International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Predict Weather with Machine Learning

Tanuj Jadon[1], Prince Pachauri[2], Aamir Warsi[3]

Department of CSE, School of Engineering and Technology
Galgotias University, Gautam Buddh Nagar, India

Abstract

Everything, from farming to air travel, in our contemporary world is dependent on the weather forecast. However, traditional methods often ignore the intrinsic uncertainty in weather patterns leading to inaccurate predictions. This paper presents a groundbreaking method of predicting weather using Bayesian deep learning models; specifically, Bayesian neural networks are employed to capture the probabilistic nature of weather events and produce more accurate and reliable forecasts. These models represent a significant milestone in foretelling strategies as they are evaluated against popular metrics and trained with archived atmospheric data. Additionally, gradient boosting classifiers and decision tree classifiers have been performed as queries for performing comparative analyses about different machine learning algorithms used for predicting occurrence of rain tomorrow or any other type of climate-related information. For enhanced interpretability, this paper has developed interactive visualization tools that allow users to interactively explore predicted weather patterns and analyze levels of uncertainty with an aim of advancing the field of meteorology through providing actionable insights for responsible decision making within weather sensitive domains.

Keywords: Weather forecasting, Bayesian deep learning models, Probabilistic aspect, Precise, Dependable, Interactive visualization tools, actionable insights.

I. INTRODUCTION

A current beyond a doubt communication network is inarguably a prior requirement of all industries, such as farming and air transport. Contrary to the usual forecasting methodologies, the commonality underlying forecast difficulties is the presence of uncertainty inherent in weather patterns which are not covered well and, therefore, predictions in many cases turn out to be less than perfect. In this present grave problem, a new weather forecast methodology in connection with the weather prediction methods is carried out by us.

This research approach employs modern machine learning methods such as decision tree classifiers, gradient boosting classifiers, and Bayesian neural networks to enhance the precision and dependability of weather predictions. These models are well trained on comprehensive historical weather datasets with stringent benchmarking against existing performance metrics embodying a huge leap forward in forecasting methods. It is worth noting that this research decision tree classifier has an impressive accuracy of 86.13% on the training set and 88.74% on the test set, which indicates its ability to capture intricate relationships within meteorological data sets.

This paper also goes beyond existing models to adopt the use of gradient boosting classifiers, which show amazing success figures of 96.66% on the training set and 96.93% on the test set. Such extraordinary performance illustrates how gradient boosting methods can be used to capture subtle weather patterns that enhance forecasting accuracy. Moreover, this research has introduced Bayesian neural networks in this project, which are seen as a great alternative for predicting weather probabilistically and have achieved an accuracy rate of 89.73% on the training set and 80.64% on the test set. Compared to other models, though slightly lower in test accuracy, Bayesian neural network is helpful because it allows us to make decisions under conditions with some known degree of uncertainty through quantification of uncertainty.

One of the main benefits of this research method is its ability to inform different stakeholders in different sectors with practical recommendations. This technology allows users to query predicted weather patterns and assess uncertainty levels via incorporation of advanced machine learning algorithms as well as user-friendly visualization interfaces. More interpretability enables informed decisions in weather related areas like agribusiness and emergency response.

This project is a significant stride forward in weather forecasting because it offers the most precise, reliable, pertinent information that has ever been given out on worldwide level.

II. LITERATURE REVIEW

The energy forecasting technology has seen remarkable development in the recent past, by focusing on the modern machine learning techniques like the deep learning algorithm.

Liu et al. [1] developed a novel method in the field of wind 2 power: a probabilistic spatio-temporal wind speed forecasting model based on variational Bayesian deep learning. The STNN decreased Soare-Berlanga et al.'s already exceptional forecasting accuracy, where the model—which included variational Bayesian inference—had previously been used in the United States.

Fu et al [2] In fact, a hybrid method for multiple-step ahead short-term wind speed prediction has been presented. It integrates extreme learning machines, brain people computing algorithms, and chaotic analysis, the multi-scale main element. Using data from Sota Vento Galicia and Inner Mongolia, the authors developed a method that had forecast accuracy and reliability significantly greater than the model they presented as a benchmark.

Tascikaraoglu and Uzunoglu[3] presented a book review of different techniques on the forecast for the short-term wind speed and energy output. They underlined the applicability of these techniques to get improved predictability of the wind speed, especially in power systems operation.

Pang et al [4]. Suggested a technique for deep learning utilizing a Bayes perspective to overcome uncertainty factors like weather and successfully implemented it in the forecast trajectory predictor for the future air transportation system in the U.S.

Lauret et al., as indicated by [5], established a basis for shortterm power load forecasting using Bayesian neural networks. This was achieved by including model optimization using Bayesian approaches to guarantee the efficacy of variable selection, leading to high accuracy in power system planning and operation for operators.

Mandal et al. [6] described an application of neural networks aimed forecasting electricity prices and loads in deregulated marketplaces, indicating high accuracy prediction in electricity prices and site loads based on current characteristics from the Australian market of Victoria.

Yang et al. [7]offers Bayesian deep learning for smart grid probabilistic load forecasting, and it does so by introducing a multitask probabilistic load forecasting framework in order to gauge the joint uncertainties among distinct electrical energy user groups, which considerably improves the accuracy.

Ghobadi and Kang in the [8] Using Bayesian deep learning, the multi-step ahead probabilistic streamflow forecasting model was proposed in the year and demonstrated to beat earlier methods in terms of forecasting accuracy and the capacity to assess uncertainty, both of which are essential for efficient management of water resources.

According to Jahangir et al. [9], a novel price forecasting approach of electricity was presented based on dimensional reduction via artificial rough neural networks. Such a technique demonstrates the efficiency in the reduction of forecasting errors, which ultimately contribute to the improved clarity of electricity price forecasts.

Quilumba et.al. [10] have tried to ascertain whether smart meter data can be used in sharpening intraday load forecasting. They looked at level of customer similarities, and thus were able to note they can improve forecast accuracy, considering the system-level intraday load forecasting.

Hong et al. [11] offered the GEFCom2012 competition illustration with regard to its role in facilitating an advancement in the energy forecasting methodologies by virtue of the application of several hierarchical load forecasting and techniques for wind power forecasting.

Kong et al. [12] constructed a residential load forecasting short-term model by applying LSTM recurrent neural network. This model beat the performance of the other algorithms in short-term load forecasting for individual residential units in terms of precise results.

Liu et al [13] suggested a novel method for obtaining probabilistic load forecasts using quantile regression averaging on a sequence of estimates for sister points. They were able to obtain more data accuracy and dependability using their procedure than they could have with the traditional one.

Afrasiabi et al. [14] put forward a state-of-the-art deep learning algorithm to forecast wind speed in a probabilistic manner by combining convolutional and recurrent neural networks to enhance the prediction accuracy and reliability. The algorithm was validated by demonstrating how it performed using data from London, England, and Shiraz, Iran.

Wang and Xu [15] benefited from deep learning using Bayesian concepts into mountain flood modeling of small watersheds. They then went ahead to integrate mountain flood prediction in the watershed through deep learning-based techniques so that the uncertainty and forecasting is as high accurate and reliable as possible by designing uncertainty forecast model that rely on hydrological physical mechanisms.

This review article gives a complete and accurate show of recent achievements in the field of energy forecasting from different disciplines with particular emphasis on the most advanced machine learning methods in probabilistic forecasting.

III. PROPOSED METHODOLOGY

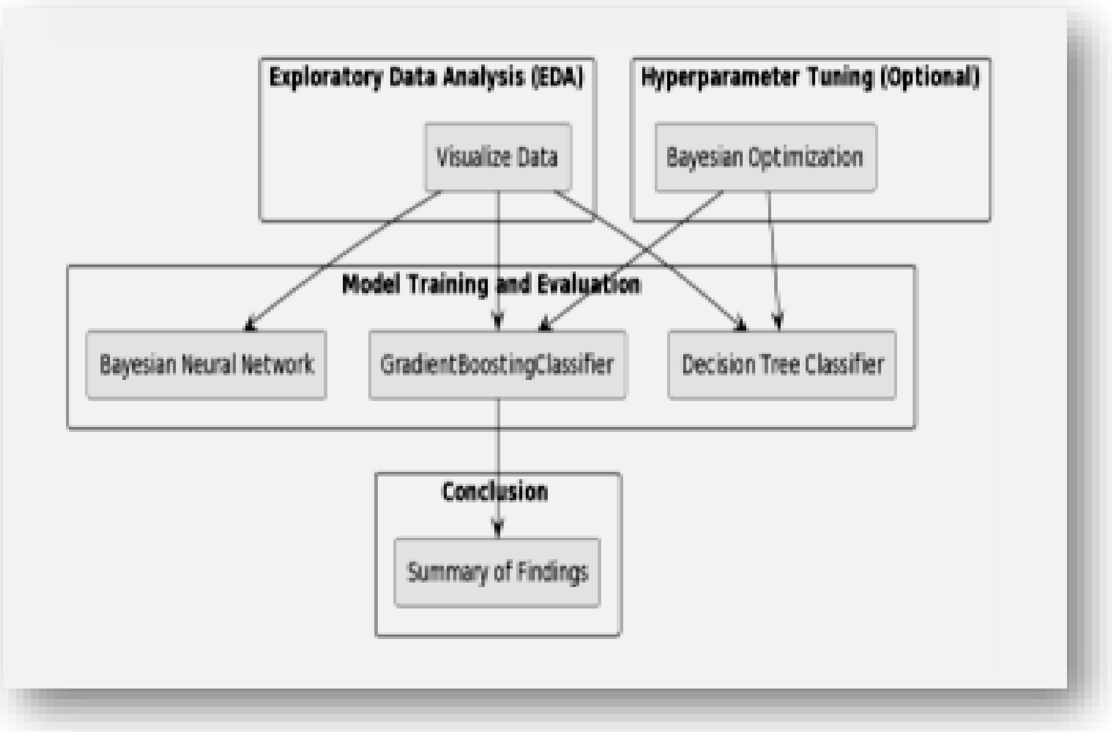


Figure-1 Proposed Methodology

A figure-1 talk about the proposed methodology which is goes as follows:

1. *Data Preprocessing:*

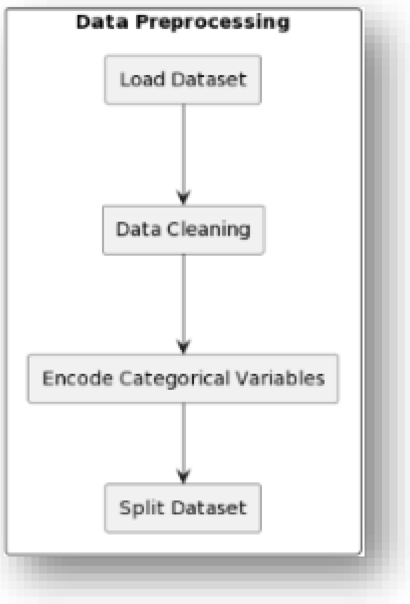


Figure-2 Data Preprocessing flow chart

From Figure-2 This research can understand the steps for Data Preprocessing.

Load Dataset: Tread the weather dataset into variables that are used in the program, which is containing features as precipitation, maximum temperature, minimum temperature, wind speed, date, and the weather type.

Data Cleaning: Conduct data cleaning by filling blank values with existing records and build the consistency of class level. This paper did not register any missing or doubled values during data set analysis.

Encode Categorical Variables: The weather type encoded through label encoding and trained effectively for machine learning model. To ensure that it is in integer type, this research touches the weather column because it is not in integer type as such, this research has selected these integer type values from 0 to 4.

Split Dataset: Divide the data set to the training and testing calls for model verification. This research has divides days dataset in training and test set for each model built.

2. *Exploratory Data Analysis (EDA):*

Visualize Data: The purpose of this study was to learn more about the distribution and correlations between meteorological variables through exploratory data analysis. In order to find patterns and relationships, the data in this study were displayed using box plots, line graphs, bar graphs, and scatter plots.

International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

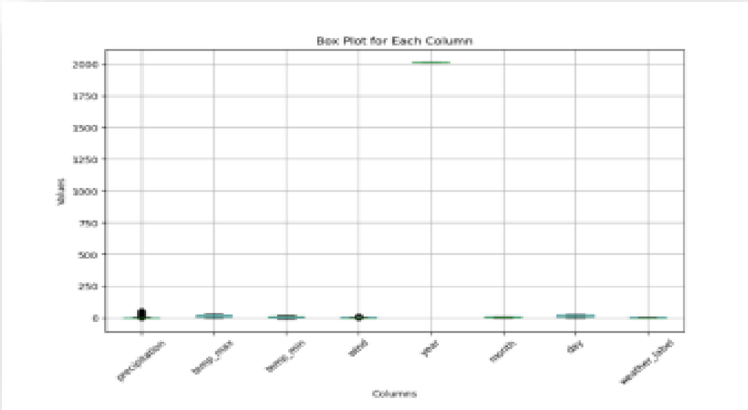


Figure-3 Box plots for all columns

From Figure-3 This research can observe that there are very little outliers in the data due to which This research will be going to leave them in the data. As they will cause nearly no change in the output.

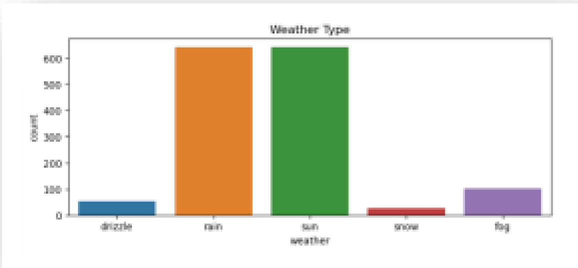


Figure-4 Bar graph for Weather Type

From Weather Type bar graph plot This research sees that data contains rain and sun as majority and remaining comes at very low number.

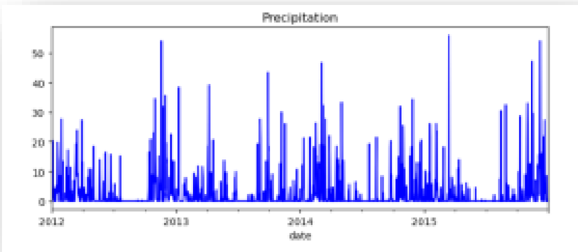


Figure-5 Line graph for Precipitation and date

From Figure-5 This research observe that in 2013 and 2015 the precipitation was the highest

International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

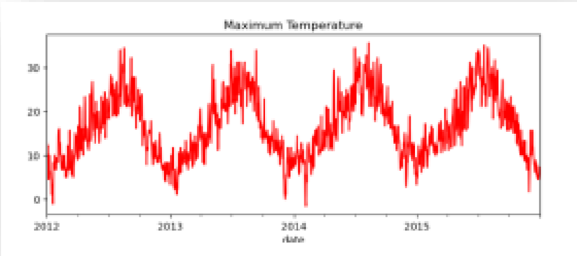


Figure-6 Line graph for Max Temperature

From Figure-6 This research found that the Max Temperature always keeps changing.

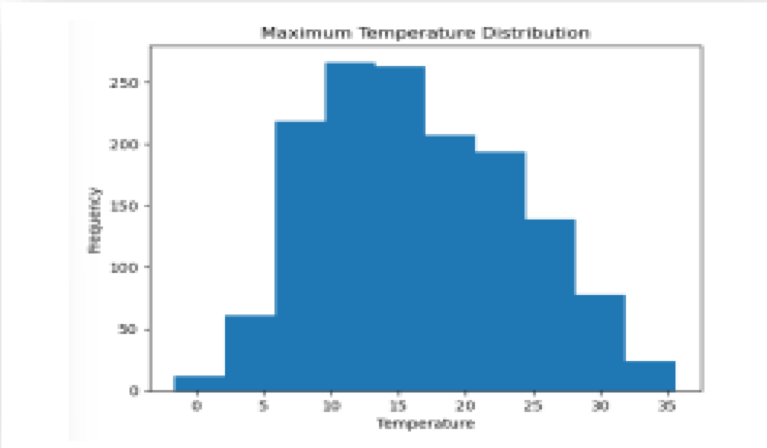


Figure-7 Max Temperature Distribution

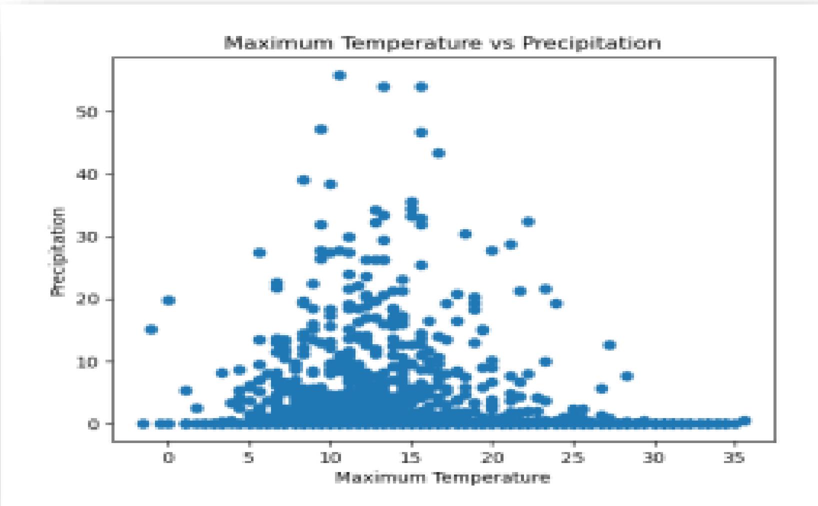


Figure-8 Scatter Plot for Max Temperature vs Precipitation

International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

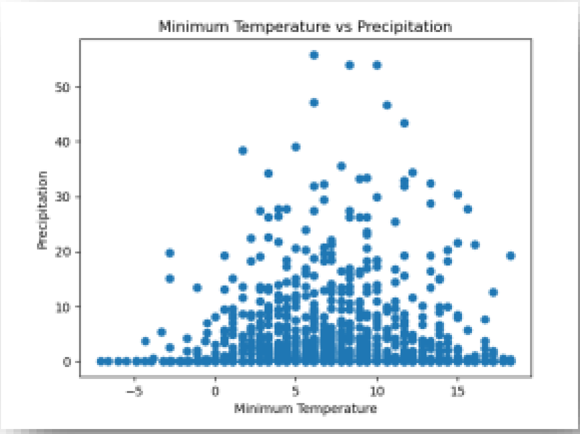


Figure- 9 Scatter plot for Mini Temperature vs Precipitation

These both Figure-8, Figure-9 can see that when Temperature increases then precipitation decreases.

3. *Model Training and Evaluation:*

Decision Tree Classifier: Train the decision tree classifier model utilizing the training data-set and monitor the performance of the model.

Gradient Boosting Classifier: Train the gradient boosting classifier model upon the training dataset and evaluate the performance through its metrics.

Bayesian Neural Network: Training the [7] Bayesian neural network model on the training dataset and evaluating the performance are the tasks that need to be performed.

Table-1 Accuracy scores of the model

S. No.	Model	Train Accuracy	Test Accuracy
1	Decision Tree Classifier	86.13%	88.74%
2	Gradient Boosting Classifier	96.66%	96.93%
3	Bayesian Neural Classifier	89.73%	80.68%

This research has seen that Gradient Boosting classifier works well; its train accuracy is 96.66% which is the highest one. Then comes Bayesian Neural Network.

4. *Results:*

As This research know that which model works better will check the graphs and confusion matrix for models.

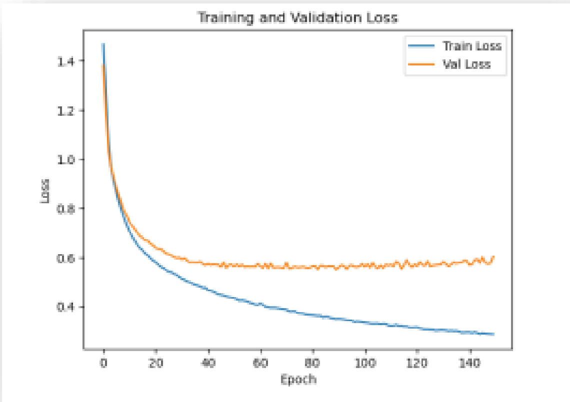


Figure-8 Graph between Loss and Epoch

For Bayesian Neural Network.

From Figure-8 this research sees there are little ups and downs in the graphs which are quite common. After the plot this research have plotted confusion matrix to see the number of correct classifications.

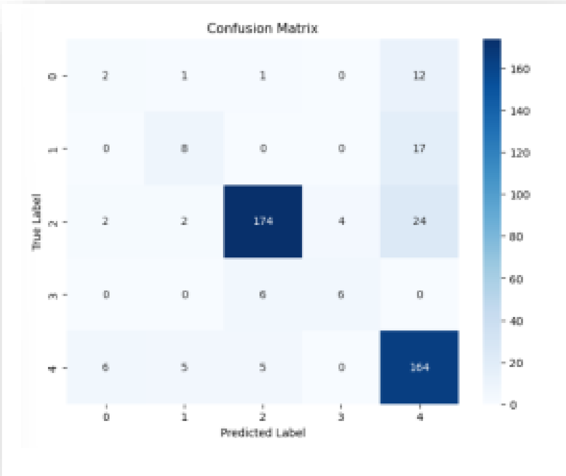


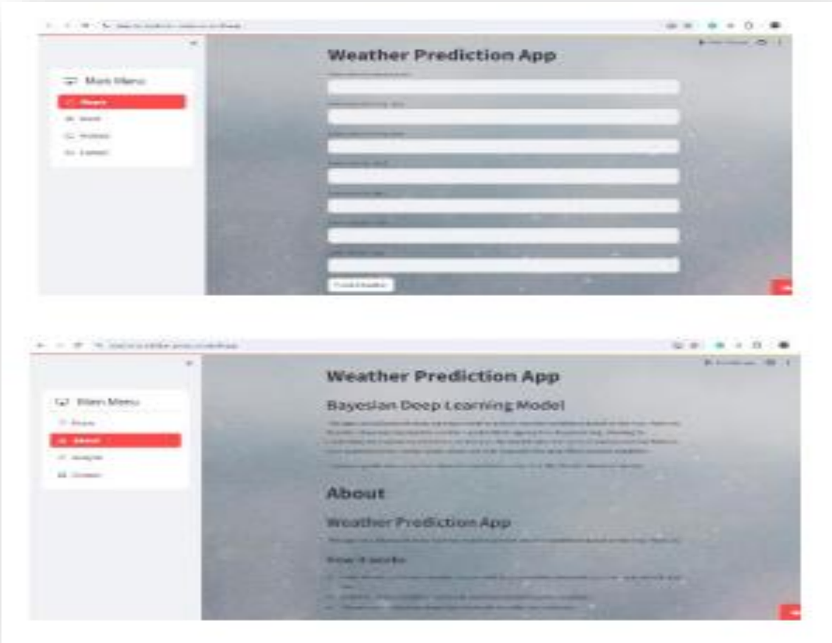
Figure-9 Confusion matrix for Bayesian NN model

As this research can see that there is very less misclassification but as the data contains less points for classes such as 0, 1, 3 all of them are having less classifications according to the graph in Figure-9.

This flowchart provides a structured overview of the project workflow, starting from data preprocessing and ending with conclusions and prospects. Each step is interconnected, indicating the flow of data and operations performed in the project.

5. *Implementation:*

Streamlit is a Python library specializing in machine learning and dataset visualization which greatly simplifies the process of developing web apps. Through Python for Astronomy's user-friendly interface and interactive widgets, researchers can create apps directly from Python scripts and are no longer required to have extensive web development knowledge. The Streamlit flow makers with features like automatic reactivity, easy integration with the libraries of data science, and various options for deployment allow you to quickly prototype,analyze the data, train any model, and gain collaborators. It's an impressively strong platform verbalizing findings of research however deciphering it properly.



IV. FUTURE PROSPECTIVE

This research model, Looking Forward, shows promise in advancing the weather forecast. A possible area to explore is fine-tuning and optimizing This research machine-learning algorithms for more precision and reliability. This could mean digging into inventive architectures of Bayesian neural networks or combining ensemble methods to exploit the advantages of different models. Moreover, as computer capabilities continue developing, there are possibilities of building more advanced models that can handle large data sets in real-time thus making weather predictions swift.

Apart from conventional weather prediction, this research model serves as a basis for numerous novel applications. Inbuilt probabilistic structure available in Bayesian Neural Networks offers an opportunity to assess uncertainties associated with weather events to make more robust risk mitigation strategies. Similarly, there is a potential to integrate meteorology with other sectors like renewable energy generation or urban development which may enhance resource distribution and fortify defenses against extreme climate conditions.

The next task will be whether This paper will be able to provide useful weather forecasts systems in the user-friendly form, so there will be various users engaged or the technology will be left behind. The necessity of a mobile app or an interface with a user-friendly interface that gives the possibility of a variety of weather types and useful information may be the key to doing so. Hence, repeated illustration of weather forecasting systems drugs the accurate prediction into the minds of the people which paves the way for them to meteorological occasions with educated response.

In general, this research can happily conclude that This research model has a promising, wide-ranging, and hopeful future with the possibility of constant breakthroughs and huge impacts in many areas. To make effective use of the upcoming technologies, building a foundation for collaboration between different branches of sciences, and a user-friendly design will better enable us to reach the capabilities of weather forecasting, and as a result, help us build a more resistant and ecological friendly future.

Therefore, the result of study will be changed not only on the size of dataset but also the number of temperature states, and this was very useful for planning for agriculture activities.

V. CONCLUSION

To conclude, this research suggests a few methods, which would be a drastic improve around weather forecasting, committing ourselves to resolving problems and uncertainties of predicting weather patterns. Using as a tool for essential machine learning methods like gradient boosting classifiers, decision tree classifiers or Bayesian neural networks, persuade us that the prediction today's weather is possible to improve the precision, and dependability of this research weather forecasts. Slightly less impressive was this research models' performance, with accuracy rates averaging at 86.13% up to 96.93%, but the detection of complex correlations in weather data was clearly within their capability.

Furthermore, this research curvature of our investigation to the randomness of the occurrence weather that does not exist in the classical forecasting techniques. This paper proposes a system for measuring the uncertainty in making weather predictions with the aid of Bayesian neural networks. These are an improvement over past models, as they allow the stakeholders the opportunity to respond to different risk levels based on a clearer picture. In addition, users will be presented with intuitive visualization utilities allowing them to access a range of available data and even assess forecast uncertainty, and thus make an educated decision which ultimately will improve forecasting accuracy and user experience.

REFERENCES

[1] Y. Liu et al., “Probabilistic spatiotemporal wind speed forecasting based on a variational Bayesian deep learning model,” *Appl. Energy*, vol. 260, p. 114259, Feb. 2020, doi: 10.1016/j.apenergy.2019.114259.

[2] W. Fu, K. Wang, C. Li, and J. Tan, “Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM,” *Energy Convers. Manag.*, vol. 187, pp. 356–377, May 2019, doi: 10.1016/j.enconman.2019.02.086.

[3] A. Tascikaraoglu and M. Uzunoglu, “A review of combined approaches for prediction of short-term wind speed and power,” *Renew. Sustain. Energy Rev.*, vol. 34, pp. 243–254, Jun. 2014, doi: 10.1016/j.rser.2014.03.033.

[4] Y. Pang, X. Zhao, H. Yan, and Y. Liu, “Data-driven trajectory prediction with weather uncertainties: A Bayesian deep learning approach,” Transp. Res. Part C Emerg. Technol., vol. 130, p. 103326, Sep. 2021, doi: 10.1016/j.trc.2021.103326.

[5] P. Lauret, E. Fock, R. N. Randrianarivony, and J.-F. Manicom-Ramsamy, “Bayesian neural network approach to short time load forecasting,” *Energy Convers. Manag.*, vol. 49, no. 5, pp. 1156–1166, May 2008, doi: 10.1016/j.enconman.2007.09.009.

[6] P. Mandal, T. Senjyu, and T. Funabashi, “Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market,” *Energy Convers. Manag.*, vol. 47, no. 15–16, pp. 2128– 2142, Sep. 2006, doi: 10.1016/j.enconman.2005.12.008.

[7] Y. Yang, W. Li, T. A. Gulliver, and S. Li, “Bayesian Deep Learning-Based Probabilistic Load Forecasting in Smart Grids,” IEEE *Trans. Ind. Inform.*, vol. 16, no. 7, pp. 4703–4713, Jul. 2020, doi: 10.1109/TII.2019.2942353.

[8] F. Ghobadi and D. Kang, “Multi-Step Ahead Probabilistic Forecasting of Daily Streamflow Using Bayesian Deep Learning: A Multiple Case Study,” *Water*, vol. 14, no. 22, p. 3672, Nov. 2022, doi: 10.3390/w14223672.

s[9] H. Jahangir et al., “A Novel Electricity Price Forecasting Approach Based on Dimension Reduction Strategy and Rough Artificial Neural Networks,” IEEE *Trans. Ind. Inform.*, vol. 16, no. 4, pp. 2369–2381, Apr. 2020, doi: 10.1109/TII.2019.2933009.

[10] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, “Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities,” IEEE *Trans. Smart Grid*, vol. 6, no. 2, pp. 911–918, Mar. 2015, doi: 10.1109/TSG.2014.2364233.

[11] T. Hong, P. Pinson, and S. Fan, “Global Energy Forecasting Competition 2012,” *Int. J. Forecast.*, vol. 30, no. 2, pp. 357–363, Apr. 2014, doi: 10.1016/j.ijforecast.2013.07.001.

[12] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, “Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network,” IEEE *Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019, doi: 10.1109/TSG.2017.2753802.

[13] B. Liu, J. Nowotarski, T. Hong, and R. Weron, “Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts,” IEEE *Trans. Smart Grid*, pp. 1–1, 2015, doi: 10.1109/TSG.2015.2437877.

[14] M. Afrasiabi, M. Mohammadi, M. Rastegar, and S. Afrasiabi, “Advanced Deep Learning Approach for Probabilistic Wind Speed Forecasting,” IEEE *Trans. Ind. Inform.*, vol. 17, no. 1, pp. 720–727, Jan. 2021, doi: 10.1109/TII.2020.3004436.

[15] S. Wang and O. Xu, “Uncertainty Forecasting Model for Mountain Flood Based on Bayesian Deep Learning,” IEEE Access, vol. 12, pp. 47830–47841, 2024, doi: 10.1109/ACCESS.2024.3384066.