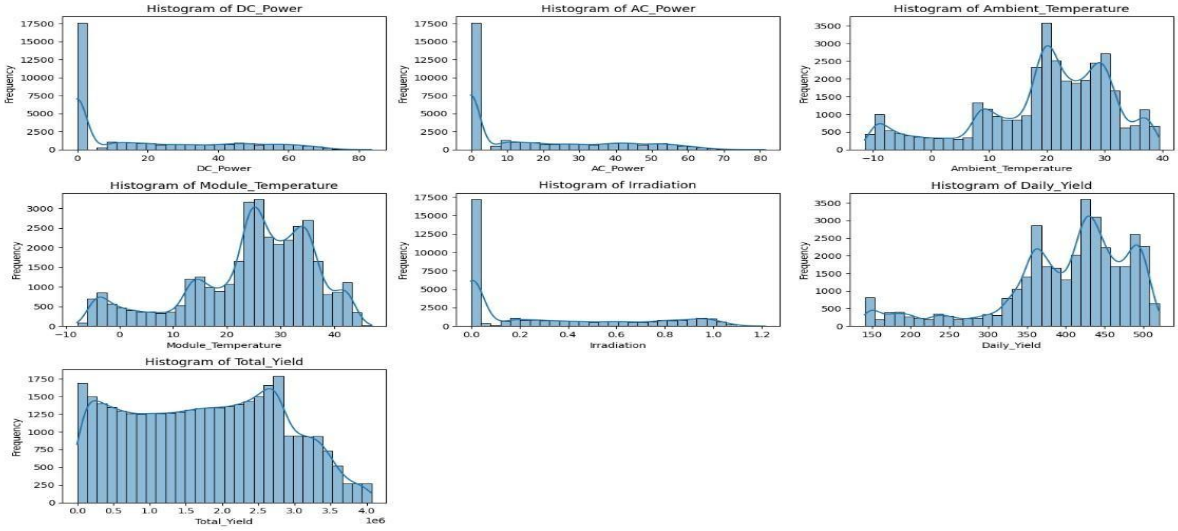
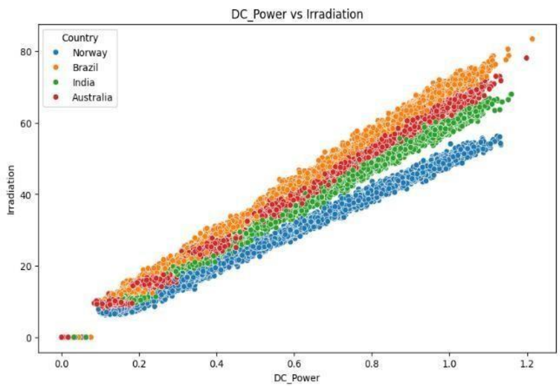
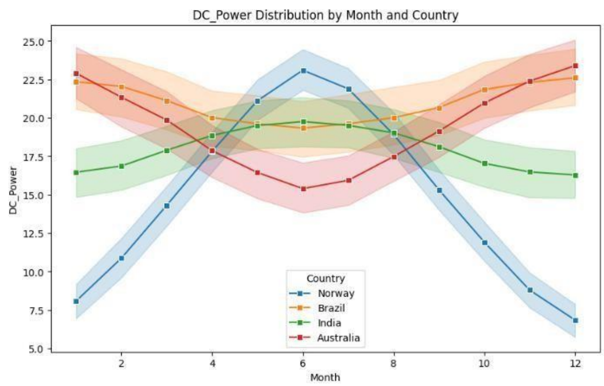


Figure 1A. Histogram Plots

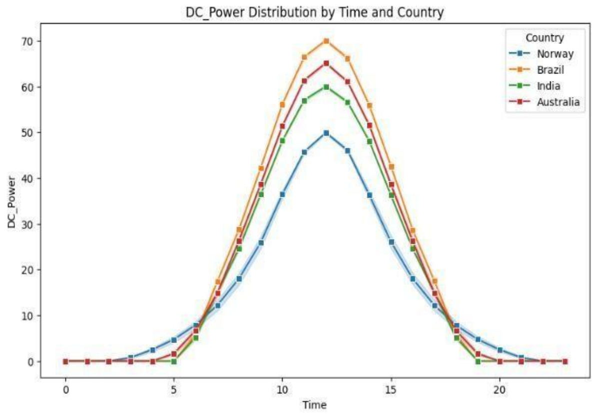
As shown in figure 1A the Histogram plots show that DC and AC power, as well as irradiation, are right skewed with most values near zero, reflecting low output during early hours. Ambient and module temperatures have varying distributions due to seasonal changes, with module temperature generally higher. Daily yield has noticeable peaks, while total yield gradually increases. These patterns highlight the data's non-linearity and the need for scaling and advanced modeling techniques.
As shown in figure 1B the scatter plot shows a clear positive relationship between DC power and irradiation across Norway, Brazil, India, and Australia confirming that higher solar radiation leads to greater power output. Country-specific clusters are visible, suggesting regional differences in efficiency and environmental conditions. Brazil shows the highest DC power for a given irradiation level, while Norway shows the lowest. These patterns highlight the impact of geography and climate on solar performance and the importance of building models that generalize well across regions.



1B. DC Power vs Irradiation



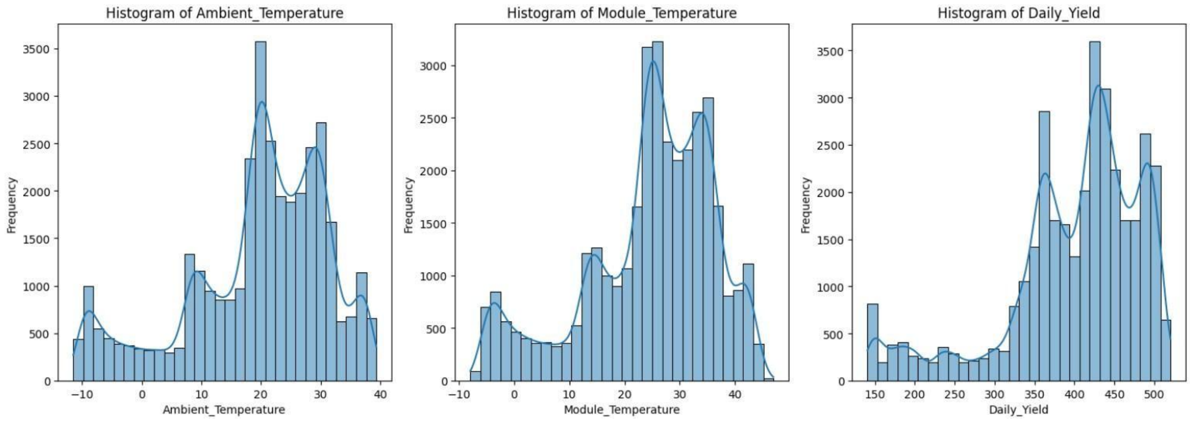
1C. Monthly DC Power Trends



1D. Hourly DC Power Output

```python
def cap_outliers_iqr(df, column):
    Q1 = df[column].quantile(0.25)
    Q3 = df[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR

    df_capped = df.copy()
    df_capped[column] = df_capped[column].clip(lower=lower_bound, upper=upper_bound)

    return df_capped

df_capped = cap_outliers_iqr(df, 'Ambient_Temperature')
df_capped = cap_outliers_iqr(df_capped, 'Module_Temperature')
df_capped = cap_outliers_iqr(df_capped, 'Daily_Yield')

def count_outliers_after_capping(df, column):
    Q1 = df[column].quantile(0.25)
    Q3 = df[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    outliers = df[(df[column] < lower_bound) | (df[column] > upper_bound)]
    return len(outliers)

print("Outliers after capping:")
for col in ['Ambient_Temperature', 'Module_Temperature', 'Daily_Yield']:
    outlier_count = count_outliers_after_capping(df_capped, col)
    print(f"{col}: {outlier_count} outliers")
```

1E. Outlier Handling



1F. Histograms After Outlier Handling

Figures 1C and 1D illustrate clear temporal patterns in solar power generation across Norway, Brazil, India, and Australia. Monthly trends (Figure 1C) show Norway peaking in summer and dipping in winter due to daylight variation, Brazil maintaining steady output year-round, India showing minor monsoon-related dips, and Australia peaking in December–January. Hourly trends (Figure 1D) reflect typical bell-shaped solar cycles, with Brazil having the highest midday output, followed by Australia, India, and Norway. These patterns highlight the need for models like GRU-attention that can adapt to regional and temporal variations. Figures 1E and 1F detail the IQR-based outlier handling applied to ambient temperature, module temperature, and daily yield. Post-capping, the histograms show cleaner distributions with reduced anomalies, improving data quality and model stability. This preprocessing ensures that critical patterns are preserved while minimizing noise—key for training accurate and efficient forecasting models.

4.2 Baseline Model Selection and Justification

In the proposed study, DC power (y) was considered the target variable that is the real-time electrical output of the solar panels. All the rest of features, such as environmental conditions, temporal attributes, and yield measurements were considered as input variables (X) to construct predictive models. . To forecast DC power, a number of high-performance ensemble learning algorithms were employed namely, LightGBM (LGBM), Random Forest Regressor (RFR), and XGBoost (XGB). The reason to select these models is that they have been proven to handle

structured, tabular data and have the ability to represent complex and non-linear dependencies between features.
In particular, LightGBM was chosen because of its ability to work with large datasets and efficiency because of histogram-based learning that reduces the time of training and memory usage. Random Forest was added since it is robust and interpretable, it gives the information on the importance of features and automatically mitigates overfitting by averaging over a large number of different models. It was trained with XGBoost which has a regularized boosting framework to enhance the predictive accuracy and generalization. All three models are suitable to regression problems like the solar power forecasting and have outperformed the conventional algorithms in other real world tasks consistently. The synergistic benefits provide them with a strong basis upon which more complicated deep learning structures can be compared at a later time.

Table 2: Performance Comparison of Machine Learning & Deep Learning Models

Algorithm	MAE	MSE	RMSE
XGBoost	4.61	30.51	5.52
RF Regressor	2.69	18.88	4.34
LightGBM	5.08	36.60	6.05
ANN	0.86	1.20	1.09
GRU	1.09	2.08	1.44
GRU + Attention	**0.53**	**0.53**	**0.73**

5 Evaluation and Results

5.1 Machine Learning

The results of the comparative analysis of the performance of the three ensembles learning algorithms XGBoost, Random Forest Regressor (RF), and LightGBM reveal different results on solar power output prediction, especially when it comes to error measures like MAE (Mean Absolute Error), MSE (Mean Squared Error), and RMSE (Root Mean Squared Error) as shown in Table 2.

The Random Forest Regressor performs the best of all three models. It has the least MAE of 2.69 meaning that its predictions are on average nearest to the true values. Also, it has the lowest MSE of 18.88 and RMSE of 4.34 which implies that the error is more widely dispersed and the overall predictive power is higher. It implies that RF is a suitable method to detect the patterns in the data without overfitting, which means that it is especially efficient in this issue.

In second place, there is XGBoost, which has an MAE of 4.61 and RMSE of 5.52. Though it is a very strong boosting method that has the potential to model complex relationships, in this

case, it performs worse than RF. This is possibly because XGBoost is sensitive to hyperparameter settings or the nature of the dataset. Nevertheless, it does much better than LightGBM in overall accuracy.
LightGBM, as fast and efficient as it is, reports the largest error values in all metrics in this case MAE of 5.08, MSE of 36.60 and RMSE of 6.05. This implies that the model either underfit or did not learn complex relationships among the features, which could be because of its default parameterization or shallow depth parameters.

All in all, Random Forest is the most reliable and accurate model to be used in this solar energy forecasting problem according to the metrics provided. The findings support the argument that the performance of model should not be measured solely based on theoretical efficiency, but also the quality of prediction on particular datasets.

5.2 Deep Learning

The three most important evaluation measures, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) were used to compare the performance of three forecasting models, Artificial Neural Network (ANN), Gated Recurrent Unit (GRU), and the proposed GRU with Attention mechanism. The outcomes are tabulated as below: The comparison of the three forecasting models Artificial Neural Network (ANN), Gated Recurrent Unit (GRU), and GRU with an attention mechanism integrated provides clear answers to the efficiency of each of the methods. When compared using three common metrics Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) it can be seen that the trend is consistent, that is, the GRU + Attention model has the lowest error rates on all three metrics, making it the most accurate and efficient model in shortterm solar power forecasting. Although the ANN and GRU models exhibit different levels of predictive performance, none of the two can be compared to the higher accuracy and robustness that the attention-enhanced GRU model exhibits.
On the other hand, the ANN model, being a feedforward architecture, lacks any inherent temporal awareness. It operates on fixed-length input windows without leveraging the sequential nature of data. Consequently, while the ANN showed moderate performance, it was outperformed by both recurrent models due to its inability to model evolving trends and dependencies inherent in solar generation. However, its relatively lower MAE compared to the baseline GRU hints that, without enhancements like attention, even recurrent models can struggle in capturing complex non-linear dynamics unless augmented appropriately.

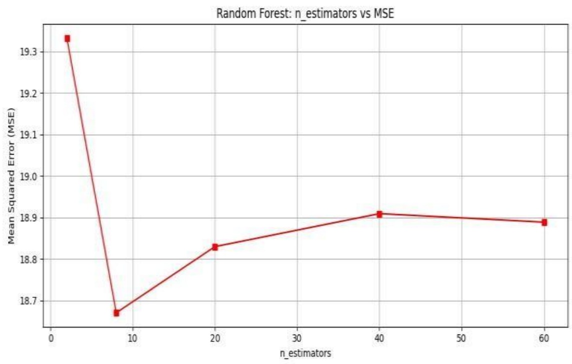
5.3 Superiority of GRU with Attention

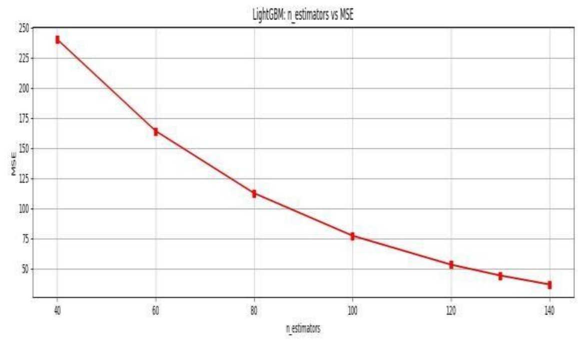
The results across all models clearly demonstrate the superior performance of the GRU with Attention architecture, establishing it as the most accurate and robust forecasting solution among both machine learning and deep learning approaches. While Random Forest emerged as the best among classical ML models, its MAE (2.69) is significantly higher than the GRUAttention's MAE of 0.53, highlighting the advantage of sequence-aware deep models in capturing temporal dynamics. Even the baseline GRU, though designed for time-series tasks, fell short due to its uniform treatment of all time steps. The attention mechanism resolves this limitation by enabling the model to dynamically focus on the most influential segments of the input sequence—such as midday solar peaks or abrupt weather changes—thereby improving precision and adaptability. In addition, GRU-Attention is characterized by a lightweight structure and a low computational overhead that allow it to be implemented in real time even in other contexts, including edge devices and microgrids. Its performance in different regions such as India, Norway, Brazil, and Australia is consistent, hence its generalizability and the ability to withstand climatic variation. All these benefits contribute to the fact that the GRUAttention model is a scalable and future-proof solution to short-term solar power forecasting.

To conclude, the empirical findings are very much in line with the effectiveness of the GRU + Attention model in short-term solar power prediction. It provides an attractive mix of low forecasting error, solid treatment of sequential data, and enhanced interpretability, with computational efficiency that is appropriate to real-time or embedded forecasting systems. These results do not only confirm the suggested architecture but also highlight the necessity of the temporal dependencies modeling that takes into account attention in domains where time sensitivity is a crucial aspect to the correct predictions and where the environment is highly variable. Future work can extend this and consider multi-head attention, cross-region transfer learning and transformer extensions to improve performance and generalization.

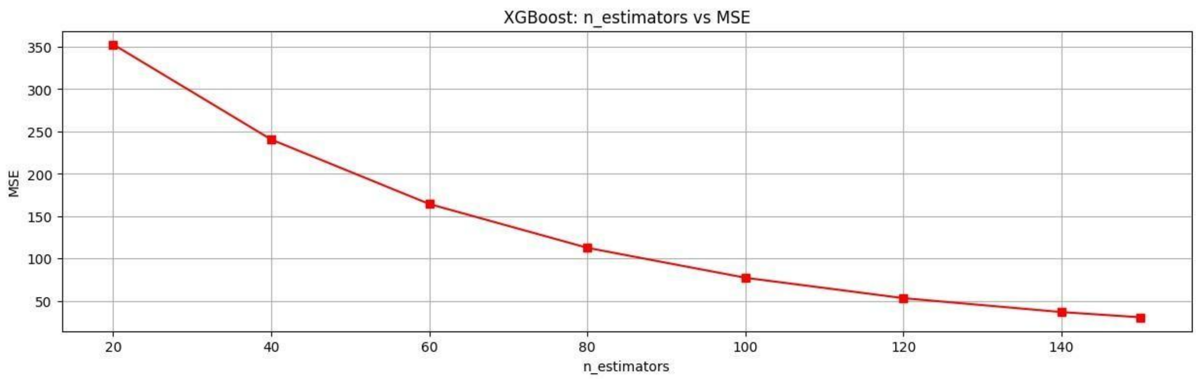
4 Hyperparameter Tuning and Conclusion

4.1 Machine Learning





2A. RFR number of estimators vs MSE 2B. LGBM number of estimators vs MSE



2C. XGB n_estimators vs MSE

Figures 2A to 2C illustrate how increasing the number of estimators (n_estimators) affects the Mean Squared Error (MSE) in Random Forest, LightGBM, and XGBoost models. In Figure

2A, Random Forest shows a sharp MSE drop from 2 to 8 estimators, after which the improvement plateaus, indicating diminishing returns. In contrast, Figure 2B reveals that LightGBM benefits consistently from more estimators, with MSE decreasing steadily from 240 to below 40 as n_estimators increase from 40 to 140. Similarly, Figure 2C shows XGBoost's MSE falling from around 350 to just under 40 as estimators rise from 20 to 150, with steep early gains and gradual improvement at higher counts. Together, these plots highlight that while all three models improve with more estimators, LightGBM and XGBoost maintain stronger, more sustained benefits, whereas Random Forest reaches an optimal threshold more quickly.

4.2 Deep Learning

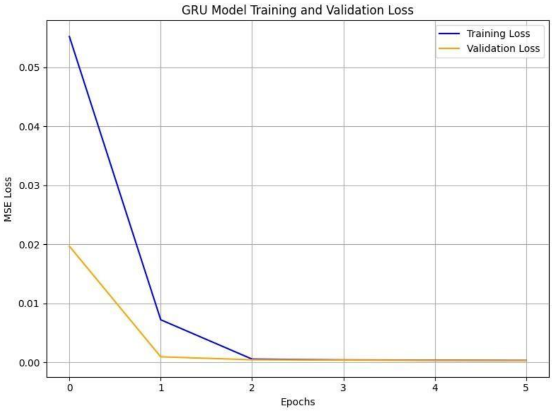
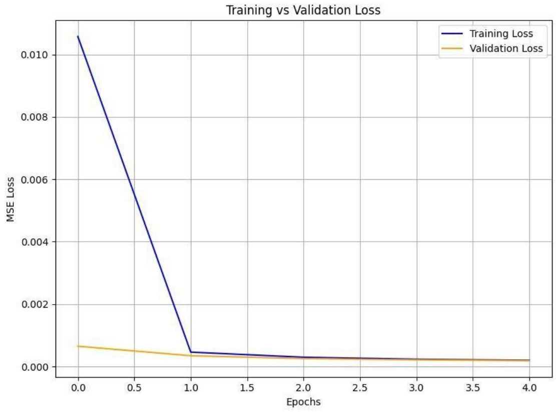
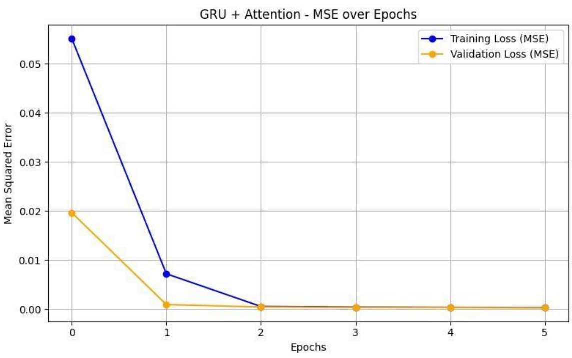


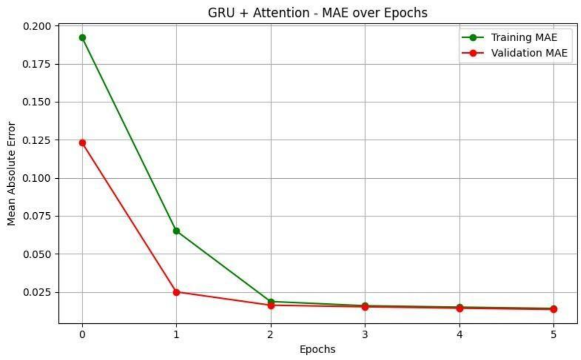
Figure 3A. GRU Validation Loss



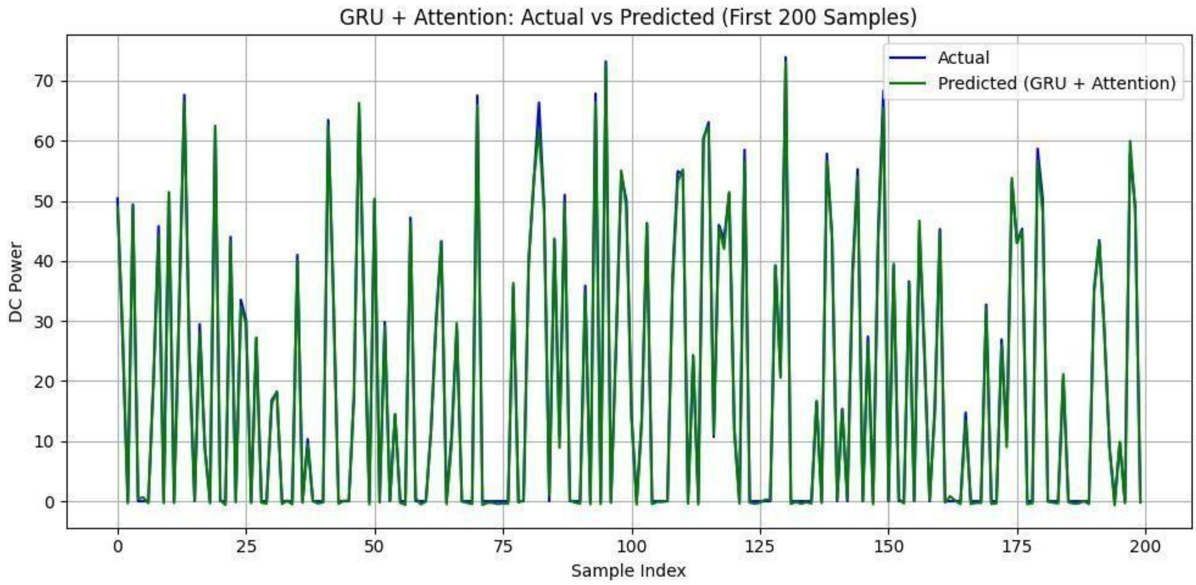
3B. ANN Validation Loss



3C. GRU + Attention MSE Over Epochs



3D. GRU + Attention MAE Over Epochs



3E. GRU + Attention Actual vs Predicted values

Figures 3A to 3D collectively illustrate the training behavior of the baseline GRU and the GRU + Attention models using MSE and MAE as performance metrics. In Figures 3A and 3B, the basic GRU model shows rapid convergence within the first few epochs, with both training and validation losses dropping sharply and stabilizing near zero. The close alignment of these curves suggests strong generalization with minimal overfitting. However, the early plateau in training loss could indicate underfitting, possibly due to the model's simplicity. The GRU + Attention model, shown in Figures 3C and 3D, achieves even better performance, with both MSE and MAE decreasing rapidly and the training and validation curves remaining nearly identical. This consistency indicates that the attention mechanism enhances the GRU's ability to capture important temporal patterns, resulting in faster convergence, improved stability, and more accurate learning of sequential data.

Figure 3E also proves the effectiveness of the GRU + Attention model by comparing actual solar power values and predicted values. The actual values are well followed by the predicted curve throughout the time series and this indicates high accuracy of prediction with slight deviations. This tight correspondence shows that the model does not just learn temporal dynamics effectively during training, but also generalizes effectively to unseen data. Together with the low and stable error rates noted above, this supports the conclusion that the GRU + Attention model is robust, accurate, and generalizable and is therefore suitable to use in realworld short-term solar forecasting applications in different climatic conditions.

6 Conclusion and Future Work

6.1 Conclusion

This study set out to answer the central research question: "***How can a Gated Recurrent Unit (GRU) model integrated with an attention mechanism outperform existing models in terms of accuracy, computational efficiency, and generalizability across multiple regions and diverse climatic conditions for short-term solar power forecasting?"*** The results provide a clear and compelling answer. Among all tested models—both traditional machine learning and deep learning approaches—the GRU + Attention architecture consistently delivered the lowest error metrics, achieving an MAE of 0.53, MSE of 0.53, and RMSE of 0.73. These figures markedly outperform those of baseline GRU (MAE: 1.09), ANN (MAE: 0.86), and even the best-performing ensemble method, Random Forest (MAE: 2.69). This demonstrates that integrating attention not only enhances the GRU's temporal learning but also makes it far more accurate for capturing the nuances of solar energy generation across time and space. The findings also reinforce the model's practical value in addressing the specific challenges outlined in the literature. While transformer-based models offer high accuracy, their resource demands limit real-time and embedded deployment. In contrast, the GRU-attention model strikes a balance between performance and efficiency, achieving fast convergence and generalization across diverse climatic zones such as India, Norway, Brazil, and Australia. The multi-regional validation, supported by exploratory data analysis and consistent results across key metrics, confirms that the proposed architecture is not only lightweight and accurate but also scalable and transferable—meeting the needs of solar forecasting applications in varied global environments.

Looking forward, the GRU + Attention framework offers a strong foundation for future research and real-world deployment. The study opens up avenues for enhancing the model further through techniques like multi-head attention, domain adaptation, and transformer-GRU hybrids. Additionally, exploring edge deployment for microgrids and mobile forecasting platforms could unlock new levels of energy intelligence in remote or infrastructureconstrained areas. Ultimately, this research demonstrates that attention-aware GRU models are a practical, interpretable, and highly effective solution for modern, data-driven energy systems—capable of driving smarter decision-making in the transition toward sustainable power grids.

6.2 Future Work

While the GRU + Attention model demonstrated superior performance in this study, several avenues remain for future enhancement. One promising direction is the integration of multihead attention mechanisms, inspired by transformer architectures, which could enable the model to capture multiple temporal dependencies in parallel. This can enhance performance in very volatile or noisy solar conditions whereby a single attention head might not be sufficient. Also, transfer learning methods can be looked at to generalize the model across geographies with minimal retraining thus not requiring large localized datasets. This would be especially useful in areas where data scarcity is an issue and therefore the model can be used more universally.

Additional studies may also be dedicated to the real-time implementation and edgeoptimization of the GRU-attention model. Running the architecture on low power devices like Raspberry Pi, NVIDIA Jetson, or microcontroller-based systems would enable real-time solar forecasting in off-grid or remote microgrids. Moreover, additional model inputs can be provided by the external satellite or sky images, along with the weather forecasts, in order to improve the accuracy of prediction in case of sudden atmospheric variations. Lastly, future studies can also be done on hybrid models that combine GRU-attention models with explainable AI (XAI) models in order to not only make predictions but also interpretable insights to grid operators to enhance trust and decision-making within energy management systems.

References

Kiasari, M. and Aly, H.H., 2024. Evaluating solar power forecasting robustness: A comparative analysis of XGBoost, RNN, KNN, RF, and LSTM with emphasis on lagged steps, sensitivity, and cross-validation techniques. *In: 2024 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. IEEE, pp.686–692.

Ledmaoui, Y., El Fahli, A., El Maghraoui, A., Hamdouchi, A., El Aroussi, M., Saadane, R. and Chebak, A., 2024. *Enhancing solar power efficiency: Smart metering and ANN-based production forecasting*. Computers, 13(9), p.235.

Lim, S.C., Huh, J.H., Hong, S.H., Park, C.Y. and Kim, J.C., 2022. *Solar power forecasting using CNN-LSTM hybrid model*. Energies, 15(21), p.8233.

Mitrentsis, G. and Lens, H., 2022. *An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting*. Applied Energy, 309, p.118473.

Mohamed, M., Mahmood, F.E., Abd, M.A., Chandra, A. and Singh, B., 2022. *Dynamic forecasting of solar energy microgrid systems using feature engineering*. IEEE Transactions on Industry Applications, 58(6), pp.7857–7867.

Hanif, M.F. and Mi, J., 2024. *Harnessing AI for solar energy: Emergence of transformer models*. Applied Energy, 369, p.123541.

Jailani, N.L.M., Dhanasegaran, J.K., Alkawsi, G., Alkahtani, A.A., Phing, C.C., Baashar, Y., Capretz, L.F., Al-Shetwi, A.Q. and Tiong, S.K., 2023. *Investigating the power of LSTM-based models in solar energy forecasting*. Processes, 11(5), p.1382.

Park, K., Yim, J., Lee, H., Park, M. and Kim, H., 2023. *Real-time solar power estimation through RNN-based attention models*. IEEE Access, 12, pp.62502–62515.

Rai, A., Shrivastava, A. and Jana, K.C., 2023. *Differential attention net: Multi-directed differential attention based hybrid deep learning model for solar power forecasting*. Energy, 263, p.125746.

Rahimi, N., Park, S., Choi, W., Oh, B., Kim, S., Cho, Y.H., Ahn, S., Chong, C., Kim, D., Jin, C. and Lee, D., 2023. *A comprehensive review on ensemble solar power forecasting algorithms*. Journal of Electrical Engineering & Technology, 18(2), pp.719–733.

Rangelov, D., Boerger, M., Tcholtchev, N., Lämmel, P. and Hauswirth, M., 2023. *Design and development of a short-term photovoltaic power output forecasting method based on random* *forest, deep neural network and LSTM using readily available weather features*. IEEE Access, 11, pp.41578–41589.

Saxena, A., Shankar, R., El-Saadany, E.F., Kumar, M., Al Zaabi, O., Al Hosani, K. and Muduli, U.R., 2024. *Intelligent load forecasting and renewable energy integration for enhanced grid reliability*. IEEE Transactions on Industry Applications, 60(6), pp.8403–8415.

Sudharshan, K., Naveen, C., Vishnuram, P., Kasagani, D.V.S.K.R. and Nastasi, B., 2022. *Systematic review on impact of different irradiance forecasting techniques for solar energy prediction*. Energies, 15(17), p.6267.

Wang, Z., Wang, L., Huang, C. and Luo, X., 2023. *A hybrid ensemble learning model for shortterm solar irradiance forecasting using historical observations and sky images*. IEEE Transactions on Industry Applications, 59(2), pp.2041–2052.

Zhang, H., Yang, J., Fan, S., Geng, H. and Shao, C., 2021. *An ultra-short-term distributed photovoltaic power forecasting method based on GPT*. IEEE Transactions on Sustainable Energy. Advance online publication.