TITLE PAGE:

Predicting the Accuracy of Global Mean Temperature using Decision tree compared with Adaboost

Krishna Kanth.V1, Fahad Iqbal2

Krishna Kanth.V1 Research Scholar,

Department of Information Technology, Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India. Pincode:602105

Email: krishnakant19@saveetha.com

Fahad Iqbal2

Project Guide,

Corresponding Author,

Department of Information Technology,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India. Pincode:602105

fahadiqbalkt.sse@saveetha.com

**Keywords:** Linear regression, Adaboost, OpenCV, accuracy, Anaconda

# ABSTRACT

**Aim:** The purpose of this work is to improve the prediction of the accuracy of global mean Temperature using Machine learning. **Materials and Methods:** Decision treeand Adaboost is executed with varying training and testing splits for predicting the accuracy of global mean Temperature using Machine learning. The Gpower test used is about 85% (g power setting parameters: α=0.05 and power=0.85). **Result:** Decision tree(86.0900%) has the increased accuracy over Adaboost (82.0500%) with a significance value of 0.008 (Two tailed, p>0.05). **Conclusion:** The accuracy of Decision treeis better when compared to accuracy of Adaboost.

Keywords: Linear regression, Adaboost, OpenCV, accuracy, Anaconda

# INTRODUCTION

We aim to use machine learning algorithms to model the mean temperature, namely Long-Short Term Memory Neural Networks (LSTM) and Random Forest Regressor (RF). For this study, we use a dataset with 30 years of radiosonde observations over the Brazilian region. In general, the results are consistent with those provided in the literature[(Brum et al. 2022)](https://paperpile.com/c/3AXnST/79pl),The proposed ensemble approach is based on three models which provide good performance in terms of model evaluation parameters like Correlation, Accuracy, R-Squared (R 2 ), Root mean square (RMSE) and Total Time to detect the predicted temperatures [(Himika et al. 2018)](https://paperpile.com/c/3AXnST/Ojlc).The purpose of this paper is to predict the most probable future global sea-level rise using advanced machine learning models. A total of 28 years' worth of sea-level rise data has been utilized for training our models using various machine learning algorithms[(Hassan et al. 2021)](https://paperpile.com/c/3AXnST/uFgL).Furthermore, the temporal and spatial variability in the modeled brightness temperatures via the SVM more closely agrees with that found in the original AMSR-E measurements. These findings suggest that the SVM is a superior alternative to the ANN for eventual use as a measurement operator within a data assimilation framework[(Forman and Reichle 2015)](https://paperpile.com/c/3AXnST/rhQq).

The research has been carried out onPredicting the accuracy of global mean Temperature using Machine learning; on an average of 44 research papers have been published in IEEE Xplore and 34 papers have been published in sciencedirect.The results were validated by in situ observations and compared with the NASA Advanced Microwave Scanning Radiometer for EOS (AMSR-E) snow water equivalent product. Satisfactory accuracy was achieved for different ecoregions with regard to daily, monthly, the Pearson correlation coefficient R ranged from 0.75 to 0.85)[(Xu et al. 2022)](https://paperpile.com/c/3AXnST/oZSx).

In this algorithm, the global precipitation measurement (GPM) product has been employed to train QPE prediction model. The real-time multiband infrared brightness temperature from Himawari-8, combined with the spatiotemporally matched numerical weather prediction (NWP) data from the global forecast system, have been used as predictor variables for QPE[(Min et al. 2019)](https://paperpile.com/c/3AXnST/e7jF). Considering that the sounding below the precipitation level becomes unreliable, the precipitation-affected observations were removed from the training dataset by means of a pre-screening test based on BT. The results show an overall ability of the algorithm to retrieve T and WV vertical profiles in line with expectations.[(Di Paola et al. 2018)](https://paperpile.com/c/3AXnST/SoBQ)

The research gap identified from the existing system is poor accuracy. This study is to improve the accuracy of classification by incorporating Decision treeand comparing its performance with Adaboost. The proposed model improves prediction of the accuracy of global mean Temperature using Machine learning.

# MATERIALS AND METHODS

The research work was done in the Soft Computing Lab, Department, college name. Sample size has been calculated using Gpower software by comparing both the controllers. Two groups are selected for comparing the process and their result is derived. In each group, 10 sets of samples and 10 samples in total are selected for this work. Two algorithms Decision treeand Adaboost are implemented using technical Analysis software. Sample size is determined as 10 for each group using GPower 3.1 software (gpower setting parameters: α=0.05 and power=0.85).

The proposed work is designed and implemented with the help of Python OpenCV software. The platform to assess deep learning was Windows 10 OS. Hardware configuration was an Intel core i7 processor with a RAM size of 4GB. System sort used was 64-bit. For implementation of code, java programming language was used. As for code execution, the dataset is worked behind to perform an output process for accuracy.

**Decision Tree:**

**Description:**

Decision Tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. A tree can be “learned” by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions. The construction of a decision tree classifier does not require any domain knowledge or parameter setting, and therefore is appropriate for exploratory knowledge discovery. Decision trees can handle high-dimensional data. In general decision tree classifier has good accuracy. Decision tree induction is a typical inductive approach to learn knowledge on classification.

**Pseudocode:**

GenDecTree(Sample S, Features F)

If stopping\_condition(S, F) = true then

a. Leaf = createNode()

b. leafLabel = classify(s)

C. return leaf

root = createNode()

root.test\_condition =findBestSpilt(S,F)

V = {v | v a possible outcomecfroot.test\_condition)

For each value v € V:

a. S1 = {s | root.test\_condition(s) = v and s €S};

b. Child = TreeGrowth (Sv,F);

C. Add child as descent of root and label the edge {root → child) as v

return root

**Adaboost**

**Algorithm:**

1. Assign equal weights to all the data points
2. Find the stump that does the best job classifying the new collection of samples by finding their Gini Index and selecting the one with the lowest Gini index
3. Calculate the “Amount of Say” and “Total error” to update the previous sample weights.
4. Normalize the new sample weights.

**Pseudocode:**

# Load data

iris = datasets.load\_iris()

X = iris.data

y = iris.target

# Split dataset into training set and test set

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3) # 70% training and 30% test

# Create adaboost classifer object

abc = AdaBoostClassifier(n\_estimators=50, learning\_rate=1)

# Train Adaboost Classifer

model = abc.fit(X\_train, y\_train)

#Predict the response for test dataset

y\_pred = model.predict(X\_test)

# Model Accuracy, how often is the classifier correct?

print("Accuracy:",metrics.accuracy\_score(y\_test, y\_pred))

# Statistical Analysis

SPSS software is used for statistical analysis of Decision treeand Adaboost. Independent variables are image, objects, distance, frequency, modulation, amplitude, volume, decibels. Dependent variables are images and objects. Independent T test analysis is carried out to calculate accuracy for both methods.

# RESULTS

The proposed Decision treeand Adaboost were run at different times in Anaconda Navigator with a sample size of 10. Table 1 represents the predicted accuracy and loss of Linear regression.Table 2 represents the predicted accuracy and loss of Adaboost. These 10 data samples are used for each algorithm along with their loss values to calculate statistical values that can be used for comparison. From the results, it is observed that the mean accuracy of Decision treewas 86.0900% and Adaboost was 82.0500%. Table 3 represents mean accuracy values for Decision treeand Adaboost. Mean value of Decision treeis better when compared with the Adaboost with a standard deviation of 2.92364 and 3.13909 respectively. Table 4 shows the Independent sample T test data of Decision treeand Adaboost with the significance value obtained is 0.781 (Two tailed, p<0.05). Figure 1 denotes the comparison of Decision treeand Adaboost in terms of mean accuracy and loss.

Mean, standard deviation and standard error mean for Decision treeare 82.0500, 2.92364 and .92454 respectively. Similarly for Adaboost, the mean, standard deviation and standard error mean are 82.0500, 3.13909 and .99267 respectively. On the other hand, the loss values of Decision treefor mean, standard deviation and standard error mean are 13.9100, 2.92364 and 0.92454 respectively. For Adaboost, the loss values of Adaboost for mean, standard deviation and standard error mean are 17.9500, 3.13909 and 0.99267 respectively.

The group statistics value along with mean, standard deviation and standard error mean for the two algorithms are also specified. The graphical representation of comparative analysis, means of loss between two algorithms of Decision treeand Adaboost are classified. This indicates that Decision treeis significantly better with 86.0900% accuracy when compared with Adaboost classified accuracy of 82.0500%.

# DISCUSSION

In the given study, the significance value obtained is 0.008 (Two tailed, p>0.05) which implies that Decision treeappears to be better than Adaboost. Accuracy analysis of the Decision treeis analyzed as 86.0900% whereas the accuracy of Decision treeis 82.0500%.

This paper presents ANN models to estimate daily global solar radiation (GSR) on a horizontal surface using meteorological variables: (mean daily extraterrestrial solar radiation intensity G 0, maximum possible sunshine hours S 0, relative humidity H, maximum air temperature T, atmospheric pressure P, and wind speed Vx) for Djelfa city in Algeria. Four input feature combinations are considered to determine how meteorological parameters affect daily global solar radiation prediction: 1) Day of the year, G0, S0, T, and Vx. 2) Day of the year, G0, S0, T, P, and Vx. 3) Day of the year, G0, S0, T, H, P, and Vx. 4) Day of the year, G0, S0, T, H, and Vx. MSE, MAE, and RMSE were used to compare these models (RMSE). Atmospheric pressure and relative humidity affect global solar radiation prediction. The relative humidity is also the most important predictor. Fourth model can forecast daily global solar radiation in other Algerian locations.[(Buchanan 2015)](https://paperpile.com/c/XWKhR8/kmBR)This paper presents temperature-based models for Ibadan's daily average global solar radiation. The International Institute of Tropical Agriculture (IITA) in Ibadan collected daily average global solar radiation, minimum and maximum daily average temperature over 9 years to develop the models. Comparing temperature model variants. Standard statistical tests—MBE, RMSE, MPE, and Correlation Coefficient—assess model suitability (R). The quadratic temperature model predicts Ibadan's global solar radiation best with MBE of 1.86 MJm -2 day -1, RMSE of 2.7, MAPE of 9.34%, and R of 0.68.[(Lima et al. 2022)](https://paperpile.com/c/XWKhR8/bxxr)This paper examines global temperature change using land-surface air temperature monthly means. After changing data sets, time series analysis worked well. We try several models before choosing ARIMA to analyse the data. Using the last 200 preserved temperature points, we test and refine our model. Finally, we predict 20 points of land-surface air temperature in 2019 and 2020. We discover that our planet is warming, so we must act to solve this terrifying global issue.[(Manabe and Broccoli 2020)](https://paperpile.com/c/XWKhR8/Da3j)This paper examines how mean weighted temperature (Tm) affects GPS-derived perceptible water vapour (PWV). Surface temperature data estimates Singapore's Tm (Ts). The estimated equation is used to calculate PWV for four consecutive months and compared to the well-known Bevis equation to determine Tm's effect. The results show that Tm and Ts relationships have little effect on PWV values.[(Houghton 2015)](https://paperpile.com/c/XWKhR8/dOIF)

The limitations of this study is that it takes a very long time to train Linear regression, especially with large datasets. The future scope of this study is that the system should be expanded to include a larger number of objects with lesser time consumption in training the data set.

# CONCLUSION

The accuracy value of the Decision treeis 86.0900% whereas the accuracy value of Adaboost is 82.0500%. Based on the analysis, Decision tree(86.0900%) performs better than Adaboost (82.0500%)

# DECLARATIONS

Conflicts of Interests

No conflict of interest in this manuscript.

# Authors Contribution

Author SA was involved in data collection, data analysis and manuscript writing. Author CPL was involved in conceptualization, data validation and critical reviews of manuscripts.

# Acknowledgement

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

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

1. Infysec Solution, Chennai
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

# REFERENCES

O. Assas, H. Bouzgou, S. Fetah, M. Salmi and A. Boursas, "Use of the artificial neural network and meteorological data for predicting daily global solar radiation in Djelfa, Algeria," 2014 International Conference on Composite Materials & Renewable Energy Applications (ICCMREA), 2014, pp. 1-5, doi: 10.1109/ICCMREA.2014.6843807.

T. R. Ayodele, A. S. O. Ogunjuyigbe, E. O. Oyediran and O. Ojo, "Temperature based model for estimating the daily average global solar irradiation of Ibadan, Nigeria," AFRICON 2015, 2015, pp. 1-5, doi: 10.1109/AFRCON.2015.7331971.

G. Naiqian, G. Yuxin and S. Xuelian, "Global Temperature Forecast Based on ARIMA Model," 2019 4th International Conference on Communication and Information Systems (ICCIS), 2019, pp. 108-112, doi: 10.1109/ICCIS49662.2019.00026.

S. Manandhar, Y. H. Lee and S. Winkler, "Mean weighted temperature for PWV estimation," 2015 IEEE 4th Asia-Pacific Conference on Antennas and Propagation (APCAP), 2015, pp. 437-438, doi: 10.1109/APCAP.2015.7374439.

[D. Brum, V. F. Rofatto, L. Gonzaga, R. De Oliveira Pena, L. F. Sapucci and M. R. Veronez, "Mean Tropospheric Temperature Estimation Using Deep Learning and Ensemble Methods," IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, 2022, pp. 6658-6661, doi: 10.1109/IGARSS46834.2022.9883387.](https://paperpile.com/c/3AXnST/79pl)

[Himika, S. Kaur and S. Randhawa, "Global Land Temperature Prediction by Machine Learning Combo Approach," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2018, pp. 1-8, doi: 10.1109/ICCCNT.2018.8494173.](https://paperpile.com/c/3AXnST/Ojlc)

[K. M. A. Hassan, M. A. Haque and S. Ahmed, "Comparative Study of Forecasting Global Mean Sea Level Rising using Machine Learning," 2021 International Conference on Electronics, Communications and Information Technology (ICECIT), 2021, pp. 1-4, doi: 10.1109/ICECIT54077.2021.9641339.](https://paperpile.com/c/3AXnST/uFgL)

[B. A. Forman and R. H. Reichle, "Using a Support Vector Machine and a Land Surface Model to Estimate Large-Scale Passive Microwave Brightness Temperatures Over Snow-Covered Land in North America," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 9, pp. 4431-4441, Sept. 2015, doi: 10.1109/JSTARS.2014.2325780.](https://paperpile.com/c/3AXnST/rhQq)

[X. Xu, X. Liu, X. Li, Q. Shi, Y. Chen and B. Ai, "Global Snow Depth Retrieval From Passive Microwave Brightness Temperature With Machine Learning Approach," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-17, 2022, Art no. 4302917, doi: 10.1109/TGRS.2021.3127202.](https://paperpile.com/c/3AXnST/oZSx)

[M. Min et al., "Estimating Summertime Precipitation from Himawari-8 and Global Forecast System Based on Machine Learning," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 5, pp. 2557-2570, May 2019, doi: 10.1109/TGRS.2018.2874950.](https://paperpile.com/c/3AXnST/e7jF)

[F. Di Paola et al., "Retrieval of Temperature and Water Vapor Vertical Profile from ATMS Measurements with Random Forests Technique," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 6014-6017, doi: 10.1109/IGARSS.2018.8518198.](https://paperpile.com/c/3AXnST/SoBQ)

**TABLES AND FIGURES**

**Table1.** Accuracy and Loss Analysis of Decision tree

| **Iterations** | **Accuracy(%)** | **Loss(%)** |
| --- | --- | --- |
| 1 | 81.6 | 18.4 |
| 2 | 82.9 | 17.1 |
| 3 | 83.7 | 16.3 |
| 4 | 84.7 | 15.3 |
| 5 | 85.8 | 14.2 |
| 6 | 86.4 | 13.6 |
| 7 | 87.4 | 12.6 |
| 8 | 88.5 | 11.5 |
| 9 | 89.1 | 10.9 |
| 10 | 90.8 | 9.2 |

**Table2.** Accuracy and Loss Analysis of Adaboost

| **Iterations** | **Accuracy(%)** | **Loss(%)** |
| --- | --- | --- |
| 1 | 77.3 | 22.7 |
| 2 | 78.4 | 21.6 |
| 3 | 79.6 | 20.4 |
| 4 | 80.7 | 19.3 |
| 5 | 81.5 | 18.5 |
| 6 | 82.1 | 17.9 |
| 7 | 83.9 | 16.1 |
| 8 | 84.7 | 15.3 |
| 9 | 85.6 | 14.4 |
| 10 | 86.7 | 13.3 |

**Table 3.** Group Statistical Analysis of Decision treeand Adaboost. Mean, Standard Deviation and Standard Error Mean are obtained for 10 samples. Decision treehas higher mean accuracy and lower mean loss when compared to Adaboost.

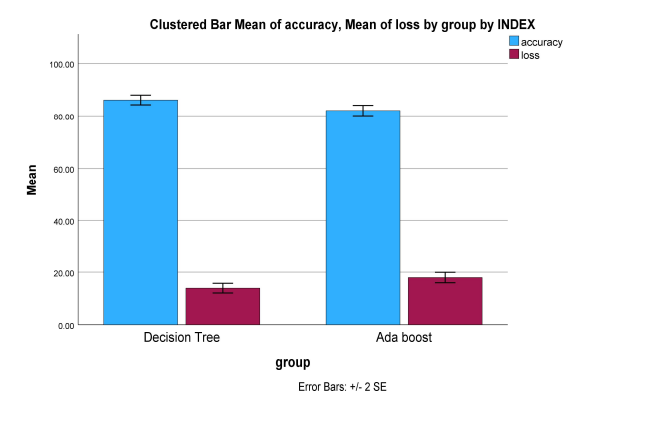
|  | **Group** | **N** | **Mean** | **Std. Deviation** | **Std. Error Mean** |
| --- | --- | --- | --- | --- | --- |
| **Accuracy** | Linear regression | 10 | 86.0900 | 2.92364 | .92454 |
| Adaboost | 10 | 82.0500 | 3.13909 | .99267 |
| **Loss** | Linear regression | 10 | 13.9100 | 2.92364 | 0.92454 |
| Adaboost | 10 | 17.9500 | 3.13909 | 0.99267 |

**Table 4.** Independent Sample T-test: Decision treeis insignificantly better than Adaboost with p value 0.008 (Two tailed, p<0.05)

|  | | **Levene’s test for equality of variances** | | **T-test for equality means with 95% confidence interval** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **f** | **Sig.** | **t** | **df** | **Sig. (2-tailed)** | **Mean difference** | **Std.Error difference** | **Lower** | **Upper** |
| **Accuracy** | **Equal variances assumed** | .079 | 0.781 | 2.978 | 18 | 0.008 | 4.04000 | 1.35652 | 1.19005 | 6.88995 |
| **Equal Variances not assumed** | 2.978 | 17.910 | 0.008 | 4.04000 | 1.35652 | 1.18902 | 6.89098 |
| **Loss** | **Equal variances assumed** | .079 | 0.781 | -2.978 | 18 | 0.008 | - 4.04000 | 1.35652 | -6.88995 | -1.19005 |
| **Equal Variances not assumed** | -2.978 | 17.910 | 0.008 | - 4.04000 | 1.35652 | -6.89098 | -1.18902 |

**Table 5.** Comparison of the Decision treeand Adaboost with their accuracy

| **CLASSIFIER** | **ACCURACY(%)** |
| --- | --- |
| **Decision tree** | 86.0900 |
| **Adaboost** | 82.0500 |



**Fig 1.** Comparison of Decision tree and Adaboost. Classifier in terms of mean accuracy and loss. The mean accuracy of Decision tree is better than Adaboost. Classifier; Standard deviation of Decision tree is slightly better than Adaboost. X Axis: Decision treeVs Adaboost Classifier and Y Axis: Mean accuracy of detection with +/-2SE.