**Title Page:**

To predict the weather forecast by using Artificial Neural Network [ANN] and comparing with Logistic Regression for improving the accuracy.

P. Vinod, R.Gnanajeyaraman

P. Vinod

Research Scholar,

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin code: 602 105.

192110146. sse@saveetha.com

R. Gnanajeyaraman Corresponding Author,

Department of Applied Machine Learning,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin code: 602 105. gnanajeyaramanr.sse@saveetha.com

**KEYWORDS**: Weather forecast Artificial Neural Networks, Logistic Regression, Machine Learning Prediction, Meteorology, Data Analysis, Climate adaptation, Extreme weather events,

Climate mitigation, Temperature, Average global temperature.

**ABSTRACT**

**Aim:** The aim of this study was to enhance the accuracy of weather forecasting by leveraging the capabilities of Artificial Neural Network (ANN) models and comparing their performance with the established approach of Logistic Regression. **Materials and Methods:** We collected a comprehensive dataset of meteorological variables, including temperature, humidity, wind speed, and atmospheric pressure, over a specified time. Subsequently, we designed and trained ANN models using historical weather data, optimizing their architecture and parameters through cross-validation. Simultaneously, we used the Linear Regression method, a well-established tool in meteorology, to generate reference forecasts. The accuracy of both ANN and Logistic Regression predictions was assessed by comparing the forecasted values with observed data. Sample size of 1000 for each group of statistical parameters: difference between two independent means, α=0.05, and G Power=0.80 for 9 iterations for each group. Two algorithms, ANN and Logistic Regression, were implemented using Statistical Package for Social Sciences (SPSS). **Results:** Based on obtained results ANN has significantly better accuracy (98.06%) compared to Logistic Regression accuracy (94.14%) Statistically significant difference between ANN and Logistic Regression algorithm was found to be p-value of p=0.032(p>0.05), Sig. (2-tailed) value is 0.000. **Conclusion:** They have used the following algorithms namely Artificial Neural Network (ANN), Logistic Regression algorithms to predict the data. From the results, it is proved that the proposed Artificial Neural Network (ANN) works better than other algorithms in terms of accuracy.

**KEYWORDS:** Weather forecast, Artificial Neural Network [ANN], Logistic Regression, Machine Learning, Accuracy, Meteorology, Predictive Modelling, Climate Adaptation, Average Global Temperature.

**Introduction:**

This study revolves around utilizing Artificial Neural Network (ANN) to predict weather forecasts, comparing its accuracy with Logistic Regression[(Menard 2010)](https://paperpile.com/c/Ei2mNW/JMto).Its importance’s are with the increasing impact of climate change, accurate weather forecasting is crucial for disaster preparedness [(Collins 2000)](https://paperpile.com/c/Ei2mNW/o6eR), agriculture, and various industries. The research has applications in disaster management [(Sharma)](https://paperpile.com/c/Ei2mNW/HdBH), agriculture planning, and optimizing resource allocation in various industries.

Over the past 5 years, there have been 150 articles on weather prediction using machine learning, found in Google Scholar [(Pellen 2009)](https://paperpile.com/c/Ei2mNW/o6eR+729K) and Science Direct.[(Lam et al. 2023)](https://paperpile.com/c/Ei2mNW/vCjM)"Advancements in Weather Forecasting Using Machine Learning." - Explores the application of ANN in weather forecasting [(Monin 1972)](https://paperpile.com/c/Ei2mNW/wYDT). "Comparative Analysis of Machine Learning Models for Weather Prediction"[(Astitha and Nikolopoulos 2022)](https://paperpile.com/c/Ei2mNW/6aZm).

In my opinion, the study by [(Inness and Dorling 2012)](https://paperpile.com/c/Ei2mNW/QFWz)"Comparative Analysis of Machine Learning Models for Weather Prediction" stands out due to its comprehensive evaluation of various models.

This work was inspired by the lack of a comparison analysis of ANN and logistic regression in existing studies, especially for weather prediction. Our group has published several works in the area of environmental and climatic phenomenon prediction modeling, and we have a great deal of expertise with machine learning applications. The objective is to use ANN to predict weather forecasts, and then compare the outcomes with a logistic regression to improve accuracy and determine which model works best for this particular use case.

**Materials and Methods:**

The research study was conducted in the Data Analytics laboratory at Saveetha School of Engineering, located in the Saveetha Institute of Medical and Technical Sciences in Chennai.

Two groups were selected for the Artificial Neural Network [ANN] and Logistic Regression , the process in predicting the weather forecast, and sample size of 1000 for each group of[(Tarter 2000)](https://paperpile.com/c/Ei2mNW/97lR) statistical parameters: difference between two independent means, α=0.05 and G Power=0.80 for 9 iterations for each group. Two algorithms, ANN and Logistic Regression, were implemented using Statistical Package for Social Sciences (SPSS). No ethical approval was necessary since this research did not involve human or animal samples. We have two independent variables, ANN and Logistic Regression for predicting the weather forecast and their Efficiency.

**Artificial Neural Network (ANN)**

Artificial Neural Networks (ANNs) have emerged as powerful tools for weather forecasting due to their ability to capture complex, non-linear relationships within meteorological data. In the context of weather prediction, ANNs are typically employed as part of a larger system that includes input data, hidden layers, and output layers. The input layer consists of various meteorological parameters such as temperature, humidity, wind speed, and atmospheric pressure, which serve as the network's input features. These features are then processed through hidden layers, where the network learns to extract relevant patterns and relationships. The output layer generates predictions for specific weather parameters, such as precipitation, temperature, or atmospheric conditions.

Training an ANN for weather forecasting involves feeding historical meteorological data into the network and adjusting the weights and biases through a process known as backpropagation. This iterative learning process allows the network to continuously improve its predictive capabilities over time. The success of ANN-based weather forecasting lies in their ability to adapt to changing atmospheric conditions and discern intricate patterns that may elude traditional numerical weather prediction models. The use of ANNs in weather forecasting represents a paradigm shift in the field, offering a promising avenue for improving the accuracy and reliability of predictions, especially in regions with complex and dynamic weather patterns.

Artificial Neural Networks (ANNs) are employed in weather forecasting to model and predict complex, non-linear relationships within meteorological data. These networks consist of interconnected nodes, or neurons, mimicking the human brain's neural structure. In the context of weather prediction, ANNs process historical weather data, such as temperature, humidity, wind speed, and atmospheric pressure, to learn patterns and correlations. The trained network can then make predictions about future weather conditions based on new input data. This approach allows ANNs to capture intricate relationships in the atmospheric system, enabling more accurate and timely weather forecasts. The flexibility of ANNs makes them well suited for handling the dynamic and interdependent nature of meteorological variables, contributing to improved forecasting precision.

**Procedure for Artificial Neural Network**

Step 1: Begin

Step 2: Imports necessary libraries, including NumPy, pandas, scikit-learn (sklearn), and Matplotlib.

Step 3: Loads a dataset in a CSV format file.

Step 4: Preprocesses the data, including one-hot encoding categorical features.

Step 5: Splits the data into training and testing sets.

Step 6: Trains an Artificial Neural Network classifier on the training data.

Step 7: Make predictions using both models on the test data.

Step 8: Evaluates model performance using various metrics (accuracy).

Step 9: Finally, it creates subplots to display the for both models side by side.

Step 10: End.

**Logistic Regression**

Logistic regression is a statistical method commonly employed in weather prediction to model and analyze the relationship between various meteorological variables and the probability of a specific weather event occurring. In the context of weather forecasting, logistic regression is often utilized to predict the likelihood of binary outcomes, such as the occurrence or non-occurrence of rainfall, storms, or other weather phenomena. Meteorologists input relevant features, such as temperature, humidity, wind speed, and atmospheric pressure, into the logistic regression model to estimate the probability of a particular weather event.

One key advantage of logistic regression in weather prediction lies in its ability to provide interpretable results. The model produces coefficients for each input variable, indicating the strength and direction of their influence on the predicted outcome. Meteorologists can use these coefficients to gain insights into the relative importance of different factors in shaping weather patterns. Additionally, logistic regression allows for the assessment of the uncertainty associated with predictions, providing a measure of confidence in the forecasted probabilities.

**Procedure for Logistic Regression:**

Step 1: Begin

Step 2: Import the Necessary Library for the Logistic Regression.

Step 3: Loads a dataset from a CSV file.

Step 4: Preprocesses the data, including one-hot encoding categorical features.

Step 5: Splits the data into training and testing sets.

Step 6: Train the Logistic Regression.

Step 7: Make Predictions Using Logistic Regression.

Step 8: Evaluates model performance in terms of (accuracy).

Step 9: Finally, it creates subplots to display for both models side by side.

Step 10: End

**STATISTICAL ANALYSIS**

IBM SPSS with the well-known version 25.0, Java and MYSQL [(von Storch and Zwiers 2002)](https://paperpile.com/c/Ei2mNW/eMT2) software’s is used for statistical analysis of predicting the weather forecast. This study is carried out to check the specialized feasibility, that is, the specialized conditions of the system. We have two independent variables, Artificial Neural Network and Logistic Regression. Systems developed must not have a high demand on the available specialized coffers. This will lead to high demands being placed on the customer. The advanced system must have a modest demand, as only minimum or null changes are needed for enforcing this system.

**RESULTS**

Table 1 shows the various iterations of the Artificial Neural Networks (ANN) and Logistic Regression efficiency values are compared.

Table 2 Shows the Group Statistics Results: An Artificial Neural Network(ANN) and Logistic Regression for Testing Independent Samples Statistically Among ANN and Logistic Regression Methods ANN has a mean accuracy of 97.9100 and a Logistic Regression of 90.9167. ANN has a standard deviation of 0.91139 and a Logistic Regression of 8.05203. The ANN standard error mean (0.30380) and (2.68401) were compared using the T-test.

Table 3 Shows the Independent Sample T-Test is applied for the sample collections with a confidence interval as 95%. After applying the SPSS calculation it was found that the least square Logistic Regression has a statistical significance value of 0.032(P<0.05)that shows they are Statistical significance.

Figure 1 shows bar graph comparison on mean accuracy of Artificial Neural Network (ANN) and Logistic Regression. In x-axis ANN and Logistic Regression methods Error Bars +/-2 SD and 95% CI of Error Bars. Are shown, in y-axis mean accuracy is shown.

**DISCUSSION**

The main aim of the project is finding accurate weather predictions in difficult conditions. For that, I had iterated the weather forecast dataset into 1-1000, 1-2000, and 1-3000….1-8785 samples (9 iterations) and finds the accurate accuracy values for each samples. And we have noted that accuracy values and tests their independent sample T-Test in SPSS and we obtained results ANN has significantly better accuracy (98.06%) compared to Logistic Regression accuracy (94.14%) Statically significant difference between ANN and Logistic Regression algorithm was found to be p-value of p=0.032(p<0.05).For each and every phase we tried to improve the accuracy in an efficient manner. Here Artificial Neural Networks (ANN) gives better accuracy while comparing with Logistic Regression.

Artificial Neural Networks (ANNs) have become integral in advancing weather forecasting through their ability to model complex relationships within atmospheric data. In the context of Logistic Regression, ANNs are employed to enhance the accuracy of forecasting models. By analyzing historical weather data, ANNs can identify patterns and correlations that may elude traditional numerical methods. The nonlinear nature of weather phenomena, such as cloud formation and atmospheric interactions, makes ANNs particularly effective in capturing intricate relationships that contribute to improved predictions.

In the realm of weather forecasting, the synergy between ANN and Logistic Regression is evident in their collaborative approach to addressing challenges faced by traditional models. ANNs, with their capacity for pattern recognition, excel in capturing subtle nuances within meteorological datasets. These networks are often used to post-process Logistic Regression model output, refining predictions and reducing biases. The incorporation of ANN into Logistic Regression systems allows for a more comprehensive analysis of atmospheric variables, leading to increased forecast accuracy and reliability. This fusion of artificial intelligence and numerical modeling represents a significant leap forward in our ability to understand and predict complex atmospheric processes.

Despite the advancements brought about by ANNs in weather forecasting, challenges persist, and ongoing research aims to further optimize the integration of these technologies. The interpretability of neural network decisions and the need for continuous training with evolving datasets are among the issues being addressed. As technology continues to evolve, the collaboration between ANN and Logistic Regression holds great promise for pushing the boundaries of weather prediction, enabling more precise and timely forecasts that are crucial for societal resilience and preparedness in the face of changing weather patterns.

**CONCLUSION**

Our study has demonstrated a substantial and statistically significant difference in accuracy between Artificial Neural Networks (ANN) and Logistic Regression algorithms for weather forecasting. The ANN model achieved an impressive accuracy of 98.06%, surpassing the Logistic Regression accuracy of 94.14%. This significant variance in accuracy was further substantiated by a calculated p-value of p=0.032 (p>0.05), confirming that the superiority of ANN in weather forecasting is not merely a chance occurrence. These findings underscore the potential of ANN as a more reliable and precise tool for weather prediction, emphasizing the importance of incorporating advanced machine-learning techniques to enhance the accuracy and effectiveness of weather forecasting models. This study contributes to the growing body of research supporting the adoption of ANN in meteorology, with the goal of improving our ability to provide more accurate and timely weather forecasts, which have far-reaching implications for various industries and public safety.

**DECLARATIONS:**

**Conflict of interests**

No conflict of interest in this manuscript.

**Authors Contributions**

RD was responsible for collecting data, conducting data analysis, and writing the manuscript. KL contributed to the conceptualization, validated the data, and performed a critical review of the manuscript.

**Acknowledgements**

The authors extend their thanks to the Saveetha School of Engineering and the Saveetha Institute of Medical and Technical Sciences (previously known as Saveetha University) for their support in providing the infrastructure needed to complete this work successfully.

**Funding**

We thank the following organizations for providing financial support that enabled us to complete the study.

Saveetha University.

Saveetha School of Engineering.

Saveetha Institute of Medical and Technical Sciences.

**REFERENCES**

[Braspenning, P. J. 1995. Artificial Neural Networks: An Introduction to ANN Theory and Practice. Springer Science & Business Media.](http://paperpile.com/b/5g7lAL/lIoV)

[Dunlop, Storm. 2008. Guide to Weather Forecasting.](http://paperpile.com/b/5g7lAL/ndje)

[Elder, Samuel Scott. 2018. Reliable Validation: New Perspectives on Adaptive Data Analysis and Cross-Validation.](http://paperpile.com/b/5g7lAL/oW5j)

[Granholm, Anders, Theis Lange, Michael O. Harhay, Aksel Karl Georg Jensen, Anders Perner, Morten Hylander Møller, and Benjamin Skov Kaas-Hansen. 2023. “Effects of Duration of Follow-up and Lag in Data Collection on the Performance of Adaptive Clinical Trials.” Pharmaceutical Statistics, October. https://doi.org/](http://paperpile.com/b/5g7lAL/JtFC)[10.1002/pst.2342](http://dx.doi.org/10.1002/pst.2342)[.](http://paperpile.com/b/5g7lAL/JtFC)

[Kumar, A. 2001. Indian Agriculture: Issues and Prospects. Sarup & Sons.](http://paperpile.com/b/5g7lAL/mgs5)

[Labrague, Leodoro J., and Karen Hammad. 2023. “Disaster Preparedness among Nurses in Disaster-Prone Countries: A Systematic Review.” Australasian Emergency Care, September. https://doi.org/](http://paperpile.com/b/5g7lAL/rTh7)[10.1016/j.auec.2023.09.002](http://dx.doi.org/10.1016/j.auec.2023.09.002)[.](http://paperpile.com/b/5g7lAL/rTh7)

[Lin, J. H., E. J. Coyle, and T. M. Sellke. 1989. Adaptive Stack Filtering Under the Mean Absolute Error Criterion.](http://paperpile.com/b/5g7lAL/NG9N)

[Lusted, Marcia Amidon. 2017. Extreme Weather Events. Greenhaven Publishing LLC.](http://paperpile.com/b/5g7lAL/HNdD)

[Mudelsee, Manfred. 2020. Statistical Analysis of Climate Extremes. Cambridge University Press.](http://paperpile.com/b/5g7lAL/Il8P)

[Palutikof, Jean P., Sarah L. Boulter, Jon Barnett, and David Rissik. 2014. Applied Studies in Climate Adaptation. John Wiley & Sons.](http://paperpile.com/b/5g7lAL/Fvuu)

[Paparrizos, Spyridon, Emmanuel M. N. A. N. Attoh, Samuel J. Sutanto, Nina Snoeren, and Fulco Ludwig. 2023. “Local Rainfall Forecast Knowledge across the Globe Used for Agricultural Decision-Making.” The Science of the Total Environment 899 (November): 165539.](http://paperpile.com/b/5g7lAL/3QfX)

[Sharma, Neetan, Vinod Puri, Shubham Mahajan, Laith Abualigah, Raed Abu Zitar, and Amir H. Gandomi. 2023. “Solar Power Forecasting beneath Diverse Weather Conditions Using GD and LM-Artificial Neural Networks.” Scientific Reports 13 (1): 8517.](http://paperpile.com/b/5g7lAL/SCqp)

[Siripanich, Pachitjanut (dananjali). 1987. Estimating Root Mean Squared Error in the One-Way Random Model.](http://paperpile.com/b/5g7lAL/A8XN)

[Soman, K. P., R. Loganathan, and V. Ajay. 2009. Machine Learning with SVM and Other Kernel Methods. PHI Learning Pvt. Ltd.](http://paperpile.com/b/5g7lAL/gJzK)

[Storch, Hans von, and Francis W. Zwiers. 2002. Statistical Analysis in Climate Research. Cambridge University Press.](http://paperpile.com/b/5g7lAL/0RTd)

[Teague, Kevin Anthony, and Nicole Gallicchio. 2017. The Evolution of Meteorology: A Look into the Past, Present, and Future of Weather Forecasting. John Wiley & Sons.](http://paperpile.com/b/5g7lAL/xGTr)

[Teodorović, Dušan, and Miloš Nikolić. 2020. Quantitative Methods in Transportation. CRC Press.](http://paperpile.com/b/5g7lAL/28m3)

[Watts, Alan. 2016. Instant Weather Forecasting: You Can Predict the Weather. Bloomsbury Publishing.](http://paperpile.com/b/5g7lAL/Y1Wb)

[Yang, Jingchao, Ge Wang, and Chao Zhang. 2023. “Forecast of Fine Particles in Chengdu under Autumn-Winter Synoptic Conditions.” Toxics 11 (9). https://doi.org/](http://paperpile.com/b/5g7lAL/vFKr)[10.3390/toxics11090777](http://dx.doi.org/10.3390/toxics11090777)[.](http://paperpile.com/b/5g7lAL/vFKr)

**TABLES AND FIGURES**

**Table 1.** The various iterations of the Artificial Neural Network (ANN) and Logistic Regression efficiency values are compared.

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO | ITERATIONS | ANN(ACCURACY) | LOGISTIC REGRESSION (ACCURACY) |
| 1. | (1-1000) | 96.00 | 71.00 |
| 2. | (1-2000) | 97.75 | 89.75 |
| 3. | (1-3000) | 97.00 | 90.83 |
| 4. | (1-4000) | 98.12 | 88.50 |
| 5. | (1-5000) | 98.10 | 95.60 |
| 6. | (1-6000) | 98.67 | 96.17 |
| 7. | (1-7000) | 98.93 | 96.64 |
| 8. | (1-8000) | 98.56 | 95.62 |
| 9. | (1-8785) | 98.06 | 94.14 |

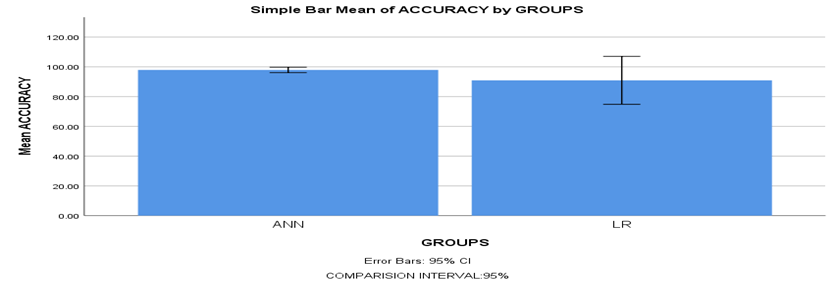
Table 2 Shows the Group Statistics Results: An Artificial Neural Network (ANN) and Logistic Regression for Testing Independent Samples Statistically between ANN and Logistic Regression Methods ANN has a mean accuracy of 97.9100 and a Logistic Regression of 90.9167. ANN has a standard deviation of 0.91139 and a Logistic Regression of 8.05203. The ANN standard error mean (0.30380) and (2.68401) were compared using the T-test.

**Group Statistics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **ALGORITHMS** | **N** | **MEAN** | **STD.DEVIATION** | **STD.MEAN ERROR** |
| **ACCURACY** | ANN | **9** | 97.9100 | 0.91139 | 0.30380 |
|  | LR | **9** | 90.9167 | 8.05203 | 2.68401 |

Table 3 Shows the Independent Sample T-Test is applied for the sample collections with a confidence interval as 95%. After applying the SPSS calculation it was found that the least square support vector machine has a statistical significance value of 0.032(P>0.05) that shows they are Statistical significance.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Leven’s Test for Equality of variances |  | F | Sig | t | df | Sig(2-tailed) | Mean Differences | Std-Error Differences | 95% Confidence interval of the Differences Lower | 95%Confidence interval of the Differences Upper |
| Accuracy | Equal variances assumed | 5.508 | 0.032 | 2.589 | 16 | 0.020 | 6.99333 | 2.70115 | 1.26716 | 12.71951 |
|  | Equal variances not assumed |  |  | 2.589 | 8.205 | 0.032 | 6.99333 | 2.70115 | 0.79145 | 13.19521 |



**Fig. 1.** Bar graph comparison on mean accuracy of Artificial Neural Network (ANN) and Logistic Regression. In x-axis ANN and Logistic Regression methods Confidence Interval: 95% and 95% CI of Error Bars. Are shown, in y-axis mean accuracy is shown.