AN ACCURATE MODEL FOR RAINFALL PREDICTION BASED ON SUPERVISED LEARNING

***Abstract* — Artificial Intelligence has been one of the most researched areas. Numerous applications and techniques are been widely used in our daily life using Artificial Intelligence. Machine Learning techniques play a vital role in numerous studies and applications. In this paper, we would be using supervised learning techniques for predicting rainfall. Numerous researchers have been working on it to explore new ways of predicting rainfall and obtain better accuracy when compared to existing systems. Two techiques were used namely Support Vector Machine (SVM) which is used to analyse and Artificial Neural Networks (ANN) is used for prediction rainfall with good accuracy. The model is designed in such a way that it fetches a sequence of images from a database including various other information about previous rainfalls in a specific area. The images are pre-processed and further segmented for feature extraction. The segmented images are then classified using neural networks. Given a study area, various parameters are evaluated such as its prediction time and efficiency and are examined to produce better accuracy when compared to the existing systems.**

Keywords— *SVM, ANN, Classification, Accuracy, Parameters, Prediction, Rainfall*

I. INTRODUCTION

One of the most dangerous climatic changes of nature is rainfall which could take away heavy loss of life and properties when the fall is heavy. It could also, in turn, lead to numerous other disasters such as floods and can cause severe damage to lives and properties. Some research that has seen a drastic growth in the era of digitization and has also in been an initiative for several new applications and techniques that are used globally by numerous people. The most common techniques which are widely deployed using artificial intelligence are Machine Learning techniques and Deep learning Approaches, which act as a base for numerous other applications being used. The medical field is one of the most prominent fields which drastically used various machine learning techniques for diagnosing various diseases[1]. Climatic changes are

observed using some of these techniques such as heavy rainfall and cyclone prediction would make a huge difference if predicted earlier[3]. Another such industry being used is robot Path Planning[2], as well as land cover classification[4] of a particular city, which is also considered as one of the most significant areas of research that used various machine learning techniques. Machine learning could be stated as a type of learning that is done without using any human intervention. The performance of any task by making use of this technique is improved by continuous learning. When it comes to learning, it is of two categories. The first category is Supervised learning[5] where the features of the dataset are given and are used for training purposes. The second category of learning is Unsupervised learning[6] where the learning [7] takes place without having any feature set. Most commonly neural networks are used for extracting the features from the dataset and then using them for further classification purposes. Feature extraction is considered to be a vital role in all the approaches as they are the inputs for the next processes and the final output could be obtained [8] [9] [10]. Classification and regression are some of the majorly done operations using machine learning techniques. During the time of usage of extremely large feature and dataset neural networks[11] are used for classification, as it produces high processing speed and accurate predictions and classifications Deep Leaning techniques come into picture when Artificial Neural Networks (ANNs) are used. They are intended to solve any kind of problems occurring in real-time scenarios and they also don't work using pre-defined algorithms.

II. RELATED WORK

Plenty of research works are being carried out in predicting rainfall with higher accuracy and less false rate prediction. The researcher is very keen on designing an efficient system with less computational time and providing higher accuracy for predicting rainfall. Numerous researchers are working on rainfall prediction out of which some of the prominent research articles [14],[15],[16].

Prediction of rainfall is done in [18], especially in Pondicherry using artificial neural networks. The author stated that the network was created using three various training algorithms. The measurement of the mean-squared error was done for each and every model. The accuracy of the model was evaluated and done by making use of the feed-forward time delay algorithm. Bhakar et al. have proposed a model for predicting rainfall in a particular area monthly using a series model [22]. J.Dou et.al has performed a survey on rainfall prediction[17]. Most of the works stated in the paper discussed predicting the rainfall using artificial neural networks. It also stated that by making use of some of the technologies like artificial neural networks was more efficient for rainfall prediction when compared to other existing traditional methods. Annual rainfall prediction was done in using the autoregressive

model(AR), which performed well compared to other existing systems in determining accurate predictions regarding rainfall when an entire year was considered[23]. Some of the simulation models are developed by Tantanee et al., which could predict long term rainfall[20]. Numerous attempts are being made on predicting the accurate rainfall prediction, for instance, marine and meteorological satellites, coupling physical satellite data, and usage of the forecasting model. In [19], the author has stated that the use of apparent characteristics in the rainfall series is often used by a stochastic process [21].

III. PROPOSED APPROACH

The proposed approach makes use of SVM and ANN classifier for classification of features and then predicting if there is a possibility of rainfall during a given time. In this research work, we have taken the Chennai rainfall dataset which comprises various features such as temperature, humidity, time, area and other details and also includes if there was rainfall or not. This file is first to read and is preprocessed to feed it as an input to the SVM and ANN classifier. The SVM classifier is used to give only two output whether it will rain or not. The features are categorized accordingly and then the output is generated when a new input is given from the learning it is done from the previous given data.

The dataset is also given to ANN where the features are not determined. ANN is usually used when a larger amount of data is used and features are not known. ANN extracts the features and learns in Fig. 1. Only 75% in the dataset is used for training purposes. The rest 25% is used for validation purposes. When the classifier is trained, new examples are given inputs from the validation set and validated for its accuracy.

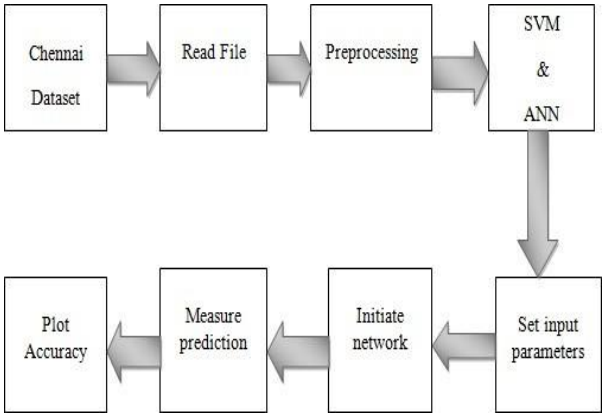

fig. 1 Block Diagram of Proposed Model

IV. EXPERIMENTAL RESULTS

The experimental results are obtained by training the SVM classifier and neural networks in Matlab R2018a. The dataset is initially divided into the ration of 75% and 25%. The initial 75% is used for training purposes and the rest for validation.

The Computer Vision toolbox is one of the toolboxes that are present in Matlab R2018a, that is actually used to create the proposed model. The model is generated by giving various parameters such as its number of iterations, the type of classifier to be used and so on after which the dataset is loaded into the classifier for evaluating the parameters.

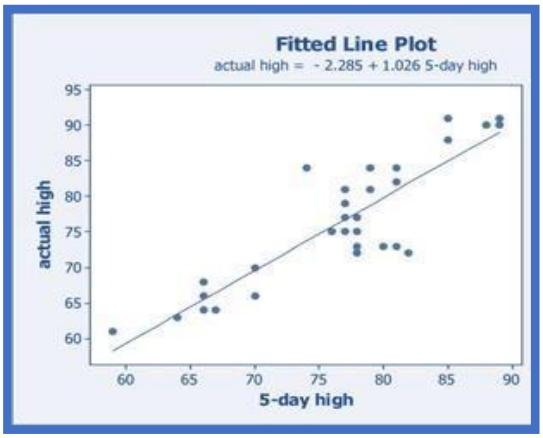

Fig. 2 SVM Classification

Firstly SVM classifier is used to predict the rainfall areas. In Fig. 2, is could be well seen that the segregation of the areas is done on the basis of rainfall. The areas with rainfall are categorized and then separated from the areas that are not affected by the rainfall. The values of predicting the rainfall using the generated model are the depicted in Fig. 3. The proposed model is highly efficient when compared with the other existing model in means of accuracy rate while predicting the occurrence of rainfall.

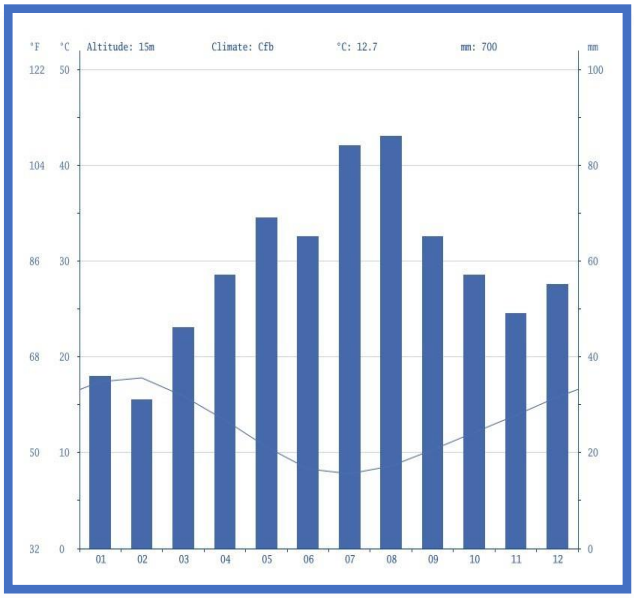

Fig. 3 Plot Analysis of Classifier

V. CONCLUSION

Rainfall is considered to be one of the most dangerous disasters that could ruin the daily livelihood of people and property damage when occurring at a massive rate. In order to preserve this destruction, rainfall prediction is one of the most important aspects, as it could avoid many life loss and property damage. In this paper, we have used the SVM classifier and ANN for predicting the rainfall. The Chennai dataset is given as the training dataset and is validated for evaluating its accuracy. The data set is categorized into two sets one for training and the other for validating. 75% is chosen for training purposes and the rest 25% is used for validating purposes. The proposed model is viewed to perform better than the existing models and provides higher accuracy of prediction rate with less

computational time. The future scope could include the use of other deep learning techniques.

REFERENCES:

1. Rahman, M. M., Bhattacharya, P., & Desai, B. C. (2007). A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. IEEE Transactions on Information Technology in Biomedicine, 11(1), 58-69.
2. Ireland, G., Volpi, M., & Petropoulos, G. P. (2015). Examining the capability of supervised machine learning classifiers in extracting flooded areas from Landsat TM imagery: a case study from a Mediterranean flood. Remote sensing, 7(3), 3372-3399.
3. Morales, M., Tapia, L., Pearce, R., Rodriguez, S., & Amato, N. M. (2004). A machine learning approach for feature-sensitive motion planning. In Algorithmic Foundations of Robotics VI(pp. 361-376). Springer, Berlin, Heidelberg.
4. Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. International Journal of remote sensing, 23(4), 725-749.
5. Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160, 3-24.
6. Cheriyadat, A. M. (2014). Unsupervised feature learning for aerial scene classification. IEEE Transactions on Geoscience and Remote Sensing, 52(1), 439-451.
7. Bauer, S., Nolte, L. P., & Reyes, M. (2011, September). Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 354-361). Springer, Berlin, Heidelberg.
8. Dumitru, D. (2009). Prediction of recurrent events in breast cancer using the Naive Bayesian classification. Annals of the University of Craiova-Mathematics and Computer Science Series, 36(2),92-96.
9. Liu, K., Tong, M., Xie, S., & Zeng, Z. (2014, August). Fusing decision trees based on genetic programming for classification of microarray datasets. In International Conference on Intelligent Computing (pp. 126-134). Springer, Cham.
10. Chou, Y. H., Tiu, C. M., Hung, G. S., Wu, S. C., Chang, T. Y., & Chiang, H. K. (2001). Stepwise logistic regression analysis of tumor contour features for breast ultrasound diagnosis. Ultrasound in medicine & biology, 27(11), 1493-1498.
11. Sarhan, A. M. (2009). Cancer classification based on microarray gene expression data using DCT and ANN. Journal of Theoretical & Applied Information Technology, 6(2).
12. Lorenc, A. C. (1986). Analysis methods for numerical weather prediction. Quarterly Journal of the Royal Meteorological Society, 112(474), 1177-1194.
13. Haupt, S. E., Cowie, J., Linden, S., McCandless, T., Kosovic, B., & Alessandrini, S. (2018, October). Machine Learning for Applied Weather Prediction. In 2018 IEEE 14th International Conference on e-Science (e-Science) (pp. 276-277). IEEE.
14. Flagg, D. D., Doyle, J. D., Holt, T. R., Tyndall, D. P., Amerault, C. M., Geiszler, D., ... & Eleuterio, D. P. (2018). On the Impact of Unmanned Aerial System Observations on Numerical Weather Prediction in the Coastal Zone. Monthly Weather Review, 146(2), 599-622.
15. Dubey, A.D. (2015). Artificial Neural Network Models for Rainfall Prediction in Pondicherry.

16. Bhakar SR, Singh RV, Neeraj C, Bansal AK (2006) Stochastic modeling of monthly rainfall at kota region. ARPN J Eng Appl Sci 1(3):36–44.
17. Dou, J., Yamagishi, H., Zhu, Z., Yunus, A. P., & Chen, C. W. (2018). TXT-tool 1.081-6.1 A Comparative Study of the Binary Logistic Regression (BLR) and Artificial Neural Network (ANN) Models for GIS-Based Spatial Predicting Landslides at a Regional Scale. In Landslide Dynamics: ISDR-ICL Landslide Interactive Teaching Tools (pp. 139-151). Springer, Cham.
18. Yevjevich V (1972) Stochastic processes in hydrology. Water Resour Pub, Colorado.
19. Tantanee S, Patamatammakul S, Oki T, Sriboonlue V, Prempree T (2005) Coupled wavelet-autoregressive model for annual rainfall prediction. J Environ Hydrol 13(18):1–8.
20. Chinchorkar SS, Patel GR, Sayyad FG (2012) Development of monsoon model for long range forecast rainfall explored for Anand (Gujarat-India). Int J Water Resour Environ Eng 4(11):322–326.
21. Hsu KL, Gupta HV, Sorooshian S (1995) Artificial neural network modeling of rainfall-runoff process. Water Resour Res 31(10):2517–2530.