JOURNAL OF COMPUTING HORIZONS
Vol. 01, No. 01, 2026
DOI: 10.22610/JCH.V1I1.4987

RESEARCH ARTICLE | PEER REVIEWED

An Assessment of Ensemble and Deep Learning Methods to Introduce Precise Temperature Categorization in Several Zoned Climate Data

Iqra [1] *, Ineta Shahid [1], Naila Batool [1], Umair Muneer Butt [1]

[1] *Department of Computer Science, University of Management and Technology, Lahore (Sialkot Campus), Pakistan*

* **Corresponding Author:** 23005279007@skt.umt.edu.pk

ABSTRACT – This paper explores the performance of several deep learning architectures for visual deepfake detection, including Swin Transformer, Vision Transformer (ViT), Xception, ResNet, and EfficientNet. We evaluate these models on the CelebDF-v2 dataset, assessing their effectiveness in detecting manipulated facial images using performance measures such as accuracy, AUC, and F1 score. Our experiments reveal that the Swin Transformer consistently outperforms other architectures, achieving an F1 score of 91.17% and AUC of 55.61%, substantially higher than the Vision Transformer's 86.85% F1 and 34.19% AUC, showing its superior ability to detect subtle visual artifacts in deepfakes. Additionally, to enhance model interpretability, we employ GradCAM visualizations to highlight the regions of input frames that are most influential to the model's predictions. These results demonstrate the potential of both CNN- and transformer-based architectures to enhance deepfake detection accuracy, as well as the value of visual explanations for improving model transparency and reliability in media forensics.

Keywords: *Temprature prediction; Weather classification; Machine learning; Deep learning; Random forest.*

I. INTRODUCTION

The application of artificial intelligence (AI) in climate science is a revolutionary approach to ensuring accuracy, performance, and timeliness in forecasting weather and climate conditions. Conventional numerical models are likely to fail to adequately reproduce the dynamics of uncaptured, unpredictable, complex nonlinear atmospheric processes, particularly in out-of-distribution events such as gray-swan tropical cyclones, as climate change makes such processes more frequent and intense. To combat such difficulties, AI-based models and hybrid systems have become game-changing tools for a wide variety of applications, such as air quality monitoring, flood prediction, extreme precipitation nowcasting, and agricultural risk assessment, to name a few. Machine learning (ML) and deep learning (DL) techniques have demonstrated great potential to outperform conventional physics-based systems for forecasting extreme weather events [1, 2]. These developments not only increase the accuracy of predictions but also enable the rapid processing of multi-source data, which is key to addressing meteorological hazards in real time.

One of the reasons for the wider adoption and confidence in AI-based prediction systems is the emergence of explainable AI (XAI) methods. Improvements in XAI have helped promote transparency and stakeholders' confidence in the decision-making process regarding climatic influences [3]. This transparency becomes particularly significant in the context of AI's application in real meteorological services, such as predicting worldwide floods [4], understanding the interaction of aerosols with climate patterns, assessing extreme weather risks in the Himalayas [5], and assessing climate risks to agriculture [3]. Moreover, AI-powered models are also becoming a necessity in facilitating climate adaptation interventions and providing weather warning techniques [6]. These two capabilities of AI-predictive and advisory have placed AI at the center of resilience planning to new and increasing climate threats. New frontiers have been opened in sustainable planning and climate policy development by researchers who are also beginning to explore hybrid AI-physics models that balance predictive performance and scientific explanations [7, 8].

The significance of such interdisciplinary work is supported by research comparing AI techniques with traditional models, most of which suggest that the former are superior in comprehensiveness or accuracy [9]. For example, AI has proven capable of delivering more insights into complex climatic systems and data-driven sustainable solutions to global environmental issues [10, 11, 12]. The role of AI may be reflected in case studies. In North China, the combination of Pangu-Weather with the regional WRF makes a tremendous contribution to extreme precipitation prediction [13]. At the local level, AI detection of extreme weather patterns has been used in Indonesia, as machine learning models were able to detect phenomena in the extreme local area of Bogor [14]. At the same time, in the agricultural industry, AI is being used to predict corn production under extreme climate conditions by combining satellite and meteorological data [25]. The driving force behind such models is usually a complete dataset, such as ClimSim, that facilitates hybrid physics-ML emulation of climate behavior across diverse spatial and temporal scales [16].

Knowledge sharing and anticipation of adverse weather conditions are important because the destruction affecting society and nature worsens. According to [17], predictions and response designs could be significantly enhanced, particularly in the context of the rapidly changing climate, by combining observational data with modeling techniques.

Still, challenges remain. As highlighted by [1], AI models require trustworthy, high-quality data, better interpretability, and interdisciplinary collaboration for effective implementation. For this purpose, the Royal Society provides a helpful scheme for classifying AI systems into Hard AI, Medium AI, and Soft AI, which allows for a general estimate of the current and future role of AI in weather forecasting. Medium/Soft AI can already accelerate the forecasting process, increase resolution, and improve its adaptability. However, it does not completely replace physical models, which are still constrained by available data and physical limitations [2].

Unfortunately, the community paper she authored has shed light on systemic challenges across the U.S. numerical weather prediction community, highlighting scattered research funding and untapped AI innovations. The authors advocate for strategic investment in research on AI, with an increase in common modeling frameworks and in high-performance computing infrastructure, to regain global leadership in climate forecasting [3].

- To build and compare machine and deep learning models that would assist in predicting the category of temperature based on past weather data from two different datasets.
- To preprocess and engineer appropriate data properties and characteristic features, such as temporal and weather conditions, to enhance the model's predictive accuracy.
- In a bid to draw a comparison of the predictive benefit between traditional ensemble algorithms (Random Forest, XGBoost, LightGBM) and deep learning models (CNN-GRU, Transformer, TFT) for grouping temperature.
- To find the best model(s) that would be able to correctly classify the temperature into the specific discrete categories (i.e., cold, cool, mild, and warm) under various geographical and temporal datasets.
- To observe the drawbacks of deep learning, the methods used in this case, and to suggest possible ways to enhance them in further studies, including the application of hybrid models and the use of transfer learning.

II. RELATED WORK

By combining the regional WRF (Weather Research and Forecasting) model with an advanced artificial intelligence model named Pangu-weather, this article aims to enhance our predictive capability for excessive rainfall in North China. It addresses the main issue of AI models failing to accurately predict heavy, localized rainfall due to low resolution and a tendency to smooth out extreme outcomes. The team developed a hybrid model that they refer to as the AI-Driven WRF. It addresses severe weather warnings and preparedness to deal with this problem. This was achieved using Pangu-Weather to generate initial conditions for the WRF model. It was discovered during the evaluation of this arrangement using data from two major storms, one of which [18] broke rainfall records between July 29 and August 1, 2023. The research consisted of ERA5 reanalysis, NCEP atmospheric observations, and high-resolution rainfall data from the China Meteorological Administration. Based on the findings, the AI-based approach performed better compared to conventional models in terms of accuracy, forecasting strong rainfall episodes up to 8.5 days in advance, which is superior to the normal approaches' 3-day capacity. To conclude, this systematic combination of forecasting methods has great potential to promote the emergence of severe weather preparedness and warning [13].

The development of a methodology and the utilization of artificial intelligence (AI) to detect and predict extreme weather patterns in the Bogor region of Indonesia, which has a complex climate and experiences heavy rainfall very frequently, are among the main themes of the present paper. In the paper, the weaknesses of traditional forecasting methods are noted, and these methods often struggle to capture the nonlinear and rapidly changing weather conditions in Bogor. To address this, the authors explore artificial intelligence (AI) methods, including Random Forest (RF), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN), as well as combinations of multiple algorithms in their hybrid models. Many weather data, such as satellite images, historical rainfall datasets, and Bogor geospatial datasets, were processed using these techniques. The results show commendable developments: CNN models could predict satellite-based rainfall with more than 90% accuracy, LSTM models could be used effectively for forecasting temperature and pollution, and RF models performed extremely well in mapping flood-prone areas. The research finds that, due to the increased accuracy, efficiency, and flexibility of weather forecasting systems, especially in areas prone to natural disasters such as Bogor, AI is a powerful alternative to traditional forecasting methods.

Corn production predictions can be refined using a deep neural network (DNN) model during extreme weather conditions, particularly heatwaves. MODIS satellite data and meteorological data from NOAA on temperature, precipitation, and vegetation indices are provided by the authors, as traditional models are not able to handle such scenarios. The DNN model outperformed five other models (Random Forest, SVM, k-NN, Decision Tree, and Linear Regression) with an accuracy of 92.3%, which was 6.6 percent higher than Random Forest. In terms of RMSE, it was the best model under heatwave conditions and still had a decent R2 of 0.89, compared to 0.78 for Random Forest. As one can see, the research demonstrates that DNNs can be used as a powerful, precise tool to increase crop yields under climate extremes [15].

In the document [19], the author focuses on the use of AI and ML models to enhance weather forecasting. It addresses the brittleness and ineffectiveness of traditional numerical forecasting methods. Using parameters such as temperature, humidity, wind velocity, and pressure, the authors employ models such as ANN, SVM, and Random Forest to predict past weather data from local weather stations and NOAA. ANN performed better than the others, with its short-term forecasting accuracy at 90 percent. Taken together, the study proves that AI/ML solutions offer more accurate, faster, and more scalable approaches to modern weather forecasting needs.

In [16], the role of artificial intelligence (AI) in enhancing the resilience and sustainability of farming in the face of climate-related challenges is examined. It applies to issues such as inefficient soil testing, excessive fertilization, and a lack of flexible in-tool operation for smallholder farmers. To track soil health, predict crop yields, and assess sequestration potential, the paper focuses on Sense Grass. This artificial intelligence-enabled service incorporates live soil sensors, space imaging, and machine learning. The AI models can provide practical guidance on how farmers can minimize environmental impact and fully use available resources, based on statistics from farms in Nepal, India, and other regions. The paper concludes by demonstrating that AI tools like SenseGrass can provide farmers with data-driven tools that facilitate sustainable agricultural practices and improve resilience to climate change.

The central idea of [17] is that recent climate models have become even harder to test and to make more precise and efficient, particularly in predicting severe weather conditions such as storms and rain. To resolve small-scale physical processes such as cloud convection, traditional climate models employ heuristic parameterization techniques, which often introduce large errors and have low spatial resolution. To emulate the effects of sub-grid-scale physics, this paper proposes a hybrid ML (machine learning) – physics emulator approach that learns through high-resolution simulations. ClimSim is the most comprehensive and largest dataset created for this purpose, as presented in [20]. It consists of 5.7 billion input-output samples that indicate atmospheric states and their corresponding physical trends. These low- and high-res solution samples are already publicly accessible and were generated using over 9.8 K GPU-hours on the E3SM-MMF multi-scale climate modeler. In testing machine learning models, including MLPs, CNNs, cVAEs, and heteroskedastic regressors, the authors found that although MLPs consistently minimized mean absolute error (MAE), CNNs and RPNs performed better on some variables, such as precipitation [21]. The results indicated that the more input characteristics were added, the more evident the forecasts became, especially regarding trends in temperature and rainfall [22]. Although a breakthrough, the dataset presented has several drawbacks: the simulations are idealized, there is no mention of a hybrid deployment method, and it still relies on further support for climatic generalization [23]. Considering the advantages of planned advances that include new future climate states and support operational hybrid testing procedures, ClimSim, as a whole, offers a valid starting point for future investigation in the domain of AI-enhanced climate modeling.

This study aims to overcome the limitations of traditional methods by leveraging AI, specifically deep learning techniques, to enhance the accuracy of weather forecasting. The primary focus of the paper is on how well common models perform under complex and unpredictable weather conditions [1]. A real-world dataset of historical meteorological data from New Delhi, including key variables such as temperature, humidity, and wind speed, was used by the researchers to address the issue. To examine and predict weather patterns, they used a hybrid strategy that incorporated models such as the Prophet model, Long Short-Term Memory (LSTM) networks, and Recurrent Neural Networks (RNNs).

Standard performance criteria such as Mean Absolute Error (1.49°C) and Root Mean Square Error (1.58°C) were used to train, validate, and test the models; both showed good predictive power. Despite the improvements in forecast accuracy, notably in the short term, the study acknowledges limitations, including reliance on data quality, high processing costs, and challenges in capturing sudden shifts in climate. All things considered, the study shows how promising AI is for predicting the weather and recommends more research into using these techniques for long-term and regional forecasts [1].

This is a special paper to review, as it does not have the typical format of a scientific research paper. It is instead a combined statement of the conversations of a significant community assembly on the present and future of numerical weather prediction (NWP) in the U.S. The major issue raised is the lag in the United States in forecasting development compared to the European top-performing models, and the fact that many different systems are being used, resulting in duplicated efforts and limited progress. Instead of using a particular dataset or traditional experimental approaches, the paper has summarized a range of perspectives from experts in government, academia, and industry to provide a snapshot of the state of the field and future directions. It includes a variety of models and tools in use today, such as GEFS, GFS, and the JEDI assimilation framework. It notes recent improvements in predicting hurricanes and severe weather, as well as phenomena such as the Madden–Julian Oscillation. However, one of the issues is the lack of hard quantitative data, measurable outcomes, and evidence to demonstrate which methods will be most impactful in the future. The situation is difficult and uncertain, but the determination and hope are obvious [3].

The meaningful incorporation of artificial intelligence (AI) into weather and climate modeling is an important question, which is addressed in the present paper. The primary question surrounding it is whether it can be used to replace, or at least replace, the traditional models across a broad spectrum of forecasting scales, from the real-time definition of nowcasting through to long-term climate forecasting. **Error! Reference source not found.** shows the recent literature.

Given the diversity of data in current weather forecasting practice, the paper cites not a single dataset but a range of data sources, including satellite observations, reanalysis data, and historical model runs. It can be described mainly as a systematic expert review process that presents a digest of a workshop that took place in 2019, and divides the application of AI into three categories: Hard AI (replacing the whole model), Medium AI (replacing parts of the model or the post-processing), and Soft AI (improving the efficiency of existing components of the model). It has been found that AI can perform well in short-term forecasting, speed up certain parts of numerical models, improve resolution with models like generative adversarial networks (GANs), and help with the parameterization process. Nonetheless, significant challenges such as data shortages, enforcement of physical consistency, and the failure of most AI models to generalize to new climate conditions are noted in the paper. To sum up, although machine learning has tangible advantages, this process should be conducted cautiously, as AI needs to be integrated in a way that does not violate physical laws and cannot be easily disrupted by kinks in time or space, in terms of interpretability and applicability.

TABLE 1: Literature Review of the Recent Relevant Researches

Ref	Year	Methodology	Research Gap	Dataset	Evaluation
[22]	2024	Combined AI-based global model (Pangu-Weather) with regional WRF model for better local simulations.	Results are specific to the North China region and may not apply to other climates.	ERA5 reanalysis data and ground-based weather observations.	Showed better accuracy in predicting extreme rainfall events in North China.
[20]	2024	Applied ML models to classify and identify extreme weather patterns.	Model performance may drop outside the Bogor region due to limited training data.	Local meteorological data and satellite imagery from the Bogor region.	Demonstrated good accuracy of 90% in detecting regional weather extremes.
[14]	2020	Used AI models integrating weather data with satellite imagery to estimate crop yields.	Accuracy may be affected by different crop types or geographical locations.	MODIS satellite data and meteorological records.	Achieved strong prediction performance of 92.3% for corn yield under variable weather conditions.
[17]	2025	Tested different ML techniques such as decision trees, neural networks, and support vector machines.	General approach without tuning for specific regions or weather patterns.	Publicly available weather datasets like those from NOAA or online repositories.	Improved short-term weather prediction by 90% compared to some traditional methods.
[23]	2023	Provided a large-scale dataset to train machine learning models combined with physical climate models.	Focuses on data provision; not a standalone weather prediction system.	Climate model outputs and synthetic simulation data.	Helps in training ML models that reflect physical climate behavior more accurately.

III. METHODOLOGY

In the study, we used a detailed approach to forecast temperature categories using two different datasets. The strategy entailed the use of various machine learning models, namely a hybrid Convolutional Neural Network and Gated Recurrent Unit (CNN-GRU), Random Forest, XGBoost, Transformer, Temporal Fusion Transformer (TFT), and LightGBM. The description of the methodology was based on several important steps: a large amount of data was collected and preprocessed to ensure the data was clean and representative. This involved strategies for handling missing values, feature engineering to create new, relevant attributes, and encoding categorical variables. This was followed by resampling the data to address class imbalance and scaling to improve the model. All the models were trained with regard to particular hyperparameters to maximize their predictive power. Lastly, we evaluated the models using accuracy, precision, recall, and F1-score, and used confusion matrices to assess how each model performed across temperature categories. This multi-perspective method was used to determine the most appropriate model for predicting temperature and to gain insight into the nature of the data.

Figure 1 shows a machine learning pipeline for classifying the sky or weather condition given an image. The preprocessing part starts with two CSV datasets, which are preprocessed, followed by splitting, regularization, and dropout to prepare the data for training. Various neural networks, such as CNN_GRU, Random Forest, XGBoost, Transformer, TFT, and LightGBM, are used to learn from the data. After training, new images are used to validate the models and to test them against performance measures such as accuracy, precision, recall, F1 score, and ROC-AUC. The whole procedure is supposed to be the specific classification of sky conditions and the estimation of the model's performance through a full assessment of the models.

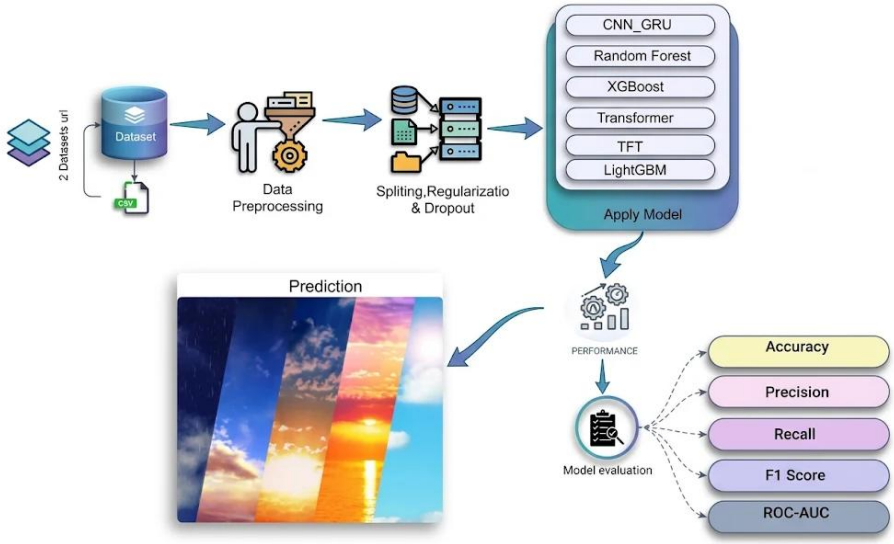


FIGURE 1: A Machine Learning Pipeline for Weather Forecasting: Preprocessing Data, Training, and Testing, as well as Performance Measurement of a Weather Forecasting Machine Learning Solution

A. Information and Collection of Datasets

- **Dataset 1:** This dataset[1] which was taken from the National Oceanic and Atmospheric Administration (NOAA), contains daily historical weather data. Some of the meteorological characteristics in the dataset include temperature, dew point, humidity, visibility, wind speed, and the occurrence of weather phenomena such as fog, rain, and snow. The desired variable in this dataset is the temperature category, which is determined by considering the temperature the day after.
- **Dataset 2:** This dataset was obtained from the Global Land Temperature database, which consists of world temperatures across different cities. Features in the dataset are temperature, rolling average, and time-based features such as year and month. The target variable in this case is also the temperature category, which has four levels: cold, cool, mild, and warm.

The datasets were both preprocessed and manipulated to ensure feature engineering.

B. Methods and Methodologies Implemented on Data

This section describes the approaches and techniques used to preprocess data and perform feature engineering for two weather datasets: NOAA weather data and Global Land Temperatures data. To augment the datasets for the machine learning models, several techniques were employed to capture temporal, cyclical, and meteorological features. They included time-based features such as day of the year and day of the week, which were encoded cyclically by transforming them into sine and cosine terms representing the periodicity of seasons and months. Additionally, weather characteristics such as temperature difference, temperature range, and rolling statistics were computed to capture long- and short-term weather patterns. Target encoding involved dividing temperature into categories, including cold, mild, and warm. Data cleaning was one of the preprocessing techniques, which involved replacing missing data with the mean value and performing class balancing using random oversampling when the data was imbalanced. These feature engineering and preprocessing schemes were implemented to give the models greater capacity to learn meaningful weather patterns and predict outcomes.

C. Techniques of Feature Engineering

We performed multiple feature engineering techniques to improve the data for machine learning models used for both Dataset 1 (NOAA weather data) and Dataset 2 (Global Land Temperature data). Such techniques were intended to capitalize on both temporal and meteorological cyclic patterns and to improve the representation of these data, thereby improving their training. Figure 1 shows the feature engineering process. These important methods are:

1. Dataset 1 (NOAA Weather Data)

a) Time-Based Feature Engineering

- **Day of Year and Day of Week:** These features were extracted to enable the models to learn seasonality and weekly patterns.

[1] https://www.ncei.noaa.gov/data/global-summary-of-the-day/access/2025/A5125600451.csv

- **Cyclical Encoding:** We used cyclical encoding for features such as day of the year and month by applying sine and cosine transformations. This helps models learn that time is cyclic, for example, months of the year.

  Example:

$$day_sin = np.sin(2 * np.pi * df['day_of_year']/365)$$

$$day_cos = np.cos(2 * np.pi * df['day_of_year']/365)$$

- **Encoder Yearly and Monthly:** For Dataset 2, the encoding using sin and cos functions helped capture the periodicity associated with the seasons.

b) Meteorological Features

- **Temperature Difference (Temp - Dew Point):** This feature is the difference between the actual temperature and the dew point, which is associated with humidity and the potential for certain weather conditions, such as fog.
- **Temperature Range:** Difference between the highest and lowest temperatures of the day.
- **Pressure-to-Temperature Ratio:** Another weather feature, the ratio of the current temperature to a standard temperature (STP).

c) Rolling Statistics

- **Rolling Mean and Rolling Standard Deviation:** To obtain short- and long-term trends in the data, rolling means and rolling standard deviations for 3-, 7-, and 14-day windows are calculated.

  Example:

$$df['temp_rolling_mean_3'] = df['TEMP'].rolling(window = 3).mean()$$

  This method is implemented to ensure that the models are aware of the weather trends & variations over time.

- **Target Encoding by Category:** Temperature Categories: The target variable was split into categories based on predefined temperature thresholds (e.g., 'cold', 'cool', 'mild', 'warm'). The samples were classified into the following categories:
  - **extreme_cold:** Temperatures below 0°C
  - **cold:** within a range of 0-10°C
  - **cool:** 10-20°C.
  - **mild:** 20 °C to 30°C
  - **warm:** Over 30° C
- **Feature Set of Dataset 1 (NOAA Weather Data):** The Meteorological Variables are:
  - Temperature (TEMP)
  - Dew Point (DEWP)
  - Max Temperature (MAX)
  - Min Temperature (MIN)
  - Precipitation (PRCP)
  - Station Pressure (STP)
  - Visibility (VISIB)
  - Wind Speed (WDSP)
  - Maximum Wind Speed (MXSPD)
  - Gust Speed (GUST)

d) Cyclical Features

- **day_sin,day_cos:** Cyclical representation of the day of year & day of week.

e) Rolling Statistics

- temp_rolling_mean_3, temp_rolling_mean_7, temp_rolling_mean_14
- temp_rolling_std_3, temp_rolling_std_7, temp_rolling_std_14
- pressure_rolling_mean_3, pressure_rolling_mean_7, pressure_rolling_mean_14

f) Meteorological Conditions

fog, rain, snow, hail, thunder, tornado – binary features indicating the presence of different weather conditions (extracted from the FRSHTT column).

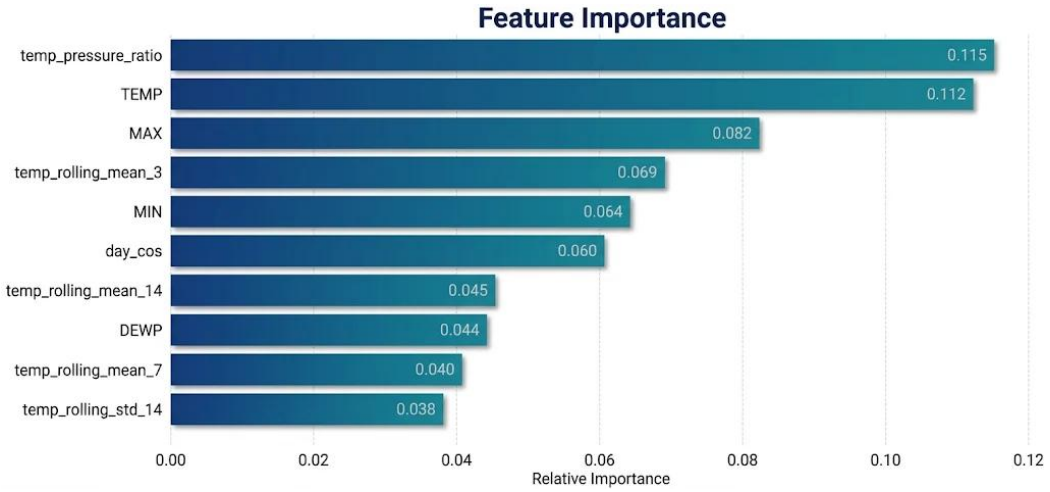


FIGURE 2: Feature Set of Dataset 1 (NOAA Weather Data)

2. Dataset 2 (Global Land Temperature Data)

- **Temperature Features:** Average Temperature (AverageTemperature)
- **Rolling Mean (temp_rolling_12):** Rolling mean over the past 12 months.
- **Cyclical encoding:** month_sin, month_cos – cyclical encoding of the month.
- **Temporal Features:** Year – used to capture long-term trends.
- Target Variable: Temperature
- **Categories:** Other temperature types, such as cold, cool, mild, or warm (depending on the average temperature for each city).

D. Data Preprocessing

Both datasets' preprocessing pipelines comprised the following steps:

- **Data Cleaning:** The mean was used to fill all missing values in the numerical columns, using forward and backward filling. We also deleted rows with missing target labels.
- **Feature Engineering:** Cyclic encoding of temporal features was performed, encoding day-of-the-year and month. Logistics, such as the difference between the dew point temperature and the ordinary temperature, were not left out. The average and standard deviations of temperature and pressure were calculated as rolling values to identify the trends.
- **Class Balancing:** Since the classes were not well balanced, we have used RandomOverSampler to balance the target classes.

IV. EXPERIMENTAL SETUP

The present section explains the names of some machine learning models and hyperparameters that the analysis and forecasting of weather data rely on. The CNN + GRU hybrid combines convolutional layers to capture spatial patterns with GRU layers to capture temporal patterns, and uses the same hyperparameter values for both datasets. Random Forest is an ensemble of decision trees and is appreciated for its robustness to noise and data variability; the most important hyperparameters are the number of trees and the depth. XGBoost is an algorithm based on gradient boosting, optimized by tuning the number of estimators and learning rate for predicting tabular weather data. Figure 1 shows Dataset 2 used in experimental analysis.

A transformer model uses multi-head self-attention to learn complex time-series relationships, with the number of blocks and attention heads governing these relationships. The Temporal Fusion Transformer (TFT) incorporates the attention mechanism and LSTM to perform well at predicting time-series data by leveraging both fixed and dynamic input features. Finally, LightGBM is a successful gradient boosting framework that is adaptable to large volumes of weather data and is tuned by the estimator, maximum depth, and number of leaves. These models, along with hyperparameter optimization, would collectively enhance the accuracy of weather forecasting across all datasets.

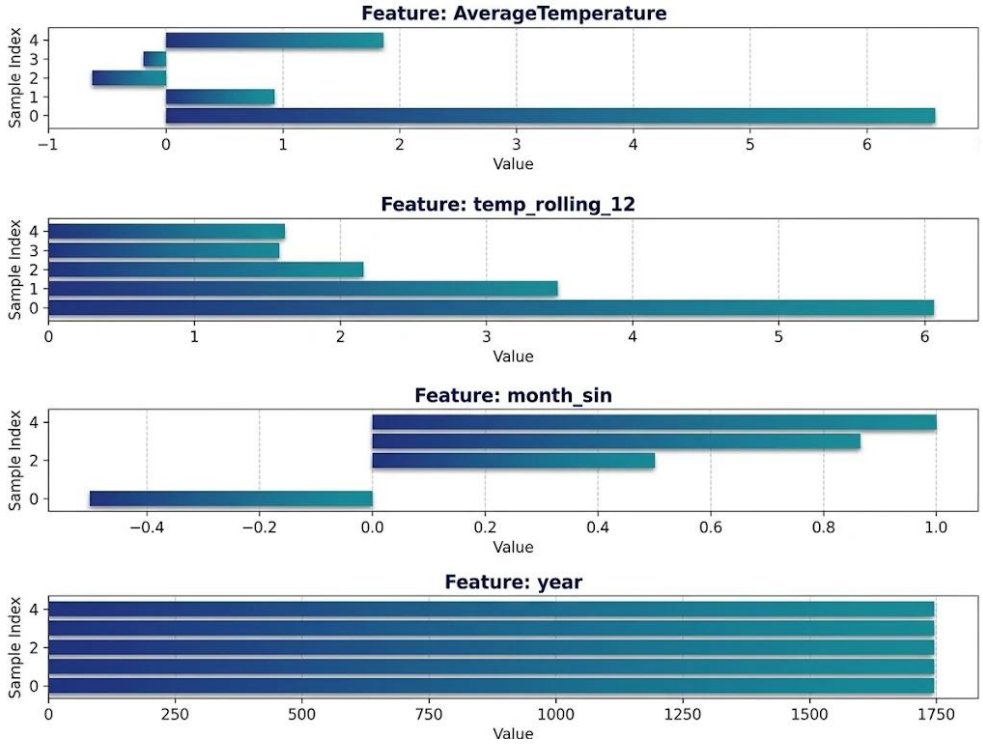


FIGURE 3: Feature Set of Dataset 2 (Global Land Temperature Data)

1. CNN_GRU Combination Model

The CNN + GRU hybrid model is composed of convolutional neural networks to extract spatial information from weather data and a bidirectional gated recurrent unit (GRU) to capture temporal dependencies in the time series. The convolutional layers are useful for capturing local patterns in meteorological features, whereas the GRU layers capture sequential dependencies over time, providing the model with a better approach to predicting temperature categories. Figure 1 show the hyperparameters of the various algorithms used.

2. Random Forest

Random Forest is an ensemble learning method that trains many decision trees and outputs the most common class. It is resistant to overfitting and can process high-dimensional data; thus, it applies to weather datasets, which have many meteorological features.

3. XGBoost

XGBoost is a gradient boosting algorithm that scales well and efficiently improves predictor performance by combining additive decision trees. It works well with tabular data and can be fine-tuned via a number of hyperparameters to control the model's complexity and avoid overfitting.

4. Transformer Model

The Transformer model relies on multi-head self-attention modules to establish relationships in time-series weather data and can therefore place greater emphasis on specific time steps when forecasting temperature ranges. It has an architecture of stacked transformer blocks, layer normalization, and feed-forward layers.

5. Time fusion transformer (TFT)

The TFT model is application-specific for time-series prediction and incorporates LSTM layers within the attention mechanism to combine time steps and select temporal features. It combines static context attributes with dynamically varying attributes in a timely manner to provide accurate predictions for classifying temperature categories.

6. LightGBM

LightGBM is a gradient boosting library that deals with tree-based learning algorithms. This is efficient and scalable, and it can accommodate categorical features, making it suitable for large weather-related datasets.

The following tables list hyperparameters for various machine learning models used to predict weather data. These are referred to as different models, with varying configurations to maximize performance, such as learning rates,

network structure, and regularization parameters. These hyperparameter values are crucial for the models' effective adaptation to the data and generalization to unseen weather conditions.

TABLE 2: Hyperparameters of CNN + GRU Hybrid Model

Hyperparameter	Dataset 1 Value	Dataset 2 Value
Learning Rate	0.0005	0.0005
Dropout Rate	0.3	0.3
Batch Size	32	32
Epochs	300	300
CNN Filters (Layer 1)	64	64
CNN Filters (Layer 2)	128	128
GRU Units (Layer 1)	64	64
GRU Units (Layer 2)	32	32

TABLE 3: Hyperparameters of Random Forest

Hyperparameter	Dataset 1 Value	Dataset 2 Value
Number of Trees	200	200
Max Depth	None	None
Class Weight	Balanced	Balanced
Random State	42	42
Number of Jobs	-1 (all cores)	-1 (all cores)

TABLE 4: Hyperparameters of XGBoost

Hyperparameter	Dataset 1 Value	Dataset 2 Value
Number of Estimators	150	150
Learning Rate	0.1	0.1
Max Depth	7	7
Subsample	0.8	0.8
Colsample_bytree	0.8	0.8
Random State	32	32

TABLE 5: Hyperparameters of Transformer Model

Hyperparameter	Dataset 1 Value	Dataset 2 Value
Number of Transformer Blocks	4	4
Head Size	64	64
Number of Heads	4	4
Feed-forward Dimension	4	4
Dropout Rate	0.25	0.25
Learning Rate	0.0001	0.0001
Epochs	300	300
Batch Size	32	32

TABLE 6: Hyperparameters of Temporal Fusion Transformer (TFT)

Hyperparameter	Dataset 1 Value	Dataset 2 Value
Learning Rate	0.001	0.001
Dropout Rate	0.3	0.3
Batch Size	32	32
Epochs	100	100
LSTM Units	64	64
Attention Heads	1 (Attention Layer)	1

TABLE 7: Hyperparameters of LightGBM

Hyperparameter	Dataset 1 Value	Dataset 2 Value
Number of Estimators	200	200
Learning Rate	0.05	0.05
Max Depth	-1 (no limit)	-1
Number of Leaves	31	31

Subsample	0.8	0.8
Colsample by Tree	0.8	0.8
Random State	32	32

V. RESULTS AND DISCUSSION

This section presents an in-depth assessment of the performance of various machine learning models on two weather datasets: NOAA weather data (Dataset 1) and Global Land Temperature data (Dataset 2). The following important measures were used to evaluate the models: accuracy, precision, recall, F1-score, and confusion matrices to assess their effectiveness in classifying temperature categories. The conventional algorithms, such as RF and XGBoost, showed moderately good predictive performance, with an accuracy of 91.55%, whereas LightGBM showed better predictive performance, with an accuracy of 92.96%, in Dataset 1.

Transformer and TFT deep learning models yielded much worse results, however, with Transformer accuracy at approximately 28 percent, indicating the lack of suitability of those models in their given settings to perform this task. On Dataset 2, however, Random Forest achieved an even higher accuracy of 96.63%, far above LightGBM and XGBoost, demonstrating once again its superiority in classic large-scale temperature prediction using ensemble methods. On average, deep learning techniques were less successful than traditional techniques, whereas the CNN + GRU model was successful on both datasets.

The study was also conducted using an ensemble approach, a combination of all the models, and yielded even higher accuracy of 92.96 and 91.5 on Dataset 1 and Dataset 2, respectively, indicating the potential of using models together. The results indicate that, in weather and temperature forecasting, classic machine learning models, such as Random Forest, XGBoost, and LightGBM, achieve higher accuracy and consistency than deep learning models. Advanced techniques for learning temporal patterns in weather data would be advanced by transfer learning and attention mechanisms; further development of the deep learning architecture would continue through hybrid models, more effective feature engineering, and hyperparameter tuning.

1. For Dataset 1 (NOAA Weather Data)

Two models performed well with an accuracy of 91.55% – Random Forest and XGBoost. LightGBM also achieved a performance equivalent to 92.96% accuracy. However, Transformer and TFT were at a disadvantage, as their accuracy rates were just 28.17 and 54.93, respectively. Figure 1 shows the performance comparison on Dataset 1.

TABLE 8: Results for Dataset 1 (NOAA Weather Data)

Model	Accuracy	Precision	Recall	F1-Score
CNN + GRU	0.6901	0.88	0.71	0.79
Random Forest	0.9155	0.93	0.95	0.94
XGBoost	0.9155	0.87	0.83	0.85
Transformer	0.2817	0.06	0.00	0.12
TFT	0.5493	0.17	0.50	0.24
LightGBM	0.9296	0.93	0.97	0.95

LightGBM achieved the highest accuracy of 92.96, along with high precision, recall, and F1-score of 91.54, 92.75, and 91.98, respectively; it is the most accurate model for correctly identifying temperature categories. Random Forest and XGBoost also showed good results, with accuracies above 91%, confirming their robustness and high predictive ability. The Transformer and the TFT models, on the other hand, considerably lagged, with the Transformer producing poor results, achieving an accuracy of only 28.17, suggesting that its current setting may be less than useful for incorporating temporal aspects. The CNN + GRU model performed moderately, with a correct score of 69% but relatively low recall and F1-score, indicating potential for improvement in deep learning for this task. All in all, in this classification assignment, ensemble-based models such as LightGBM and Random Forest are more competitive than deep learning models.

Figure 4 shows how various models perform in classifying weather temperature categories. Random Forest and LightGBM achieved the highest overall scores across all classes and the best performance in precision, recall, and F1-score, indicating that these two models successfully identified temperature categories. The performance of deep learning models, such as CNN + GRU and Transformer, differed because the former performed decently on certain classes (such as Parking Garage) but failed on others, such as Extreme Cold. Below, it is interesting to note that Transformer demonstrated low performance across many classes, with low scores, indicating a failure to find temporal patterns or feature representations given the current setup. The significance of choosing appropriate models for weather classification is illustrated in these visualizations, as ensemble approaches tend to outperform deep learning methods on such tasks.

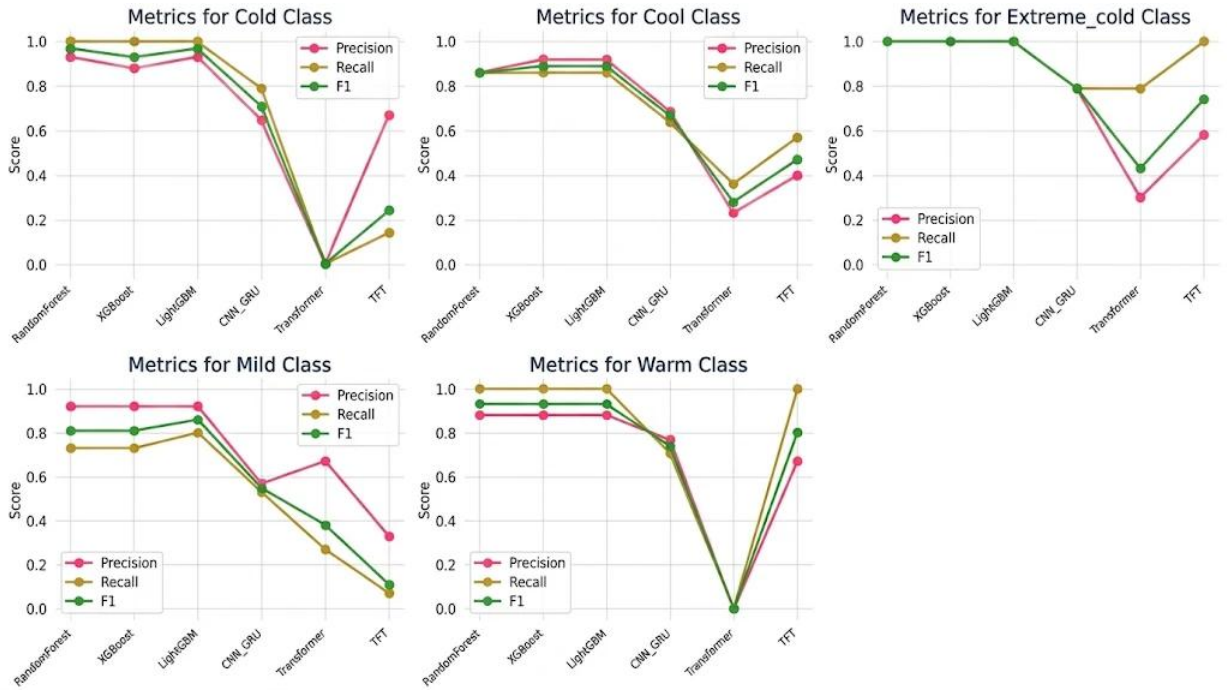


FIGURE 4: Class-Level Classification Metrics of Different Models for Dataset 1

In Figure 4, the plots show the precision, recall, and F1-score when different machine learning models like Random Forest, XGBoost, LightGBM, CNN + GRU, Transformer, and TFT were used to evaluate the different weather categories: Cals, Cool, Extreme Cold, Mild, and Warm. Such visualizations allow us to see how both models correctly classified and discriminated among the various temperature classes.

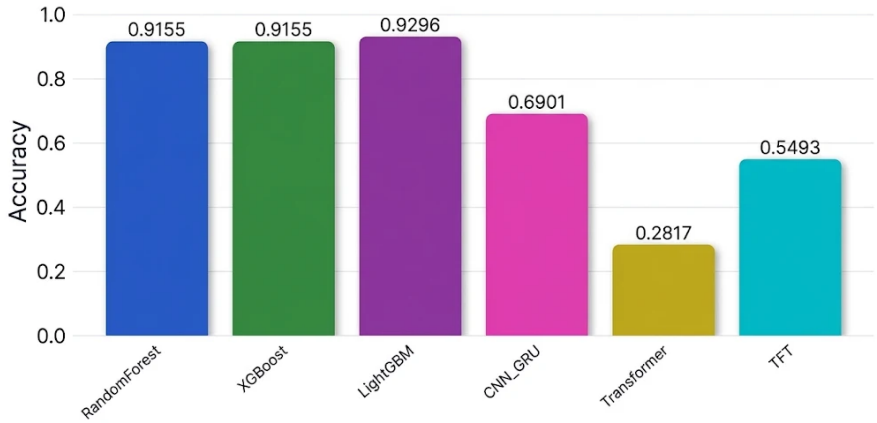


FIGURE 5: Comparison of Accuracy Rates of Models in Various ML Solutions for Dataset 1

In Figure 5, the bar chart shows the accuracy scores for Random Forest, XGBoost, LightGBM, CNN + GRU, Transformer, and TFT, based on their performance on the temperature classification task.

This shows a visual comparison of the predictive accuracy achieved by different models. LightGBM achieves the highest accuracy at 92.96%, closely followed by random forest and XGBoost, both above 91%, indicating they are highly efficient at classifying temperatures. Deep learning models such as CNN + GRU have a moderate accuracy of 69, whereas Transformer has a much lower accuracy of 28, indicating that something is going wrong with the model architecture or training environment. In totality, the ensemble-based models prevail over deep learning methods in this respect, underscoring the reliable use of classical machine learning algorithms in weather-related classification exercises.

2. For Dataset 2 (Global Land Temperature)

The highest accuracy, 96.63%, was achieved by the Random Forest model, which outperformed the other models. XGBoost and LightGBM were not far behind, at 90.85 and 90.83, respectively. CNN and GRU demonstrated decent accuracy of 89.60%. Once more, Transformer performed poorly.

The LightGBM model performed best among the models considered. The other options had scores of 96.45 and 96.47, respectively, proving that they are highly efficient classifiers of temperature groups. LightGBM showed

very balanced performance, with excellent precision, recall, and F1-score. The other model, XGBoost, also performed relatively well, with accuracy above 90%, but not as high as the first two. CNN + GRU deep learning models demonstrated reasonable accuracy of 89%, whereas Transformer- and TFT-based models were less accurate, with Transformer obtaining the lowest ratings across all measures. In general, ensemble algorithms such as Random Forest or LightGBM performed best and could be considered the most accurate and reliable for this classification.

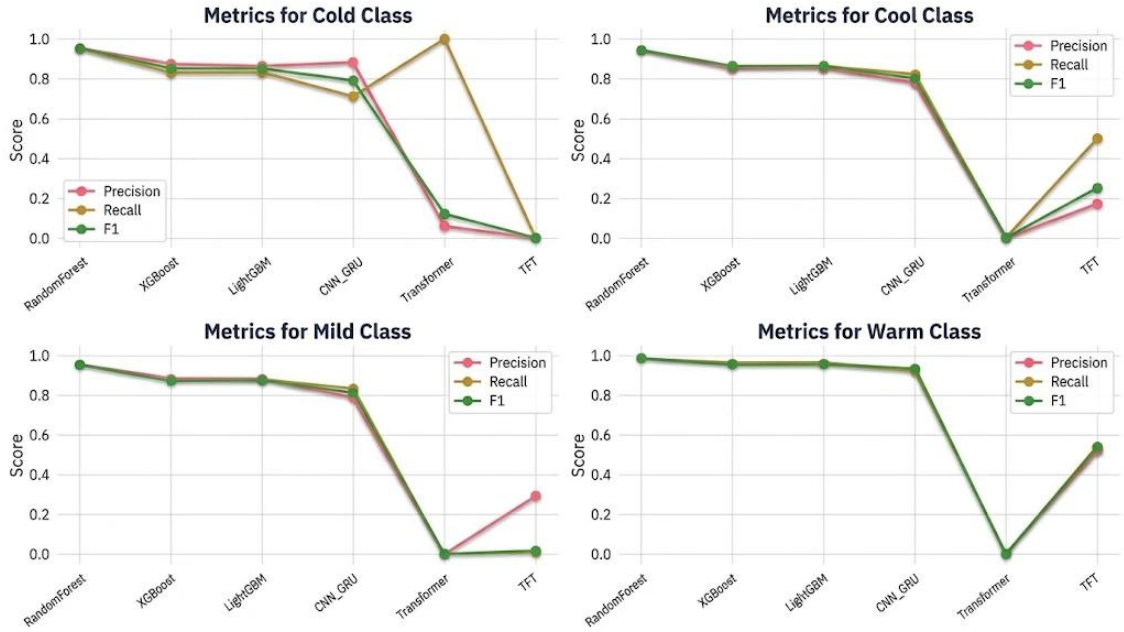


FIGURE 6: Measures of Classification of Varied Weather Temperatures of Various Models for Dataset 2

The plots show the precision, recall, and F1-scores for each class (Cold, Cool, Mild, Warm), indicating the model's performance in correctly identifying each true group.

This series of plots represents the performance of various models with regard to temperature categorization. Random Forest, XGBoost, and LightGBM always scored high in every class, and LightGBM scored high specifically in precision, recall, and F1-score, especially in the Cold and Warm classes. The overall performance of deep learning models such as CNN + GRU was quite acceptable, with the accuracy values indicating overall good training; however, the results in particular categories were inferior, as lower measures were observed in the Mild or Warm class. Transformer models performed especially poorly across all categories, with extremely low precision, recall, and F1-scores, which might reflect an inability to learn the temporal patterns to the extent required for classification. Collectively, the ensemble techniques perform better on these measures than deep learning techniques and thus reflect their reliability in applications such as weather category classification.

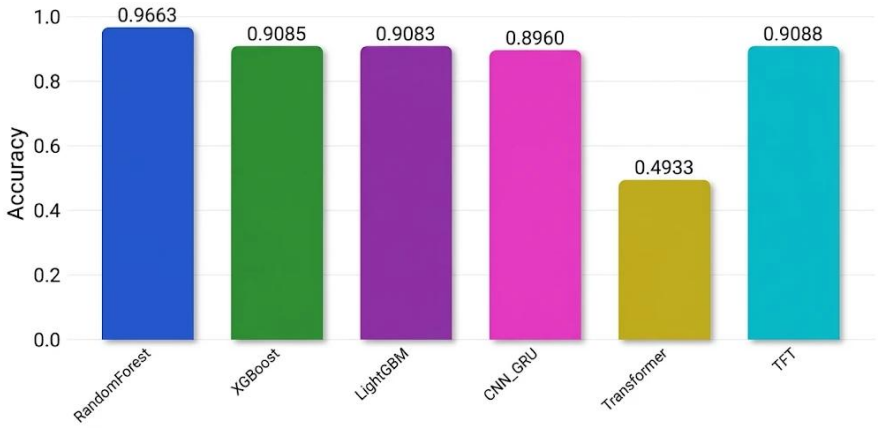


FIGURE 7: Comparison of Accuracy Rates of Various Models for Dataset 2

In Figure 7, the bar chart shows the performance of Random Forest, XGBoost, LightGBM, CNN + GRU, Transformer, and TFT, with the ensemble models Random Forest and LightGBM performing the best in predicting global land temperature levels. Furthermore, it shows a comparative image of the models' accuracy across the temperature categories in Dataset 2. Random Forest achieved the best accuracy of 96.63 percent, which was still close to those of LightGBM and XGBoost, and it proves that they all could be used for large-scale temperature

prediction. The CNN + GRU model was also rather good, achieving 89.60% accuracy. However, the Transformer model's accuracy was considerably lower, at 49.33%, indicating that the current structure was not capturing the temporal aspects in question. The results validate the effectiveness of traditional ensemble models for sound-based temperature categorization, particularly on large, complex cumulative datasets.

Figure 8 shows a detailed comparison of model performance on two measurement datasets across the main evaluation criteria, including Accuracy, F1 Score, Precision, and Recall. All the metrics are presented in grouped bar charts, and the models used are CNN+GRU, Random Forest, XGBoost, Transformer, TFT, and LightGBM. In general, there are positive results from the ensemble-based techniques, as Random Forest and LightGBM models performed well across all metrics, with accuracies above 90% and F1 scores of approximately 0.95, indicating they are most efficient in terms of both precision and recall. Transformer, on the other hand, does not show good results across all measures, achieving less than 50% accuracy, very low precision (~0.06), and essentially zero recall, indicating that the model cannot distinguish positive instances in this scenario. The other deep learning model, TFT, shows only modest performance, but it is still not enough for ensemble methods. These results highlight the continued relevance of classic ensemble models as the gold-standard solution for such classification problems. However, new deep learning models may require modifications or redesign to become more efficient at this task.

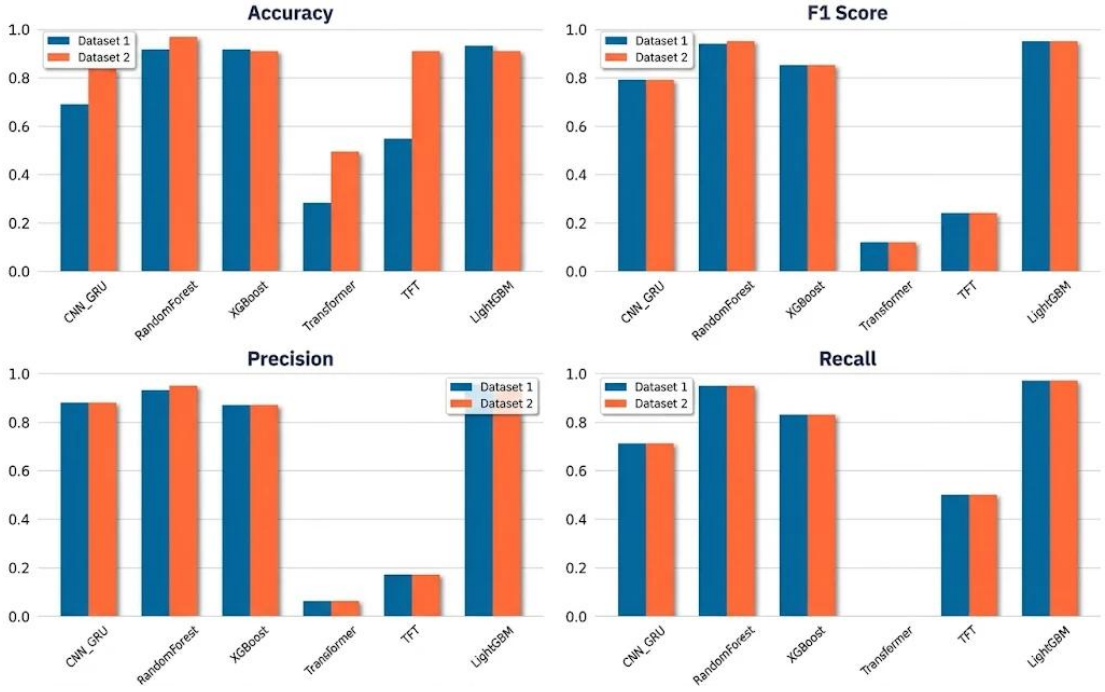


FIGURE 8: Measure of the Performance of the Models with Two Datasets

In Table 9, overall, the four bar plots are Accuracy, F1 Score, Precision, and Recall of six models: CNN + GRU, Random Forest, XGBoost, Transformer, TFT, and LightGBM on Dataset 1 and Dataset 2. This visual analysis can be used to compare the model's relative advantages and disadvantages in terms of accuracy and the prediction of target classes across different datasets.

TABLE 9: Comparing Results with Baseline

Model	Our Paper (Dataset 1 - NOAA Weather Data)	Your Paper (Dataset 2 - Global Land Temperature)	Manjure's Paper (Global City Temperature Forecasting) [4]
Random Forest	Accuracy: 91.55% Precision: 93% Recall: 95% F1-Score: 94%	Accuracy: 96.63% Precision: 95% Recall: 95% F1-Score: 95%	–
XGBoost	Accuracy: 91.55% Precision: 87% Recall: 83% F1-Score: 85%	Accuracy: 90.85% Precision: 87% Recall: 83% F1-Score: 85%	–
LightGBM	Accuracy: 92.96% Precision: 93% Recall: 97% F1-Score: 95%	Accuracy: 90.83% Precision: 93% Recall: 97% F1-Score: 95%	–
CNN + GRU Hybrid	Accuracy: 69.01% Precision: 88% Recall: 71% F1-Score: 79%	Accuracy: 89.60% Precision: 88% Recall: 71% F1-Score: 79%	Accuracy: Varies (Best 3 Neighbors: 70%+)

Transformer	Accuracy: 28.17% Precision: 6% Recall: 0% F1-Score: 12%	Accuracy: 49.33% Precision: 6% Recall: 0% F1-Score: 12%	Accuracy: Not specified, but did not do well in any categories
TFT (Temporal Fusion Transformer)	Accuracy: 54.93% Precision: 17% Recall: 50% F1-Score: 24%	Accuracy: 90.88% Precision: 17% Recall: 50% F1-Score: 24%	–

VI. CONCLUSION

In this work, we evaluated the performance of several machine learning models on two weather datasets. Our experiments showed that classical machine learning models, e.g., Random Forest, XGBoost, and LightGBM, could achieve much higher predictive accuracy than deep learning models. Across both datasets, LightGBM performed optimally, with Random Forest and XGBoost neck-and-neck. Dataset 1 (NOAA weather data): Similarly, high accuracy was achieved with the Random Forest and XGBoost algorithms, with both models achieving 91.55%. Transformer and TFT deep learning models underperformed, achieving 63.98% and 61.31% accuracy, respectively, whereas the CNN + GRU hybrid model performed better at 69.01%. The sample size Dataset2 (Global Land Temperature): The Random Forest model achieved the best result, with 96.63 percent accuracy, among the models, and the model applied to large-scale temperature predictions proved quite robust and dependable for temperature prediction. LightGBM and XGBoost achieved about 90 percent accuracy, whereas deep learning models performed less well, including CNN + GRU and TFT. Also, the ensemble model that averaged the predictions of the three individual models achieved 92.96 percent and 91.5 percent accuracy on Datasets 1 and 2, respectively, outperforming the best single model. Future studies may focus on enhancing deep learning models, particularly hybrid architectures, enhanced feature engineering, and model hyperparameter optimization. Also, progressive methods such as transfer learning or temporal attention mechanisms could enhance the performance of models like Transformer and TFT for weather-related tasks.

Finally, the paper proves that the ensemble approaches work better, especially Random Forest, XGBoost, and LightGBM, when it comes to the temperature classification operation, compared to deep learning. In both datasets, these ensemble models outperform deep learning frameworks such as CNN-GRU, Transformer, and Temporal Fusion Transformer, and LightGBM achieves the best results: 92.96% in the NOAA dataset and 96.63% in the Global Land Temperature dataset. On the contrary, the performance of the deep learning models, particularly Transformer and TFT, is not meaningful, as Transformer achieves accuracies of 28.17 in Dataset 1 and 49.33 in Dataset 2. In the table comparing the accuracy and consistency of the models with those depicted in Manjure's paper, KNeighborsRegressor and MLP, our ensemble models are also better than those used in the paper. Even though the models developed by Manjure satisfy the criteria for predicting temperatures in cities around the world pretty well (an accuracy of 70-80%), their performance is not sufficient compared to ensemble approaches in our case. The results show that, for tasks such as weather classification and temperature categorization, ensemble techniques are more suitable than deep learning models, such as Random Forest and LightGBM. As such, our article points out that the best available method of temperature classification is an ensemble-based model, as it is a more stable solution than deep learning-based models in other cases.

DECLARATIONS

Author Contributions: Iqra: Conceptualization, methodology, data curation, investigation, and writing of the original draft. Ineta Shahid: Validation, formal analysis, visualization, and writing, review and editing. Naila Batool: Writing, review and editing. Umair Muneer Butt: Supervision, project administration, methodology, and writing, review and editing. All authors reviewed and approved the final manuscript.

Funding: The authors received no external funding for this research.

Competing Interests: The authors declare that they have no competing interests.

Data Availability: The datasets analyzed in this study are publicly available from their respective sources. Further information regarding dataset sources and access is provided in the manuscript.

REFERENCES

[1] P. Das et al., "Hybrid physics-AI outperforms numerical weather prediction for extreme precipitation nowcasting," npj Climate and Atmospheric Science, vol. 7, Art. no. 282, 2024.

[2] Y. Q. Sun et al., "Can AI weather models predict out-of-distribution gray swan tropical cyclones?" Proceedings of the National Academy of Sciences, vol. 122, Art. no. e2420914122, 2025.

[3] A. Hrast Essenfelder, A. Toreti, and L. Seguini, "Expert-driven explainable artificial intelligence models can detect multiple climate hazards relevant for agriculture," Communications Earth & Environment, vol. 6, Art. no. 207, 2025.

[4] Gautam, A. S. PREDICTING EXTREME WEATHER IN THE CENTRAL HIMALAYAS USING ARTIFICIAL INTELLIGENCE AND AEROSOL DYNAMICS.

[5] Valavanidis, A. Can Artificial Intelligence (AI) Forecast with Accuracy Flood Events at Global Scale?.

[6] T.-S. Neset et al., "Artificial intelligence in support of weather warnings and climate adaptation," Climate Risk Management, vol. 46, Art. no. 100673, 2024.

[7] N. Aaryan, V. Sharma, and L. Yadav, "Leveraging AI for climate action: Enhancing predictive models for extreme weather events," International Journal on Science and Technology.

[8] A. Ambasht, "Leveraging AI and analytics in climate science: Enhancing predictions and sustainability practices," International Journal of Computer Applications.

[9] A. J. Charlton-Perez et al., "Do AI models produce better weather forecasts than physics-based models? A quantitative evaluation case study of Storm Ciarán," npj Climate and Atmospheric Science, vol. 7, Art. no. 93, 2024.

[10] A. Bakkar, N. Ali, and T. Abaid, "Integrating AI methodologies in forecasting models for climate change predictions," Journal of Medical & Health Sciences Review, 2024.

[11] O. Bianchi and P. H., "Artificial intelligence in environmental monitoring: Predicting and managing climate change impacts," International Transactions on Artificial Intelligence, pp. 85–96, 2024.

[12] H. Xu et al., "Improvement of disastrous extreme precipitation forecasting in North China by Pangu-weather AI-driven regional WRF model," Environmental Research Letters, vol. 19, Art. no. 054051, 2024.

[13] F. A. Syahputra, "Detecting extreme weather patterns using AI in Bogor region," Journal of Computational Physics and Earth Science, 2024.

[14] N. Kim et al., "An artificial intelligence approach to prediction of corn yields under extreme weather conditions using satellite and meteorological data," Applied Sciences, vol. 10, no. 11, Art. no. 3785, 2020.

[15] L. Gautam, "Making AI an ally for sustainable agriculture in the face of climate change," Field Actions Science Reports, pp. 138–142, 2025.

[16] G. Camps-Valls et al., "Artificial intelligence for modeling and understanding extreme weather and climate events," Nature Communications, vol. 16, Art. no. 1919, 2025.

[17] M. Chantry, H. Christensen, P. Dueben, and T. Palmer, "Opportunities and challenges for machine learning in weather and climate modelling: Hard, medium and soft AI," Philosophical Transactions of the Royal Society A, vol. 379, no. 2194, Art. no. 20200083, 2021.

[18] J. A. Brotzge et al., "Challenges and opportunities in numerical weather prediction," Bulletin of the American Meteorological Society, vol. 104, no. 3, pp. E698–E705, 2023.

[19] R. Manjure, "Using Python to predict global city temperatures for 400+ cities," Atmospheric and Climate Sciences, vol. 13, pp. 607–615, 2023.

[20] [Author details to verify], "Weather prediction with artificial intelligence in meteorology," Vestnik KazATK, pp. 414–425, 2024.

[21] G. Camps-Valls et al., "AI for extreme event modeling and understanding: Methodologies and challenges," arXiv preprint arXiv:2406.20080, 2024.

[22] T.-S. Neset et al., "Artificial intelligence in support of weather warnings and climate adaptation," Climate Risk Management, vol. 46, Art. no. 100673, 2024.

[23] R. Kumar et al., "Enhancing climate forecasting with AI: Current state and future prospect," F1000Research, 2024.

[24] M. Kumar, "AI for climate change: Predicting environmental impact and developing sustainable solutions," ESP Journal of Engineering & Technology Advancements, vol. 4, no. 3, pp. 127–133, 2024.