Machine Learning Prediction Models applied to Weather Forecasting: A survey

Israa Jasim Mohmmed*, Bashar Talib AL-Nuaimi **, Asst . Prof. Dr. Dher Intisar Bakr**

*Department of Computer Science, College of Science, Diyala University, Iraq

** College of Computer Science, University of Diyala, Iraq

Abstract

Using scientific knowledge, weather forecasters can predict what the atmosphere will be like in a particular place. It predicts snow, cloud cover, rain, temperature and wind speed. Since weather predictions consist of multidimensional and nonlinear data, they are one of the world's most challenging problems. Various machine learning algorithms and methods have been used in data mining for weather prediction, including Support Vector Machines, supervised and unsupervised machine learning algorithms, artificial neural networks, FPGrowth Algorithms, Hadoop with Map Reduce, K-medoids, and Naive Bayes. This survey briefly explains the methods used to build weather forecasting models to assist researchers in choosing the appropriate method for their model.

Keywords- *weather forecasting, Precise Forecasts, Outlying Regions.*

I. INTRODUCTION

Weather forecasting aims to anticipate atmospheric conditions in a certain location and time frame [1]. "Climate change" is the term used to describe significant, long-term changes in the global climate [2] . The earth, sun, winds, oceans, snow, rains, deserts, savannas, woods, and everything people do all contribute to the global climate. Several topical articles provide in-depth analyses of machine learning algorithms, the most significant class of AI techniques (Figure 1) in the field of meteorology. Types of weather forecasting is illustrated in Figure 2. Many techniques and their categories are described in these books. Supervised learning (Figure 3), which is the most prominent set of approaches in the most recent publications in the area, was judged to be the most intriguing for atmospheric scientists. If one has access to labeled data, they may use it to train a function that maps inputs to desired outcomes. After the model has been tested on a separate dataset (the "testing" dataset), it may be utilized for classification and regression in any necessary applications if the results are satisfactory. Methods like Decision Trees, Random Forest [3], XG Boost, Deep Learning, and Support Vector Machines are all included in this category (SVM). Second, unsupervised learning as shown in Figure 2, is a subfield of machine learning in which algorithms are not provided with labeled data for training and must instead determine how to partition or decrease the dimensionality of a dataset before proceeding with analysis. K-means Clustering (K-means) and Principal Component Analysis (PCA) are two of the most widely used techniques in the field of meteorology and climate science, respectively.

Several studies presented on weather prediction, Mark Holmstrom *et al*, [4] In Using meteorological data from the previous two days, this paper demonstrates machine learning algorithms used to weather predicting based on the highest temperature and the minimum temperature for seven days. A functional regression model variant that could identify weather trends was utilized in addition to a linear regression model. Professional weather predicting services beat both of our models, but the difference between the two quickly closed for projections of later days, suggesting that models may be able to outperform professional ones for even longer time scales.

Anusha *et al*, [5]. In this study, the weather state is forecasted using the given features. Most current systems depend on statistical techniques like support vector machine (SVM), which are unable to produce precise forecasts because they do not take into account rapidly changing meteorological conditions. The model used the multi-linear regression concept to achieve better results than current approaches.

A.S. Alhumaima and S.M. Abdullaev,[16] The suggested system uses site-specific characteristics (latitude, longitude, minimal, maximum, and mean heights, as well as the type of landcover) and seasonal meteorological factors to anticipate the vegetation's composition in the Diyala River basin (precipitation sum, and minimal, maximal and average daily temperatures). The best perceptron's forecasting performance showed a significant spatial heterogeneity: $R2$ was 0.76-0.80 in the upper part of the basin, which is classified by Köppen-Trewartha bioclimatic classification as landscapes of temperate mountain climate and the subtropical climate with dry summers, while $R2$ was 0.6-0.7 in the dry steppe and semi-desert landscapes of Diyala downstream.

K. Ramesh *et al*, [15]: They created a method called Linear Regression Based Lead Seven Days Minimum and Maximum Air Temperature Prediction, which develops numerous linear regression models based on numerical meteorological factors for lead seven day minimum and maximum temperature predictions. The analysis confirms that regression-based minimum temperature prediction models are more precise than maximum temperature forecast models by the dataset's lowest MAE and RMSE, highest R2, and highest R2 values. The experiment also showed how forecast accuracy improves at shorter lead times for both minimum and maximum temperatures and quickly deteriorates at longer lead times.

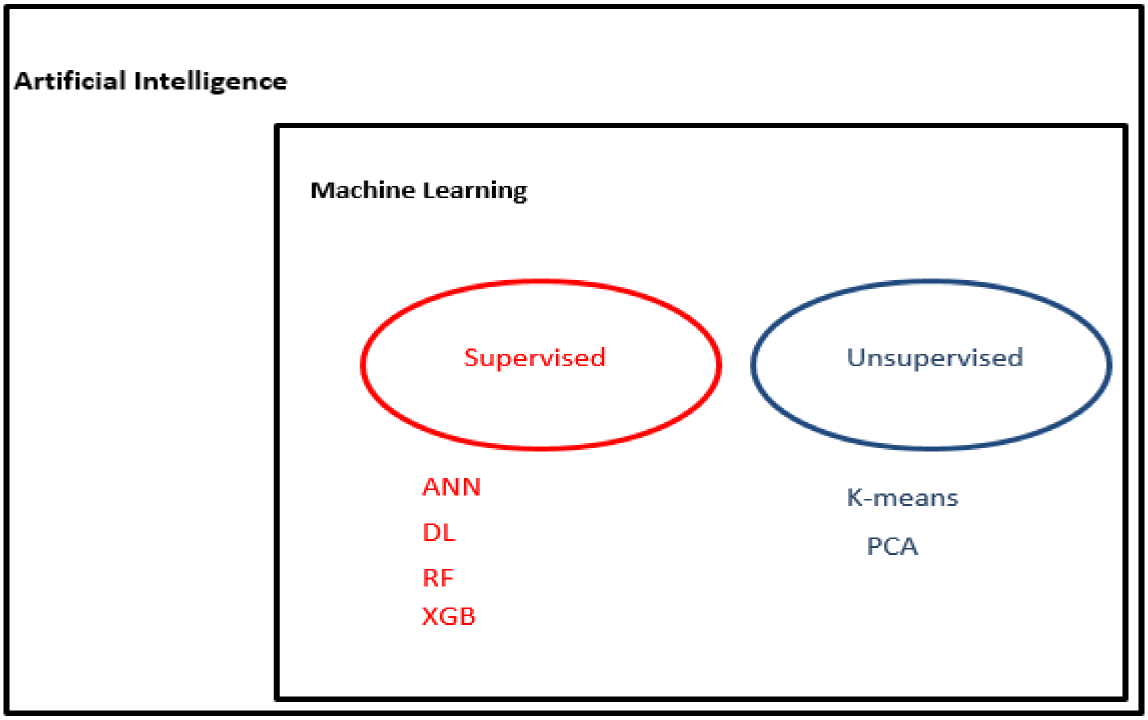

Fig 1. Artificial intelligence, machine learning, and their classification[3].

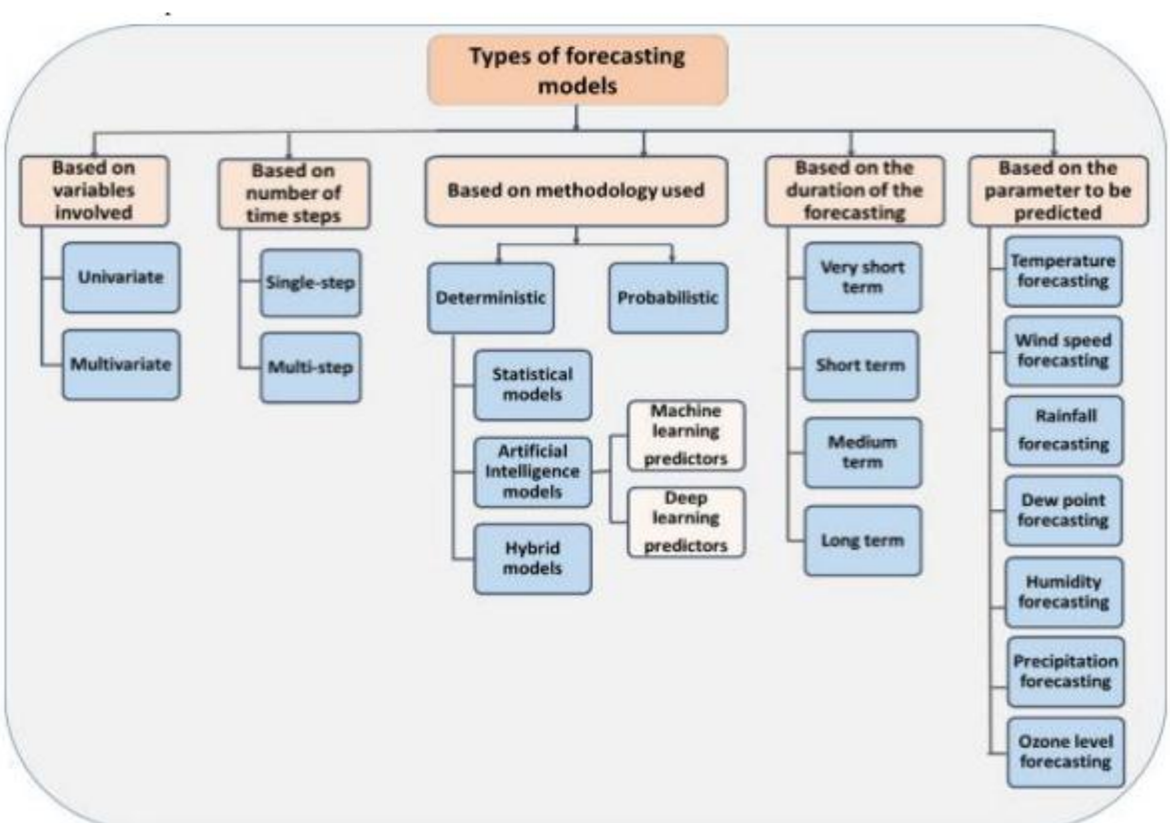

Fig 2. Types of weather forecasting[7]

II. ARTIFICIAL INTELLIGENCE MODELS

An increase in self-learning models has been prompted by the development of increasingly powerful AI. This one performs better than statistical models. Artificial intelligence models are more accurate at making predictions because of how well they handle nonlinear data. Deep and machine learning predictors [8].

2.1 Machine Learning Prediction Models

Deep learning and machine learning are more suitable for dealing with nonlinear datasets. Machine learning frequently uses the predictors SVM, ANN, ELM, and Random Forests. These indicators can also be utilized to forecast the weather. This section presents different machine learning tactics. There are models based on ML, ANN, and SVM [9].

2.1.1. Artificial Neural Network Model (ANN)

determine connections between present and future weather patterns, they are a widely used method of forecasting. It's a fast-learning nonlinear regression model that can map inputs to outputs and analyze data in a streamlined fashion. Connected nodes represent individual neurons in an Artificial Neural Network. A neural network has an input layer, a hidden layer, and an output layer. The input layer node performs some sort of transformation on the data before passing it on to the next node. The activation function is applied to the evaluated value at each node, transforming the overall result. The input layer receives the values to be altered, which then pass through the hidden levels and out to the output layers. Here in, many ANN-based prediction models are constructed as shoe in Figure 3 [7].

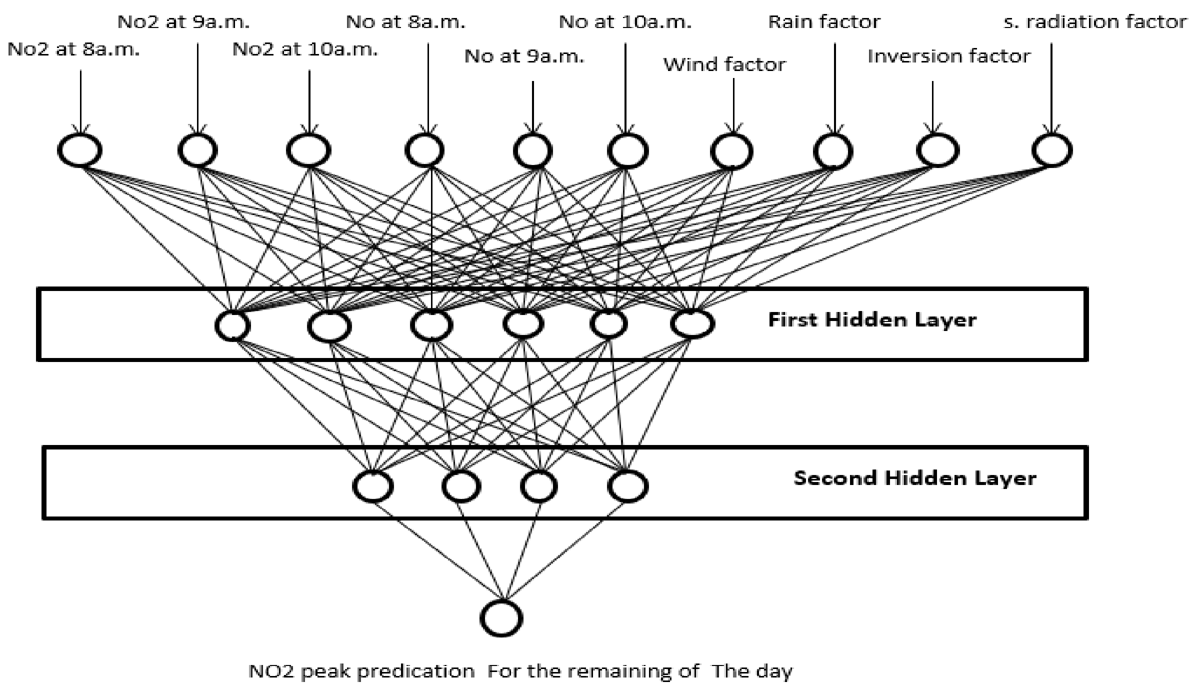

Fig 3. ANN model [10].

2.1.2. SVM based models

A supervised machine learning technique, SVM may be applied to both classification and regression tasks. The minimal distance hyper plane is used to define SVM, which is then used to sample training data. Maximum margin in training data is a hallmark of a good hyper plane. By employing the "kernel technique," SVMs are able to identify nonlinear solutions and they also perform well by high -dimensional data. Mathematical of functions known as "kernels" transform data into the required form. Linear, nonlinear, polynomial, and sigmoid functions are all solvable by SVM algorithms [11].

2.1.3. Decision Tree

Similarly, to a flowchart, a Decision Tree organizes possible outcomes in a hierarchical fashion. Attribute tests are represented via internal nodes. The test's results are displayed as branches. Class membership is represented by leaf nodes[9].

As opposed to abstract mathematical formulae, the findings may be readily understood thanks to the decision tree structure's provision of an explicit set of "if-then" rules. Leaves in the tree diagrams stand for different categories, while the branches stand for the many combinations of characteristics that together form those categories. A decision tree is a useful tool in decision analysis because it provides a clear and concise representation of the decision-making process. At a node deep within the network, the splitting value is determined by using the idea of information gain. Most information is gained by selecting the splitting value that maximizes this factor. Entropy provides a formal definition of information gain. Various methods, like as boosting and pruning, have been developed to enhance the performance of classification and regression trees by increasing their precision and adaptability.

By repeatedly applying the function and merging the results of each application with weighting to reduce the overall prediction error, or by simultaneously developing a number of independent trees and integrating them once all the trees have been generated, "boosting" is a technique for improving the accuracy of a predictive function. Overfitting, when the model starts to fit noise in the data, is an issue in big, single-tree models, and may be mitigated by tree pruning. A non-generalizable model is one that fails to produce accurate results when applied to data that was not used to train the model. There are a variety of decision tree algorithms available, such as the Alternating Decision Tree, Logitboost Alternating Decision Tree (LAD), C4.5, and the Classification and Regression Tree (CART)[12]. Figure 4 show decision tree.

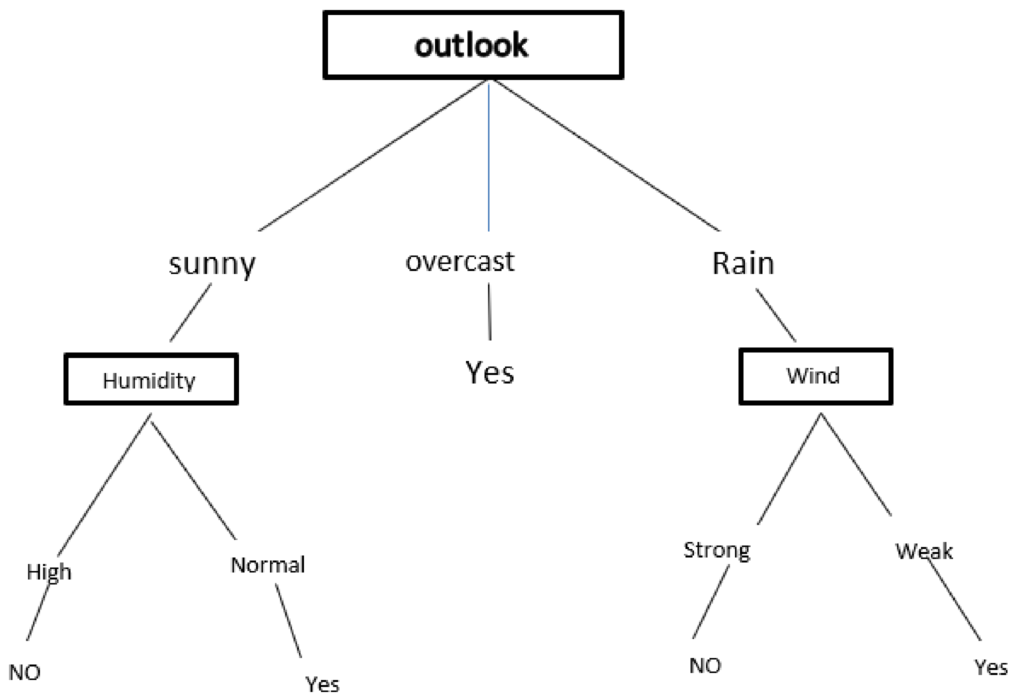

Fig 4. decision tree model [10].

2.1.4. Deep Learning Predictors

Deep learning [13], a subfield of machine learning algorithms, takes a neural approach to learning, employing supervised and unsupervised learning in deep structures to construct hierarchical representations of data. Speech recognition, time series analysis, and genomics are just few of the domains where deep learning has been successfully used. With their superior capacity to understand the temporal dependency inherent in time series data, deep learning approaches provide a more solid foundation for time series prediction. Microsoft, IBM, Google, and Facebook are just a few of the major IT firms actively researching this area [8].

III. Weather Forecasting And Its Types

Global crowds of onlookers and professional meteorologists collaborate on a daily basis to provide an accurate forecast. In reality, meteorologists use a variety of approaches. Because they are:

A- **Persistence Forecasting**:- It's the simplest method for trying to guess the weather. Forecasting tomorrow's weather is impossible without knowing the current weather.

Predictions about the weather will be vetted by meteorologists.

B- **Synoptic Forecasting**: - This method is based on fundamental gauging concepts. To make an instantaneous forecast, meteorologists read reports from the public and use those reports as criteria.

C- **Statistical Forecasting**: - Forecasters can get an idea of what the weather "should resemble" at any given moment thanks to the records kept by meteorologists of average temperatures, normal precipitation, and typical snowfall throughout the years.

D- **Computer Forecasting**: Forecasters add their own prejudices and presumptions to the hazy situation . In order to create computer "models," a few powerful computers replicate these scenarios. These computer "models" are then used to test a hypothesis over the course of the following several days. In addition, meteorologists should always use the other anticipated methods in addition to this kind of forecasting technology since often unique situations produce different results. Using the aforementioned methods, meteorologists come up with their "best theory" for what the next weather could be like and use it to make predictions.

IV . Conclusion

This paper explains different weather prediction techniques and methodologies. Based on the results, the deep learning algorithm provides better weather predictions. Unsupervised and supervised algorithms suffers from time constraints in issues with verifications and learning. FPGgrowth algorithm is expensive in built and suffers from memory issues.

REFERENCES

[1]N. Jaiswal and B. O. F. Technology, "A MACHINE LEARNING MODEL FOR WEATHER FORECASTING," no. July, 2021.
[2]M. Abdaki, A. Al-Iraqi, and R. M. Faisal, "Predicting long-term climate changes in Iraq," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 779, no. 1, 2021, doi: 10.1088/1755-1315/779/1/012053.
[3]N. Guyennon, F. Salerno, D. Rossi, M. Rainaldi, E. Calizza, and E. Romano, "Climate change and water abstraction impacts on the long-term variability of water levels in Lake Bracciano (Central Italy): A Random Forest approach," *J. Hydrol. Reg. Stud.*, vol. 37, no. April, p. 100880, 2021, doi: 10.1016/j.ejrh.2021.100880.
[4]S. M. M. I. Diaz E. F. Combarro et al, "Machine Learning Applied to Weather Forecasting," *Springer*, vol. 15, no. 2, p. 99999, 2017.
[5]N. Anusha, M. Sai Chaithanya, and G. Jithendranath Reddy, "Weather Prediction Using Multi Linear Regression Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 590, no. 1, 2019, doi: 10.1088/1757-899X/590/1/012034.
[6]K. Ramesh and R. Anitha, "MARSpline model for lead seven-day maximum and minimum air temperature prediction in Chennai, India," *J. Earth Syst. Sci.*, vol. 123, no. 4, pp. 665–672, 2014, doi: 10.1007/s12040-014-0434-z.
[7]K. U. Jaseena and B. C. Kovoor, *Deterministic weather forecasting models based on intelligent predictors: A survey*, vol. 34, no. 6. King Saud University, 2022. doi: 10.1016/j.jksuci.2020.09.009.
[8]G. M. Dharsan, "A survey on weather forecasting and their techniques," no. 02, pp. 12–17, 2022.
[9]M. Q. Mohammed and J. M. Al-Tuwaijari, "A Survey on various Machine Learning Approaches for thalassemia detection and classification," *Turkish J. Comput.* ..., no. xxxx, 2021, [Online]. Available: https://turcomat.org/index.php/turkbilmat/article/view/11284%0Ahttps://turcomat.org/index.php/turkbilmat/article/download/11284/8350
[10]G. K. Kang, J. Z. Gao, S. Chiao, S. Lu, and G. Xie, "Air Quality Prediction: Big Data and Machine Learning Approaches," *Int. J. Environ. Sci. Dev.*, vol. 9, no. 1, pp. 8–16, 2018, doi: 10.18178/ijesd.2018.9.1.1066.
[11]C. K. , et. al., "Prediction of Climate Change using SVM and Naïve Bayes Machine Learning Algorithms," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 2, pp. 2134–2139, 2021, doi: 10.17762/turcomat.v12i2.1856.
[12]F. Olaiya and A. B. Adeyemo, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies," *Int. J. Inf. Eng. Electron. Bus.*, vol. 4, no. 1, pp. 51–59, 2012, doi: 10.5815/ijieeb.2012.01.07.
[13]A. Wunsch, T. Liesch, and S. Broda, "Deep learning shows declining groundwater levels in Germany until 2100 due to climate change," *Nat. Commun.*, vol. 13, no. 1, pp. 1–13, 2022, doi: 10.1038/s41467-022-28770-2.