TITLE PAGE:

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

Krishna Kanth. V1, Fahad Iqbal 2

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

krishnakanthv19@saveetha.com

Fahad Iqbal 2

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, Decision tree, 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:** Linear regression and Decision tree 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:** Linear regression (88.7590%) has the increased accuracy over Decision tree (86.0900%) with a significance value of 0.255 (Two tailed, p>0.05). **Conclusion:** The accuracy of Linear regression is better when compared to accuracy of Decision tree.

Keywords: Linear regression, Decision tree, 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 Linear regression and comparing its performance with Decision tree. 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 Linear regression and Decision tree 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.

**Linear regression**

**Description**

The machine learning algorithm Linear Regression is based on supervised learning. It does a task called regression. In regression, the independent variables are used to model a target prediction value. It is mostly used to figure out how different things relate to each other and make predictions.

**Pseudo code:**

Require: Training data D, number of epochs e, learning rate ŋ, standard deviation σ

Ensure: Weights ω0 , ω1,.....ωk

Initialise weights ω0 , ω1,.....ωk from standard normal distribution with

zero mean and standard deviation σ

for epoch in 1...e do

for each (x, y) **∈** D in random order do

**y ‘**←ω0 + Σki=1ωixi

if (**y ‘**>1 and y = 1) or (**y ‘**<-1 and y=-1) then continue

ω0← ω0 - η 2(**y ‘ -y)**

**For i in 1…..k do**

ωi← ωi - η 2(**y ‘ -y)**xi

end for

end for

return ω0 , ω1,.....ωk

**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

# Statistical Analysis

SPSS software is used for statistical analysis of Linear regression and Decision tree. 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 Linear regression and Decision tree 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 Decision tree. 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 Linear regression was 88.7590% and Decision tree was 86.0900%. Table 3 represents mean accuracy values for Linear regression and Decision tree. Mean value of Linear regression is better when compared with the Decision tree with a standard deviation of 6.55630 and 2.07329 respectively. Table 4 shows the Independent sample T test data of Linear regression and Decision tree with the significance value obtained is 0.011 (Two tailed, p<0.05). Figure 1 denotes the comparison of Linear regression and Decision tree in terms of mean accuracy and loss.

Mean, standard deviation and standard error mean for Linear regression are 88.7590, 6.55630 and 2.07329 respectively. Similarly for Decision tree, the mean, standard deviation and standard error mean are 86.0900, 2.07329 and .92454 respectively. On the other hand, the loss values of Linear regression for mean, standard deviation and standard error mean are 11.2890, 6.47235 and 2.04674 respectively. For Decision tree, the loss values of Decision tree for mean, standard deviation and standard error mean are 13.9100, 2.92364 and .92454 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 Linear regression and Decision tree are classified. This indicates that Linear regression is significantly better with 88.7590% accuracy when compared with Decision tree classified accuracy of 86.0900%.

# DISCUSSION

In the given study, the significance value obtained is 0.255 (Two tailed, p>0.05) which implies that Linear regression appears to be better than Decision tree. Accuracy analysis of the Linear regression is analyzed as 88.7590% whereas the accuracy of Linear regression is 86.0900%.

A study examined the relationship between INSAT 3D Land Surface Temperature (LST) and ground meteorological station Tair. Winter results show good correlation, but monsoon results decrease due to extreme temperatures and data unavailability. We found low root mean square error (RMSE) of ~1.5 °C for winter months and ~4.5 °C in June. LST and air temperature agree well, despite their different physical meanings and atmospheric responses.[(Leroux 2005)](https://paperpile.com/c/XWKhR8/sIYX)This paper examines a three-component temperature inversion scheme that uses airborne multiangle thermal infrared observations to reduce the difference between retrieved data and subpixel temperature distribution. The matrix of component effective emissivity, which links multiangular observations and component temperatures, is calculated using the FR97 model, an analytical directional brightness temperature model modified by dividing the soil component into sunlit and shaded portions. Simulations from the Scattering by Arbitrarily Inclined Leaves (4SAIL) model evaluate the new forward model and inversion scheme. The results show that the modified FR97 model's simplicity, accuracy, and low noise sensitivity make it suitable for inversion. Airborne data from the wide-angle infrared dual-mode line/area array scanner over maize fields and ground measurements from the Heihe Watershed Allied Telemetry Experimental Research campaign validate the inversion scheme. The leaf, sunlit soil, and shaded soil component temperatures had root mean square errors of 0.72 °C, 1.55 °C, and 2.73 °C, respectively.[(Singer and Avery 2007)](https://paperpile.com/c/XWKhR8/Morn)Radiosonde observations in Anqing from 2004 to 2007 determined the weighted mean temperature-surface temperature relationship. When the surface temperature rises, the weighted mean temperature estimated from the regressed formula is higher than that from the Bevis formula, resulting in a 1.9% difference in GPS precipitable water vapour at 40 °C and a smaller difference at lower temperatures. The GPS PWV differs by 0.72% at 40 °C and 2.17% at −10 °C due to regressions between years. The regressed formula to estimate the weighted mean temperature improves GPS PWV estimation in summer but not winter.[(Horne et al. 2016)](https://paperpile.com/c/XWKhR8/Zp4r)From 1982 to 2006, Shaanxi province's annual mean temperature and NDVI were examined using Climatic Research Unit (CRU) TS3.0 and NOAA/NDVI AVHRR's datasets. Temperature and NDVI increased. The correlation coefficient between temperature and NDVI was 0.534 (P<0.01), indicating a strong relationship between the two. NDVI was also positively correlated with the prolonged growing period, which is aided by rising temperatures.[(Mansinhos et al. 2022)](https://paperpile.com/c/XWKhR8/mGvd)

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 Linear regression is 88.7590% whereas the accuracy value of Decision tree is 86.0900%. Based on the analysis, Linear regression (88.7590%) performs better than Decision tree (86.0900%)

# 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

N. Doshi, T. Turakhia, A. S. Nair, M. Pandya and R. Iyer, "Estimating Air Temperature using Land Surface Temperature products of INSAT-3D satellite," 2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS), 2020, pp. 177-180, doi: 10.1109/InGARSS48198.2020.9358919.

Z. Bian et al., "Retrieval of Leaf, Sunlit Soil, and Shaded Soil Component Temperatures Using Airborne Thermal Infrared Multiangle Observations," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 8, pp. 4660-4671, Aug. 2016, doi: 10.1109/TGRS.2016.2547961.

Yunchang Cao, Feifei Zheng, Yifeng Xie and Yanmeng Bi, "Impact of the weighted mean temperature on the estimation of GPS precipitable water vapor," 2008 International Conference on Microwave and Millimeter Wave Technology, 2008, pp. 799-801, doi: 10.1109/ICMMT.2008.4540519.

Li Junyuan, Guo Jifeng and Xu Weixin, "Response of vegetation cover over Shaanxi province to global warming," 2011 International Symposium on Water Resource and Environmental Protection, 2011, pp. 2304-2306, doi: 10.1109/ISWREP.2011.5893727.

[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 Linear regression

| **Iterations** | **Accuracy(%)** | **Loss(%)** |
| --- | --- | --- |
| 1 | 88.4 | 11.6 |
| 2 | 89.7 | 10.3 |
| 3 | 90.8 | 9.2 |
| 4 | 91.01 | 8.99 |
| 5 | 92.1 | 7.9 |
| 6 | 91.8 | 8.2 |
| 7 | 93.6 | 6.4 |
| 8 | 94.8 | 5.2 |
| 9 | 95.4 | 4.6 |
| 10 | 96.06 | 3.94 |

**Table2.** 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 |

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

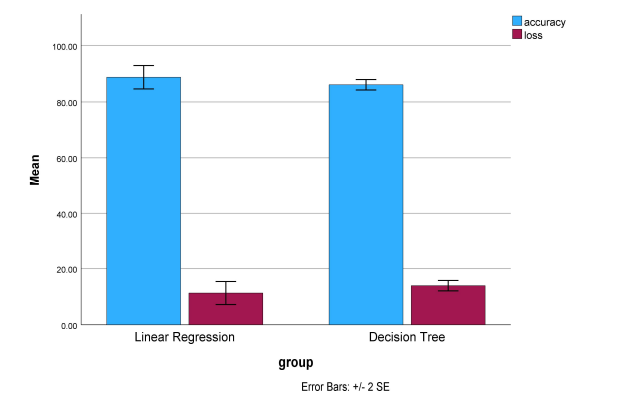
|  | **Group** | **N** | **Mean** | **Std. Deviation** | **Std. Error Mean** |
| --- | --- | --- | --- | --- | --- |
| **Accuracy** | Linear regression | 10 | 88.7590 | 6.55630 | 2.07329 |
| Decision tree | 10 | 86.0900 | 2.07329 | .92454 |
| **Loss** | Linear regression | 10 | 11.2890 | 6.47235 | 2.04674 |
| Decision tree | 10 | 13.9100 | 2.92364 | .92454 |

**Table 4.** Independent Sample T-test: Linear regression is insignificantly better than Decision tree with p value 0.255 (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** | 8.155 | 0.011 | 1.176 | 18 | 0.255 | 2.66900 | 2.27008 | -2.10027 | 7.43827 |
| **Equal Variances not assumed** | 1.176 | 12.443 | 0.255 | 2.66900 | 2.27008 | -2.25762 | 7.59562 |
| **Loss** | **Equal variances assumed** | 8.155 | 0.011 | - 1.176 | 18 | 0.255 | -2.66900 | 2.27008 | -7.43827 | 2.10027 |
| **Equal Variances not assumed** | - 1.176 | 12.443 | 0.255 | -2.66900 | 2.27008 | -7.59562 | 2.25762 |

**Table 5.** Comparison of the Linear regression and Decision tree with their accuracy

| **CLASSIFIER** | **ACCURACY(%)** |
| --- | --- |
| **Linear regression** | 88.7590 |
| **Decision tree** | 86.0900 |



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