International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 13, Issue 9, September 2024)

Review of Weather Forecasting Using Machine Learning Techniques

Abhishek Singh[1], Hitesh Gupta[2]
[1]Research Scholar, Department of Computer Science & Engineering, Lakshmi Narain College of Technology, Bhopal (M.P)
[2]Professor, Department of Computer Science & Engineering, Lakshmi Narain College of Technology, Bhopal (M.P)

***Abstract*— Weather forecasting plays a vital role in various industries and everyday life. Traditional weather prediction methods rely on numerical weather prediction models, which have their limitations in accuracy and computational complexity. With the advancements in machine learning techniques and the availability of large-scale weather datasets, researchers have started exploring the application of machine learning in weather forecasting. This paper presents the progress and challenges in weather forecasting prediction using machine learning. The objective is to leverage machine learning algorithms to improve the accuracy and efficiency of weather predictions.**

Keywords*— *Machine Learning, Weather Forecasting, Prediction.

I. Introduction

Predicting Weather forecasting is the process of predicting the atmospheric conditions at a particular location and time. It involves analyzing various atmospheric variables such as temperature, humidity, air pressure, wind speed and direction, and precipitation patterns. Meteorologists use scientific models, historical data, and observations from weather stations, satellites, and radar systems to make these predictions.

The foundation of weather forecasting lies in numerical weather prediction (NWP) models. These models use complex mathematical equations to simulate the behavior of the atmosphere. By inputting initial conditions and applying the laws of physics, they can estimate how the atmosphere will evolve over time. These models are run on powerful supercomputers to generate forecasts. Observations from weather stations, weather balloons, satellites, and other instruments are crucial for initializing the models accurately. These observations provide data on current weather conditions, which serve as the starting point for the models' calculations. Weather radars are also used to track precipitation, storms, and severe weather events in real-time. As technology advances, weather forecasting has improved significantly over the years. Short-range forecasts, covering a few hours to a few days, tend to be more accurate because they rely on recent observations. Medium-range forecasts, spanning several days to a week, are less accurate due to increasing uncertainty. Long-range forecasts, extending beyond a week, are the least accurate but can still provide some general guidance.

Weather forecasting plays a crucial role in various sectors, including agriculture, aviation, transportation, emergency management, and daily planning for individuals. It helps us prepare for severe weather events, make informed decisions, and mitigate potential risks.

Weather forecasting using machine learning techniques is an emerging field that aims to improve the accuracy and efficiency of weather predictions. Traditional numerical weather prediction models have limitations in handling complex relationships and capturing intricate patterns present in weather data. Machine learning algorithms, on the other hand, have the potential to address these challenges by learning from historical weather data and making predictions based on learned patterns.

Several machine learning techniques have been applied to weather forecasting with promising results. These techniques include:

1. Artificial Neural Networks (ANN): ANN models are widely used for weather forecasting tasks. They consist of interconnected nodes or neurons that mimic the structure of the human brain. ANN models can capture non-linear relationships in weather data and learn complex patterns, making them effective for predicting variables such as temperature, humidity, precipitation, and wind speed.

2. Support Vector Machines (SVM): SVM is a supervised learning algorithm that is commonly used for classification and regression tasks. In weather forecasting, SVM can be trained to predict weather conditions based on historical data. It works by mapping weather data into a higher-dimensional feature space, where a hyperplane is constructed to separate different weather patterns.

3. Random Forests: Random forests are an ensemble learning method that combines multiple decision trees to make predictions. In weather forecasting, random forests can handle high-dimensional datasets and capture complex interactions between weather variables. They are known for their robustness and ability to handle noisy and missing data.

4. Deep Learning: Deep learning techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown promise in weather forecasting. CNNs excel in processing spatial data, such as satellite images, to predict weather patterns. RNNs, on the other hand, are effective in capturing temporal dependencies in weather data, making them suitable for predicting time-series variables.

These machine learning techniques rely on historical weather data, including variables such as temperature, humidity, pressure, wind patterns, and satellite imagery. The models are trained on these datasets, and the learned patterns are then used to predict future weather conditions.

II. LITERATURE SURVEY

Prathyusha, Zakiya et al.,[1] this article is to devise an accurate method by which to estimate the next temperature. Due to the dynamic nature of climate change, there are still substantial challenges to overcome in order to broaden the use of weather predictions in the agricultural industry. Despite the progress that has been achieved in this area, this is the case. These include the need for increased model accuracy, quantitative proof of the value of climate projections as tools for agricultural risk management, and addressing the most significant possibilities of disease occurrence, which are often seasonal and rely on characteristics such as temperature and rainfall. This research aims to anticipate factors that might assist farmers in making an educated choice so that they can avoid losses by adopting necessary preventative steps. This article presents a comprehensive review of several approaches to weather forecasting and investigates potential directions for future study in this area.

C. Li et al.,[2] suggested and compared one deep learning model called LSTM to two classic machine learning models called XGBoost and Polynomial Regression. According to the findings of our experiments, the LSTM model should be considered the most reliable option for forecasting the temperature of the surrounding air. This model has the quickest running time and can even generate predictions using discontinuous temporal information; it has an R-squared accuracy of 98.4% and an RMSE of 1.04. In addition, the other two standard machine learning models that are often utilized are also capable of doing very well when it comes to the prediction of continuous air temperature characteristics.

Y. -L. Li et al.,[3] The forecasting of wind power has garnered an increasing amount of interest over the last several decades. The intermittent nature of wind energy combined with its inherent unpredictability makes it susceptible to producing an unpredictable quality of power supply, which in turn may pose significant problems for industry. As a result, an approach to wind power forecasting that is reliable and precise is particularly vital for a system that uses wind energy. In this study, a unique method for estimating the force of the wind is suggested. The data on the wind that is given by the Central Weather Bureau of Taiwan is used to make the first determination of key elements.

M. Hostetter et al.,[4] The unbalanced class problem is an inherent part of solar flare forecasting, as are other difficulties that we discover in data-driven forecasting problems that are often concealed inside an imbalanced dataset. This problem is also shared by other issues that we find in data-driven forecasting challenges. When dealing with unbalanced data, one approach is to produce synthetic instances of the underrepresented group via the process of synthetic oversampling and then use those examples to correct the imbalance in the data.

J. Sleeman et al.,[5] presented the proposed impact of using a Machine Learning derived Planetary Boundary Layer Height (ML-PBLH) using ground-based Ceilometer observing systems. The inspiration for this paper came from the significance of Planetary Boundary Layer Heights (PBLH) for deducing Air Quality assessments and the unsatisfactory performance of current weather forecasts of PBLH. In the field of air pollution research, the PBLH is an essential tool for determining the degree to which contaminants (such as particles, trace gases, and so on) mix vertically.

A. Doroshenko et al.,[6] provided a concise review of recent developments in numerical weather prediction, obstacles, and the nature of their occurrence, as well as current and prospective solutions to solve these challenges. The design of neural networks has been suggested as a potentially fruitful method that might improve the accuracy of the temperature prediction that the COSMO regional model provides for the 2m timescale. Because of this architecture's design, it is possible to foresee flaws in atmospheric model predictions and their subsequent adjustments. Experiments are performed using histories of regional model mistakes that vary from one another. After a certain number of epochs, network overfitting will occur, and this number may be determined. It has been shown that it is feasible to produce an improvement of a 2m temperature prediction using the architecture that has been suggested roughly 50% of the time.

A. Golder et al.,[7] explored many potential models, some of which are Support Vector Machines (SVM), Multi-Layer Perceptrons (MLP), and Long Short-Term Memory (LSTM). Both the prediction of the Load Demand and the prediction of the PV production were

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 13, Issue 9, September 2024)

accomplished with the help of the models. The dataset that was taken into consideration for the Load Demand prediction models was the aggregated load demand for forty households in Austin, Texas that were picked at random, as well as the weather in Austin, Texas. The data set that was used for the PV generation Model was that of the Yulara Plant in Australia. Additionally, the climatic conditions for the site were also taken into consideration. The MLP model provided us with the greatest results for the Load Prediction, while the LSTM model provided us with the best results for the PV Prediction. Both models were closely followed by the LSTM model.

N. L. et al.,[8] In general, weather is made up of a number of different factors, some of which include rainfall, precipitation, wind speed, and so on. Forecasting the environmental weather is a challenging task for academics, and in recent years it has attracted a significant lot of interest from the scholarly community. Our research takes into account a broad range of different ways for calculating weather figures. These methods may monitor weather in the middle of a season, from month to month, or yearly by taking into consideration the meteorological data that is accessible. Therefore, providing an accurate prediction of meteorological parameters is proving to be a difficult undertaking as a result of the characteristics of their dynamic environment.

C. Feng et al.,[9] purpose of this study is to design a unique dynamic model selection (DMS) technique for STLF that is based on reinforcement learning. The first step is to construct a forecasting model pool, which will ultimately consist of 10 advanced machine learning based forecasting models. After that, a Q-learning agent analyzes the performance of the forecasting models to determine the optimum strategy for choosing the most accurate model to use for the subsequent time step. Using a moving window and using the optimum DMS policy allows for the selection of the best model at each time step in the process. The Q-learning system converges quickly, as shown by numerical simulations conducted on data spanning two years worth of load and weather, which leads to an effective and efficient DMS. When compared to the most advanced machine learning-based STLF models, the newly created STLF model that uses a DMS that is based on Q-learning achieves an improvement in predicting accuracy of roughly fifty percent.

Y. Jiang et al.,[10] The local agricultural and transportation sectors have suffered significant setbacks as a direct result of the regular occurrence of severe thunderstorms and high winds. This article offers a technique for predicting local thunderstorms and severe weather using a multisource convolution neural network that is built to extract the characteristics of weather-related data with many kinds from Doppler radar. The approach is used to make predictions about the occurrence of local thunderstorms and severe weather. Following the training phase in which Center-Loss and Softmax were simultaneously used as loss functions, the features that were acquired were merged with SVM in order to do classification. This was done so that the discriminative power of the features could be improved.

III. CHALLENGES

The Weather forecasting using machine learning techniques faces several challenges that need to be addressed for the development of reliable and accurate forecasting systems. Here are some of the key challenges:

- Data Quality and Availability: Weather forecasting relies on high-quality and diverse datasets. However, ensuring data quality, including accurate measurements, consistent formats, and absence of biases, can be challenging. Additionally, obtaining comprehensive and up-to-date weather data from various sources can be a complex task.

- Missing Data and Data Imputation: Weather data often contains missing values due to sensor malfunctions or data collection issues. Dealing with missing data requires appropriate imputation techniques to ensure the integrity and reliability of the datasets used for training the machine learning models.

- Interpretability and Explainability: Machine learning models, especially complex ones like deep learning architectures, can be difficult to interpret and understand. Interpreting the predictions made by these models and understanding the underlying reasoning or features influencing the forecasts is crucial for gaining trust and confidence in the results.

- Overfitting and Generalization: Overfitting occurs when a machine learning model performs exceptionally well on the training data but fails to generalize to unseen data. Weather forecasting models need to strike a balance between capturing complex patterns in the training data and avoiding overfitting, ensuring accurate predictions for unseen weather conditions.

- Computational Complexity and Efficiency: Weather forecasting requires processing large volumes of data and performing computationally intensive tasks, especially when dealing with high-resolution models or real-time predictions. Balancing computational complexity and efficiency is crucial to provide timely and practical forecasts.

- Uncertainty Quantification: Weather forecasts inherently come with uncertainties. Communicating the uncertainty associated with predictions is essential for decision-making and risk assessment. Developing methods to quantify and communicate uncertainty in machine learning-based weather forecasts is an ongoing challenge.

- Incorporating Physical Constraints: Weather is governed by physical laws and processes. Integrating these domain-specific knowledge and constraints into machine learning models can help improve their accuracy and ensure that the predictions align with known atmospheric behavior.

IV. CONCLUSION

The machine learning has emerged as a promising tool for weather forecasting prediction. Continued research and development in this area have the potential to enhance the accuracy, speed, and reliability of weather forecasts, benefiting industries, emergency management, and individuals in making informed decisions and mitigating risks associated with weather events. Addressing challenges requires collaboration between meteorologists, data scientists, and machine learning experts. By combining domain knowledge, data expertise, and algorithmic advancements, it is possible to overcome these challenges and develop robust and reliable machine learning-based weather forecasting systems.

REFERENCES

1. S. Addu, P. R. Ambati, S. R. Kondakalla, H. Kunchakuri and M. Thottempudi, "Predicting Delay in Flights using Machine Learning," 2022 International Conference on Applied Artificial

Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 374-379, doi: 10.1109/ICAAIC53929.2022.9793243.

2. B. T. L. S. S., H. Al Ali, A. A. A. M. Majid, O. A. A. A. Alhammadi, A. M. Y. M. Aljassmy and Z. Mukandavire, "Analysis of Flight Delay Data Using Different Machine Learning Algorithms," 2022 New Trends in Civil Aviation (NTCA), Prague, Czech Republic, 2022, pp. 57-62, doi: 10.23919/NTCA55899.2022.9934398.
3. M. -T. Vo, T. -V. Tran, D. -T. Pham and T. -H. Do, "A Practical Real-Time Flight Delay Prediction System using Big Data Technology," 2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), Solo, Indonesia, 2022, pp. 160-167, doi: 10.1109/COMNETSAT56033.2022.9994427.
4. R. A. Sugara and D. Purwitasari, "Flight Delay Prediction for Mitigation of Airport Commercial Revenue Losses Using Machine Learning on Imbalanced Dataset," 2022 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM), Surabaya, Indonesia, 2022, pp. 1-8, doi: 10.1109/CENIM56801.2022.10037369.
5. M. Guimarães, C. Soares and R. Ventura, "Decision Support Models for Predicting and Explaining Airport Passenger Connectivity From Data," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 9, pp. 16005-16015, Sept. 2022, doi: 10.1109/TITS.2022.3147155.
6. S. Sharan, M. Sriniketh, H. Vardhan and D. Jayanth, "State-Of-Art Machine Learning Techniques to Predict Airlines Delay," 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS), Bengaluru, India, 2021, pp. 1-6, doi: 10.1109/FABS52071.2021.9702590.
7. C. Y. Yiu, K. K. H. Ng, K. C. Kwok, W. Tung Lee and H. T. Mo, "Flight delay predictions and the study of its causal factors using machine learning algorithms," 2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Changsha, China, 2021, pp. 179-183, doi: 10.1109/ICCASIT53235.2021.9633571.
8. R. Hendrickx, M. Zoutendijk, M. Mitici and J. Schäfer, "Considering Airport Planners' Preferences and Imbalanced Datasets when Predicting Flight Delays and Cancellations," 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 2021, pp. 1-10, doi: 10.1109/DASC52595.2021.9594367.
9. J. Huo, K. L. Keung, C. K. M. Lee, K. K. H. Ng and K. C. Li, "The Prediction of Flight Delay: Big Data-driven Machine Learning Approach," 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, Singapore, 2020, pp. 190-194, doi: 10.1109/IEEM45057.2020.9309919.
10. M. Bardach, E. Gringinger, M. Schrefl and C. G. Schuetz, "Predicting Flight Delay Risk Using a Random Forest Classifier Based on Air Traffic Scenarios and Environmental Conditions," 2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 2020, pp. 1-8, doi: 10.1109/DASC50938.2020.9256474.