An Efficient Hybrid Machine Learning Classifier for Rainfall Prediction

P. Asha¹*, A. Jesudoss²,S. Prince Mary³,K. V. Sai Sandeep⁴, K. Harsha Vardhan⁵

¹,²,³,⁴,⁵Sathyabama Institute of Science and Technology, India, Chennai-119

ashapandian225@gmail.com

Abstract. The most leading applications of Artificial Intelligence that seems to witness an immense Progression in the digital era are the Machine Learning (ML) Techniques. It learns itself from the past experiences and attempts at the best prediction of future instances or trends. Such progressive learning does not demand any explicit programming structures. Machine learning finds a wide range of application areas, and out of which accurate real time weather prediction gains importance. An interactive neural network based classification model for better prediction of rainfall has been put forth in this paper; we have proposed an interactive model for predicting rainfall using neural classification. The model is premeditated in a way, that it fetches feature extraction from a database including information about previous rainfalls in a specific area. The features were then pre-processed and further segmented by employing the random forest. The segmented outputs are then classified using neural networks. A comparison of spatial interpolation scheme is done with existing systems by deploying the hybrid classifier. The efficiency of the proposed model is evaluated and is compared with the traditional Deep Learning process and it is observed that the Random forest based interactive model provides better performance. Results of the model seem to be more accurate as the model uses an iterative approach for feature extraction.

1. Introduction

Artificial Intelligence (AI) has gained a drastic growth in the era of digital world and has also in turn given birth to several new applications and techniques that are being widely used in the day to day life. Some of the technologies being used widely are Machine Learning techniques and Deep learning approaches that are being widely used in numerous areas. When it comes to machine learning approaches, they are widely used in areas such as a medical field for diagnosing any particular type of tumor cells, or segmentation or classification of any type of disease or cell [1]. It is also been used in other industries such as robot path planning [2], prediction of climatic changes especially rainfall prediction and cyclone prediction. The occurrence of floods and its detection [3] as well as land cover classification [4] of a particular city also one of the prominent areas of research by making use of 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 types. One is supervised learning where the features of the dataset are given and are used for training purposes [5]. The second type of learning is Unsupervised learning where the learning takes place without having any feature set [6]. Here,
majorly, neural networks are employed for extricating the features from the dataset and then using them for further classification purposes. Feature extraction plays a vital role in all the approaches as they are the inputs for the next processes and the final output could be obtained. Some of the approaches being used are for classification and regression techniques. Some of the classifiers used predominantly in machine learning techniques are SVM classifiers [7], Decision Trees [8], Naive Bayes [9] and Random Forests [10]. When it comes to larger datasets, mostly neural classification is preferred due to its processing speed and accurate predictions and classifications [11]. Deep Learning techniques come into picture when Artificial Neural Networks (ANNs) are used. Here all the layer present in the network uses the output of the previous layer as its input process the features and then send it output as an input for the next layer. This is similar to the biological functioning of the neurons in the human body [12]. ANNs are considered to be an information processing structure that is majorly used for solving any kind of problems arising in real-time scenarios by learning though examples rather than using pre-defined algorithms [13]. In this paper, we could be using the Hybrid Neural Network, which issued for predicting rainfall in a particular study component. Rainfall is one of the major climatic changes that could lead to numerous other disasters such as floods and can cause severe damage to lives and properties. The rest of the paper consists of some related works proposed by various researchers, the methodologies and algorithms used in the proposed technique and experimental results.

2. Literature Survey

Weather prediction has become very popular nowadays and various researchers are working on it to make it more efficient. Some of the popular weather prediction researchers are stated in [14], [15], [16]. In [17], the author has performed a survey on rainfall prediction. 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. Another work [18] aimed at the prediction of rainfall, 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 every model. The accuracy of the model was evaluated and done by making use of the feed-forward time delay algorithm. The MSE value obtained was as low as 0.0083. Rainfall is considered to be a very complex atmospheric process as it is entirely dependent on time and space. Hence, it becomes one of the toughest ever prediction. In [19], the author has stated that the use of apparent characteristics in the rainfall series is often used by a stochastic process. When it comes to other applications that fully depend on resource planning purposes, the prediction of rainfall is one of the most important criteria that is to be done. In [20], a simulation model was developed, which could predict long term rainfall. 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. The inclusion of these techniques also includes the use of artificial neural networks and fuzzy logic were also proposed.

3. Proposed Methodology

Numerous feature extraction methods and classification techniques are used widely to identify and predict the rainfall for a specific location or a particular country's annual rainfall. Various classifiers, such as, SVM, K-means clustering and Random Forest Variables are used widely for the image segmentation applications that have been already discussed. In this paper, we majorly focus on the comparative analysis of various existing algorithms and their performance evaluation was computed by using an application-oriented model. The hybrid classifier designed in the current paper first extracts the important characteristic features from the database using a random forest. The outputs of the previous machine learning algorithm are directly inserted into neural networks. The validation of the dataset is done frame by frame for a particular period.
Fig. 1. System Architecture

Fig. 1 displays the architecture design of the entire proposed system. The dataset inputted consists of all the necessary attributes of the rainfall for the past 18 years. This is then fed to the Matlab tool where the processing takes place. Random forest classification is first applied to the dataset, which is then fed to Back propagation Neural Networks, where the errors are fed backward to the previous layers and weights and biases are adjusted until the required or appreciable result is retrieved.

The model includes the prediction of rainfall in a particular area (month wise) using a series model. Annual rainfall prediction was done using the autoregressive model (AR), which performed well compared to other existing systems in determining accurate predictions regarding rainfall when the entire year was considered.

4. Results and Discussion

The experimental results are obtained by training the neural networks in Matlab R2018. In Fig. 2, the neural network toolbox is shown where the rainfall dataset is being analyzed and is trained. The number of iterations being done is 8 and the performance of the neural network is being tested.

The regression plot for rainfall is shown in Fig. 3. This data is used for training the neural network. Training, Testing and Validations are depicted in Fig. 4. In, Fig. 5 the prediction is done which gives the output as highest rainfall for the month in a particular year and amount of rainfall in centimetre. This output with numerous parameters are being fed and trained to the network using the back propagation method to the classifier and hence it can predict the upcoming rainfall. The accuracy, precision and recall for the proposed classifier is shown in Table 1.
Fig. 2. Neural Network ToolBox

Fig. 3. Regression Plot of Rainfall
5. Conclusion

Weather prediction has gained rapid growth in the upcoming technical era. Machine learning techniques and Deep learning Techniques play a vital role in the prediction of weather related to a specific area or
a country for a period of time. The proposed interactive random forest based model, which is the Hybrid Neural Network model helped at predicting the annual rainfall of Chennai city for the next 365 days. The training time and the test time of the classifier are computed and the average accuracy of 100% is obtained. The pr designed Hybrid model serves better when contrasted with the existing models. Few Security issues pertaining in the work will be focused as a future enhancement towards the pertaining model.

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), Pp. 58-69.

[2] Asha P., Sravani B., SatyaPriya P. (2018), “Heart Block Recognition Using Image Processing and Back Propagation Neural Networks”, Soft Computing Systems, Communications in Computer and Information Science, Vol 837, Springer, Singapore.

[3] Ireland, G., Volpi, M., and 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), Pp. 3372-3399.

[4] Huang, C., Davis, L. S., and Townshend, J. R. G. (2002), “An assessment of support vector machines for land cover classification”, International Journal of remote sensing, 23(4), Pp. 725-749.

[5] Kotsiantis, S. B., Zaharakis, I., and Pintelas, P. (2007), “Supervised machine learning: A review of classification techniques”, Emerging Artificial Intelligence applications in Computer Engineering, 160, Pp. 3-24.

[6] D. S. and P. Asha, (2019), "Earlier Diagnosis and Survey of Diabetes Mellitus Using Machine Learning Techniques", Third International Conference on Inventive Systems and Control (ICISC), Pp. 37-38.

[7] Dumitru, D. (2009), “Prediction of recurrent events in breast cancer using the Naïve Bayesian classification”, Annals of the University of Craiova-Mathematics and Computer Science Series, 36(2), Pp. 92-96.

[8] Chou, Y. H., Tiu, C. M., Hung, G. S., Wu, S. C., Chang, T. Y., and Chiang, H. K. (2001), “Stepwise logistic regression analysis of tumor contour features for breast ultrasound diagnosis”, Ultrasound in Medicine and Biology, 27(11), Pp. 1493-1498.

[9] P. Asha, E. Vandana, E. Bhavana and K. R. Shankar, (2020), "Predicting University Dropout through Data Analysis," 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Pp. 852-856.

[10] Sarhan, A. M. (2009), “Cancer classification based on microarray gene expression data using DCT and ANN”, Journal of Theoretical & Applied Information Technology, 6(2).

[11] Steven K. Rogers, Dennis W. Ruck, Matthew Kabrisky, (1994), “Artificial Neural Networks for early detection and diagnosis of Cancer”, Elsevier Scientific Publishers, Pp. 79-83.

[12] Lorenc, A. C. (1986), “Analysis methods for Numerical Weather Prediction”, Quarterly Journal of the Royal Meteorological Society, 112(474), Pp. 1177-1194.

[13] Asha P., Lahari T., Kavya B. (2018), “Comprehensive Behaviour of Malware Detection Using the Machine Learning Classifier”, Soft Computing Systems, Communications in Computer and Information Science, Vol 837. Springer, Singapore.

[14] Haupt, S. E., Cowie, J., Linden, S., McCandless, T., Kosovic, B., and Alessandrini, S. (2018), “Machine Learning for Applied Weather Prediction”, IEEE 14th International Conference on E-Science, Pp. 276-277.

[15] P. Asha, P. H. Kiran Reddy, P. V. S. Manikanta and K. S. Sai (2020), "Cloud computing based Drug Delivery System in the Manufacturing of Drugs and Online Purchase," 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Pp. 847-851.
[16] Flagg, D. D., Doyle, J. D., Holt, T. R., Tyndall, D. P., Amerault, C. M., Geiszler, D., and 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), Pp. 599-622.

[17] Dou, J., Yamagishi, H., Zhu, Z., Yunus, A. P., and Chen, C. W. (2018), “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”, Landslide Dynamics: ISDR-ICL Landslide Interactive Teaching Tools, 1, Pp. 139-151.

[18] Dubey. (2015), "Artificial neural network models for rainfall prediction in Pondicherry", International Journal of Computer Applications, Vol. 120 (3), 2015.

[19] Chinchorkar, S. S., Patel, G. R., and Sayyad, F. G. (2012), “Development of monsoon model for long range forecast rainfall explored for Anand (Gujarat-India)”, International Journal of Water Resources and Environmental Engineering, 4(11), Pp. 322-326.

[20] Ramana, R. & B. Krishna & Kumar, S. & Pandey, N.. (2013), “Monthly Rainfall Prediction Using Wavelet Neural Network Analysis”, Water Resources Management. 27.