Journal of Theoretical and Applied Information Technology
30th April 2024. Vol.102. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

NOVEL TECHNIQUE FOR PREDICTION OF WEATHER FORECASTING USING MACHINE LEARNING

*[1]KATTA TRINADHA RAVI KUMAR, [2]DR P SURESH VARMA ,[3]DR M V RAMA SUNDARI

[1]Research Scholar,Dept of CSE, Adikavi Nannaya University,Rajamahendravaram,&Assoc Professor,Dept of Computer Science,SVKP&DrKSRaju Arts & Science College(A),Penugonda-534320
[2]Professor, Department of CSE, Adikavi Nannaya University, Rajamahendravaram, India
[3]Professor, Dept of AIML, Gokaraju Ranga Raju Institute of Engineering and Technology, Hyderabad,

Email: trinadhaknu9@gmail.com sureshvarmap@gmail.com ,mvramasundari@gmail.com

ABSTRACT

Many primary sector operations, including farming, rely on the weather to be productive. Weather forecasting has a significant impact on both life and productivity. The need of accurately predicting the serious repercussions of climate change has increased. Weather forecasts are produced by analyzing vast amounts of data that are sent from satellites for certain uses. Analysis of such a large amount of data takes time. The forecast of meteorological conditions, such as rain, wind, heat, humidity, etc., is possible with this innovative approach. It is helpful in agriculture as well. Therefore, with this new method (KNN+ RF), the occurrence, forecast time, and accuracy of sandstorms are compared with Decision Trees. As a consequence, our model outperforms the current approach in terms of results.
.

Keywords: *Atmospheric Condition, Decision Tree (DT), K-Nearest Neighbours (KNN), Random Forest (RF)*

1. INTRODUCTION

Predicting the weather and climate has been crucial throughout human history. From individual decision-making to large-scale industrial planning, weather forecasting is a vital instrument that supports many aspects of human existence and societal processes. Its ability to direct personal safety measures—such as avoiding risky outdoor activities during bad weather or adopting health precautions in extremely hot or cold temperatures—demonstrates its importance on an individual basis. Forecasts are used to guide planting, harvesting, and irrigation schedules in the agricultural sector, which ultimately helps to maximize crop yields and maintain stable food supply chains [1].

The contagious effects of precise forecasting .This efficiency are echoed in the transportation industry, where the planning and scheduling of flights, train routes, and maritime activities hinge on weather conditions. Accurate weather forecasts are essential for reducing delays and improving safety procedures [2]. Beyond these industries, weather forecasting is crucial to the building and infrastructure development sectors. Since unfavorable circumstances can lead to project delays and quality degradation, precise forecasting is essential to efficient project management. In addition, the ability to predict severe weather phenomena such as hurricanes and typhoons is crucial for disaster relief efforts since it provides early alerts, potentially reducing casualties and property damage [3]
Climate prediction is closely related to life on Earth, even though humans tend to overlook it in the near term. Sea level rise brought on by global warming poses serious problems with far-reaching effects for the planet's future[4]. By utilising advanced climate modeling and forecasting methodologies, we may acquire significant understanding of the possible consequences of these occurrences, which will facilitate the creation of focused mitigation plans. For example, accurate projections of sea level rise in coming decades might guide sensible urban design and catastrophe mitigation strategies in coastal towns. Over a long period of time, climate change is expected to cause significant changes in the geographic range of many species, endangering biodiversity. Modern climate models incorporate a variety of factors—such as atmospheric pressure, ocean currents, land ecosystems, and biosphere interactions—to provide a detailed understanding of environmental changes [5]. The development of successful national, international, and local policies

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

targeted at protecting ecological variety requires an integrated approach. Tourism, fishing, and agriculture are three industries that are particularly vulnerable to the unpredictable effects of climate change. Increased temperatures might cause agriculture yields to fall, and a rise in extreme weather events could have a negative effect on tourism. The use of longitudinal climate projections to inform commercial and governmental adaptation plans to these unavoidable changes is crucial. Furthermore, long-term climate forecasts are also helpful for sustainable resource management, which includes land, water, and forests. Predictive models with high accuracy may anticipate future water shortages in particular areas, which enables the proactive adoption of wise water management practices [6]. Numerous public health emergencies, from the spread of infectious illnesses to an increase in heat wave occurrences, are also linked to climate change. Thorough long-term climate models may provide public health organizations with the information they need to allocate resources and create efficient response plans[7].

Weather forecasting is the practice of projecting future weather conditions. In this research, real-time temperature, humidity, and pressure data from many sensors are used to predict rain. Without human programming, machine learning enables computers to learn from experience and become more efficient. Data analysis and prediction have become more easier since the machine learning idea was introduced. Machine learning uses historical data to forecast future data rather than requiring an understanding of the physical mechanisms controlling the environment. Consequently, this procedure might be used to weather forecasting. [8].

Humans are facing a number of issues as a result of weather changes. One strategy to reduce harmful effects is to forecast the weather and climate. Regretfully, even accurate climate and weather prediction models take a long time to provide forecasts and are not very accurate for longer than a week [9]. Lately, current numerical simulation models have been enhanced with the application of machine learning and deep learning models. A large-scale EuroHPC2 project called MAELSTROM1 aims to enhance machine learning's application in weather and climate modeling in three areas: workflow, machine architectures suitable for ML-augmented Workload Characterization modeling, and applications amenable to ML augmentation [10]. Six distinct deep learning and machine learning applications are available on MAELSTROM. The majority of applications, such as temperature downscaling, weather forecasts to assist energy production, and forecast post-processing for improved local weather forecasts, use neural networks to predict weather more quickly. Second application is still being worked on. There is a lot of data in the majority of the apps for testing and training. Each application is expected to collect an average of 10 TB of data in the future, which will make training, testing, and operating this application more challenging. The main challenge is getting all six apps to efficiently use the computer technology and provide results quickly. Within the MAELSTROM project, our goals are to create performance prediction models that enable us to explore the design space for appropriate future architectures without having to construct them, and to get a full understanding of the features of the MAELSTROM applications on contemporary hardware..

Therefore, a unique machine learning-based weather forecasting prediction approach is described in this paper. The remaining content is arranged as follows: The literature review is described in Section II. The weather interactive prediction system's machine learning technique is shown in Section III. The outcome analysis of the suggested technique is covered in section IV. Section V serves as the work's conclusion

2. LITERATURE SURVEY

The project Fleet Weather Map, presented by M. Hellweg, J. -W. Acevedo-Valencia, Z. Paschalidi, J. Nachtigall, T. Kratzsch, and C. Stiller, et al. [11], looks at the possibility of employing data from floating cars as a source for meteorological information. A larger network of measurements is required to improve the temporal and geographical resolution of weather predictions and, consequently, provide safe autonomous driving features. Moreover, the necessity of raw signal quality control and bias adjustments is demonstrated. The approach's potential seeks to increase the forecast step width to five minutes and yields first positive results.

The current paper [12] by G. Molinar, J. Bassler, N. Popovic, W. Stork, et al. examines current-carrying capacity forecast models using online Numerical Weather Prediction (NWP) data. Feed forward and convolutional neural networks have been used for this job. In the first, the accuracy of the ampacity forecast is directly optimized by interpolating the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

NWP to the overhead line. The second method treats the NWP findings' spatial grid as though it were an image's pixels. Because convolutions may identify pertinent spatial and temporal patterns from the data and integrate them into the ampacity forecast performance, they are crucial to this method. In this work, the ampacity prediction from the closest NWP grid point is directly calculated against the output of these machine-learning-based forecast models. For this case study, a standard open-source dataset was created as a guide for further research in this field.

According to Hassina Ait Issad, Rachida Aoudjit, Joel J.P.C. Rodrigues, et al. [13], agriculture is still an important industry in the majority of nations. It offers the world's population their primary food supply. Its main task, though, is to produce more and better while boosting sustainability and using natural resources sensibly, minimizing environmental damage, and adjusting to climate change. Therefore, it is crucial to transition from traditional to contemporary agricultural practices. One way to achieve environmental standards and address the rising need for food is through smart agriculture Information is becoming more and more important in smart agriculture. Information about insects, diseases, soils, seeds, fertilisers, and other related topics is crucial to the sector's sustainable and profitable growth. Data collection, transmission, selection, and analysis are the components of smart management. Robust analytical tools capable of processing and analyzing massive volumes of data are crucial in order to generate more precise forecasts and more trustworthy information, as the amount of agricultural data is increasing considerably. It is anticipated that data mining would be crucial to handling real-time data analysis with vast data in smart agriculture.

The notion of crowd sensing was explained by Federico Montori, Luca Bedogni, Luciano Bononi, and others [14]. In this method, individuals exchange data from their smartphones with environmental phenomena. They unveiled Sen-Square, an architecture that manages data from crowd sensing platforms and IoT sources and presents it to subscribers in a unified manner. The environment of smart cities is monitored using this data. But none of these pieces make advantage of the notion of merging information from nearby locations.

Using the data from the previous two days, Mark Holmstrom, Dylan Liu, Christopher et al. [15] suggested a method to predict the maximum and lowest temperatures of the upcoming seven days. They used a functional linear regression model that was modified in addition to a linear regression model. They demonstrated that for up to seven days of prediction, professional weather forecasting services beat both models. Their approach, however, does a better job at predicting later dates or longer time horizons.

In their study, C. Feng, J. Zhang, W. Zhang, B.-M. Hodge, et al. [16] used the deep convolutional neural network model for long-term time series solar energy forecasting. According to experimental data, CNNS routinely outperform shallow machine learning models when it comes to weather forecasting, with an average improvement rate of around 7%.

Based on numerical weather prediction analysis, B. He, L. Ye, M. Pei, P. Lu, B. Dai, Z. Li, and K. Wang et al. [17] suggested a combination model for short-term wind power forecasting. under this model, wind power was predicted using both CNN and LSTM networks under varying weather scenarios. The prediction outcomes from the two models were then combined using the IOWA operator. The findings of the experiment demonstrate that the suggested technique may significantly increase the accuracy of wind power prediction under various weather conditions when compared to the Radial Basis Function (RBF), Extreme Learning Machine (ELM), and Support Vector Machine (SVM) methods. Currently, as a result of extensive study on ensemble learning, academics are progressively accepting of its broad meaning. It describes a method of teaching several student groups without recognizing the differences in the types of learners.

An autonomous visual categorization method was proposed by X. Zheng, W. Chen, Y. You, Y. Jiang, M. Li, T. Zhang, et al. [18] by combining deep learning with ensemble learning. To increase the model's capacity for generalization, the technique utilizes the Bagging algorithm and incorporates the Swish activation function into the LSTM network.

A stacking learning framework was developed by Y. Lu and S. Z. Zheng et al. [19] based on five base classifiers: nearest neighbour, logistic regression, naïve Bayes, decision trees, and rule learning for the classification ensemble problem. It was then compared to techniques like voting, AdaBoost, Bagging, Random Forest, and Cross-Validation. According to the experimental findings, the stacking method is better suited for scenarios involving a high number of samples and has the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

strongest generalizations ability. A novel approach based on support vector machines was put out by L. Shi, J. Zhang, D. Zhang, T. Igbawua, Y. Liu, and others[20] to automatically identify sandstorms using data from remote sensing. The experimental findings demonstrate the effectiveness of the SVM-based supervised classification strategy for SDS detection.

W. Wang, P. De Maeyer, Y. Ge, A. Samat, J. Abuduwaili, and T. In an effort to address the low efficiency of manually labelled samples, Van De Voorde Wei et al. [21] suggested a new technique for mixed identification of sandstorms based on MODIS data of the GEE platform to assist in automatically labelling training samples. With this approach, the false positive rate can be significantly decreased and the sandstorm detection task's accuracy rate may exceed 98%..

Turkey electric load time forecasting experiments were conducted by A. Tokgoz and G. Unal et al. [22], who also investigated the application study of RNN in the electric load area. They employed RNN-based variant networks, LSTM, and GRU. The experimental findings demonstrate that this method's forecasting success rate is raised by 2.6% and 1.8%, respectively, when compared to the current power load forecasting techniques based on ARIMA and artificial neural networks..

A neural network prediction model based on long short-term memory (LSTM) was suggested by T. G. Huang, L. Yu, et al. [23] to address the long-term reliance and complexity of financial time series prediction. The model extracts features from the fundamental market data and financial time series technical indicators using the stacked denoising self-encoding process. The experimental findings demonstrate that the prediction model based on LSTM neural network has greater prediction accuracy when compared to standard neural networks.

The deep convolutional neural network model was utilised by C. Feng, J. Zhang, W. Zhang, B.-M. Hodge, et al. [24] in their study on long-term time series solar energy forecasting research. According to experimental data, CNNS routinely outperform shallow machine learning models when it comes to weather forecasting, with an average improvement rate of around 7%.In order to determine the origin of the sand-dust storm in Khuzestan Province, southwest Iran, H. Gholami, A. Mohamadifar, and A. L. Collins, et al. [25] employed eight machine learning techniques, including Random Forest, Support Vector Machine, BART, Radial Basis Function, XGBoost, RTA, BRT, and EM algorithms. The EM algorithm has the best prediction accuracy, according to the data, with an AUC index of 99.8%.

3. FRAMEWORK OF NOVEL TECHNIQUE FOR PREDICTION OF WEATHER FORECASTING USING MACHINE LEARNING

Figure 1 in this part shows a block schematic of a unique machine learning-based weather forecasting approach. This comprises satellite and ground-based cloud imagery as well as statistical characteristics used in weather attribute (rainfall) forecasts. Sky Finder is one of the datasets used for training and testing purposes.

Preparing the data for statistical parameter cleaning, preprocessing satellite and ground-based picture data to eliminate any noise, and preparing the image for a subsequent cloud classification method that will be applied to rainfall forecasts are all included in preprocessing. Picture Pre-processing steps might include actions to enhance the quality of the picture by removing undesired distortions and preparing it for feature extraction in a later stage. When weather forecasters discuss humidity, they may use the phrases absolute humidity and relative humidity interchangeably.

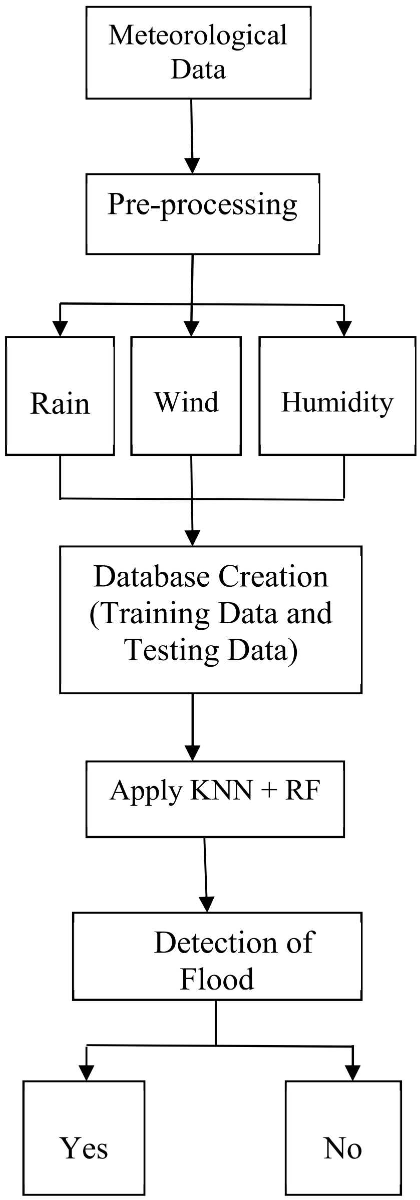

Fig.1: Block Diagram Of Novel Technique For Prediction Of Weather Forecasting Using Machine Learning

The ratio of water vapor to dry air in a given volume of air at a certain temperature is known as absolute humidity. The air's capacity to contain water vapor increases with temperature. A database will be created by storing the meteorological data that was received during picture pre-processing and statistical parameters that were needed for the model's training and testing. The model created for rainfall forecasting will be trained and tested using the data in the database. Next, use a hybrid KNN and RF model to predict whether or not a flood will occur.

4. RESULT ANALYSIS

This section presents the findings of an innovative machine learning-based weather forecasting system.

Table1. Weather Forecasting Parameters

Parameters	Hybrid (KNN+RF)	DT	KNN	RF
Accuracy	86.7	80.2	83	84.6
Sandstorm Occurrence	91.2	87.4	85	86.2
Prediction Time	86.9	91.7	88.5	90.5

The below graph represents the comparison between the proposed algorithms KNN, RF, DT, and Hybrid model

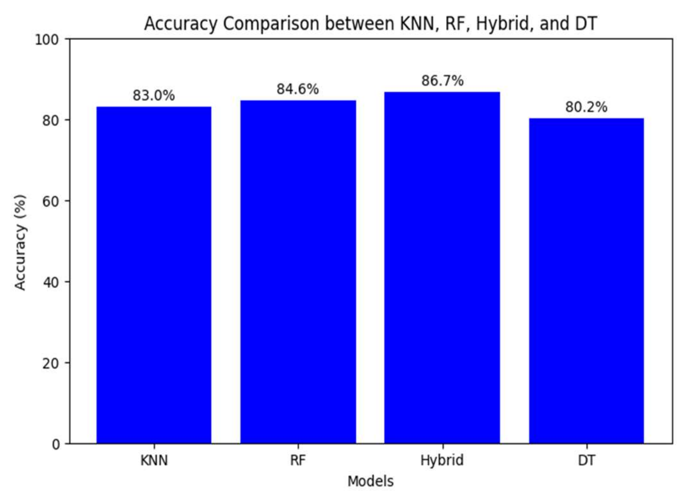

Fig.2: Accuracy Comparison Graph

The below figure compares the occurrence of sandstorms using KNN, RF, Hybrid, and DT.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

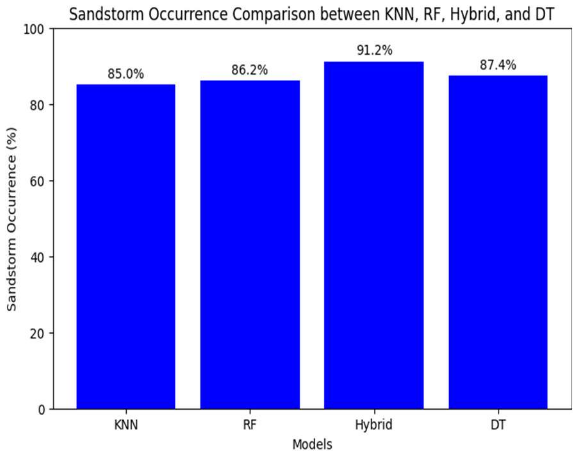

Fig.3 Sandstorm Comparison Graph

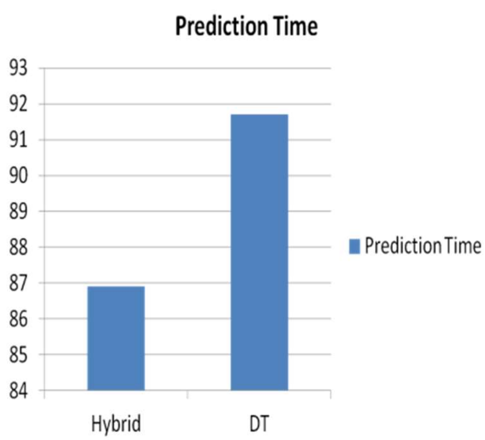

Fig.4 Prediction Time Comparison Graph

Figure 4 compares the prediction times of KNN, RF, Hybrid, and DT.

5. CONCLUSION

Since the effects of climate change are becoming more severe, it is critical to make extremely precise projections. The sheer amount of data involved in the traditional weather forecasting method, which depends on a thorough study of enormous amounts of satellite data, makes it time-consuming. This innovative method offers considerable time savings in the analysis of huge datasets by merging Random Forest (RF) with K-Nearest Neighbours (KNN). It has proven useful in forecasting meteorological variables such as rain, wind, heat, and humidity. Its utility is increased by the application's extension to agriculture. After a thorough comparison with Decision Tree (DT) in terms of sandstorm incidence, forecast time, and accuracy, this hybrid model has proven to perform better than previous approaches, which is a noteworthy development. Combining KNN with RF demonstrates improved prediction performance, which makes it a viable option for more accurate and timely climate forecasts.

REFERENCES

[1] Garima J and Bhawna M, “A Review on Weather Forecasting Techniques,” in International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, no. 12, pp. 177–180, December, 2016.

[2] H. Saima, J. Jaafar, S. Belhaouari and T. A. Jillani, "Intelligent methods for weather forecasting: A review," 2011 National Postgraduate Conference, Kuala Lumpur, pp. 1-6, 2011.

[3] Gujanatti R, Nataraj V, “Intelligent Methods Used for Obtaining Weather Derivatives: A Review,” Engineering and Applied Sciences. Vol. 4, No. 6, 2019, pp. 144-148. doi: 10.11648/j.eas.20190406.12 [4] Gujanatti Rudrappa and Nataraj Vijapur, “Cloud Classification using K-Means Clustering and Content Based Image Retrieval Technique,” International Conference on Communication and Signal Processing – 2020 (iccsp - 20), July 28 - 30, 2020, India

[5] V K Didal, Brijbhooshan, Anita T and K Choudhary, "Weather Forecasting in India: A Review", Intertnational Journal of Current Microbiology and Applied Science, Vol. 6, no. 11, pp-577-590, 2017.

[6] Wei-Ta C, Kai-Chia H and Ali B, "Visual Weather Temperature Prediction," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, pp. 234-241, 2018.Available: https://arxiv.org/abs/1801.08267

[7] Wei-Ta C, Xiang-You Z, Ding-Shiuan D, “Camera as weather sensor Estimating weather information from single images,” Journal of Visual Communication and Image Representation, Vol. 46, pp-233-249, July 2017.

[8] Cewu L, , Di L, J Jiaya J and Chi-Keung T, "Two-Class Weather Classification," IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 39, Issue 12, pp. 2510-2524, 1 December 2017

[9] Sandeep K M, Anamika U and Channabasva G, "Rainfall prediction based on 100 years of meteorological data," 2017 Int. Conf. on

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Computing and Communication Technologies for Smart Nation, Gurgaon, pp. 162-166, 2017.

[10] R. Venkata Rao, "Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems," in International Journal of Industrial Engineering Computations, volume 7, pp.19–34, 2016

[11] M. Hellweg, J. -W. Acevedo-Valencia, Z. Paschalidi, J. Nachtigall, T. Kratzsch and C. Stiller, "Using floating car data for more precise road weather forecasts," *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, Antwerp, Belgium, 2020, pp. 1-3, doi: 10.1109/VTC2020-Spring48590.2020.9129401.

[12] G. Molinar, J. Bassler, N. Popovic and W. Stork, "Ampacity forecasting from Numerical Weather Predictions: a fusion of the traditional and machine learning methods," *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, The Hague, Netherlands, 2020, pp. 514-518, doi: 10.1109/ISGT-Europe47291.2020.9248877.

[13] Hassina Ait Issad, Rachida Aoudjit, Joel J.P.C. Rodrigues, "A comprehensive review of Data Mining techniques in smart agriculture," Engineering in Agriculture, Environment and Food, Volume 12, Issue 4, 2019, Pages 511-525, ISSN 1881-8366, https://doi.org/10.1016/j.eaef.2019.11.003.

[14] Federico Montori, Luca Bedogni, and Luciano Bononi. 2017. A Collaborative Internet of Things Architecture for Smart Cities and Environmental Monitoring. IEEE Internet of Things Journal (2017).

[15] Mark Holmstrom, Dylan Liu, and Christopher Vo. 2016. Machine Learning Applied to Weather Forecasting.

[16] C. Feng, J. Zhang, W. Zhang, and B.-M. Hodge, ``Convolutional neural networks for intra-hour solar forecasting based on sky image sequences," *Appl. Energy*, vol. 310, Mar. 2022, Art. no. 118438.

[17] B. He, L.Ye, M. Pei, P. Lu, B. Dai, Z. Li, and K.Wang, ``Acombined model for short-term wind power forecasting based on the analysis of numerical weather prediction data," *Energy Rep.*, vol. 8, pp. 929_939, Nov. 2022.

[18] X. Zheng,W. Chen, Y. You, Y. Jiang, M. Li, and T. Zhang, ``Ensemble deep learning for automated visual classi_cation using EEG signals," *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107147.

[19] Y. Lu and S. Z. Zheng, ``A comparative study learning and general ensemble method," *Chin. Sci.*, vol. 11, no. 4, pp. 372_379, Feb. 2018.

[20] L. Shi, J. Zhang, D. Zhang, T. Igbawua, and Y. Liu, ``Developing a dust storm detection method combining support vector machine and satellite data in typical dust regions of Asia," *Adv. Space Res.*, vol. 65, no. 4, pp. 1263_1278, Feb. 2020.

[21] W. Wang, A. Samat, J. Abuduwaili, Y. Ge, P. De Maeyer, and T. Van De Voorde, ``A novel hybrid sand and dust storm detection method using MODIS data on GEE platform," *Eur. J. Remote Sens.*, vol. 55, no. 1, pp. 420_428, Dec. 2022.

[22] A. Tokgoz and G. Unal, ``A RNN based time series approach for forecasting Turkish electricity load," in *Proc. 26th Signal Process. Commun. Appl. Conf. (SIU)*, Izmir, Turkey, May 2018.

[23] T. G. Huang and L. Yu, ``Application of SDAE-LSTM model on _nancial time series forecasting," *Comput. Eng. Appl.*, vol. 55, no. 1, pp. 142_148, Aug. 2019.

[24] C. Feng, J. Zhang, W. Zhang, and B.-M. Hodge, ``Convolutional neural networks for intra-hour solar forecasting based on sky image sequences," *Appl. Energy*, vol. 310, Mar. 2022, Art. no. 118438.

[25] H. Gholami, A. Mohamadifar, and A. L. Collins, ``Spatial mapping of the provenance of storm dust: Application of data mining and ensemble modelling," *Atmos. Res.*, vol. 233, Mar. 2020, Art. no. 104716.