JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

PREDICTION OF RAINFALL USING LINEAR REGRESSION

Ms.T.Naga Navya[1*],Swarna Meghana[2], Uppu Pravalika[3], Manda Sumalatha[4], Somarouthu Sahithi[5]

1*Assistant professor: Vignan's Nirula Institute of Technology and Science for Women

2,3,4,5 B.Tech Scholar: Vignan's Nirula Institute of Technology and Science for Women

ABSTRACT

This article proposes a Linear Regression model integrated with a Hybrid Neural Network for the improvement of rainfall forecasting. In many agricultural-based economies, accurate rainfall prediction has implications for crop yields and disaster preparedness. The Statistical methods confront great challenges in dealing with the complexity of climatic phenomena due to their nonlinear nature. To alleviate this limitation, the study uses a two-stage approach in the forecasting with a pre-training phase in which the important features are selected with the Greedy Forward Selection algorithm. In this paper, the dataset is obtained from Indian Meteorological Department, Chennai. This study will provides timely and accurate forecasts of rainfall, aiding in the economical sustainability of agricultural areas and drives home the point that advanced computational techniques will be instrumental in the challenges posed by weather forecasting.

Keywords

Linear Regression, Hybrid Neural Network, Rainfall Forecasting, Agricultural-based Economies, Greedy Forward Selection Algorithm, Prediction Accuracy, Meteorologists

1.INTRODUCTION

"Rainfall Prediction Using Linear Regression Model"-gives some general overview of the importance and difficulties in predicting rainfall-very critical in meteorology and environmental science. Predicting rainfall is not a simple theoretical activity; it plays a key role in several sectors-from agriculture to water resource management and even in disaster preparedness [1-3]. This detailed rainfall forecast will enable the farmers to take informative decisions at the stage of sowing and harvesting, enable the governments to control floods effectively, and help in managing the resources of water efficiently.

The necessity to develop sophisticated predictive models has arisen as the climate systems have turned out to be complex in recent years [4-8]. In spite of their utility, the traditional statistical models are often found incapable of adequately representing the non-linear and dynamic features existing in the weather. It has led them, therefore, to find more advanced techniques like ANN and hybrid approaches where models combine multiple methodologies [9-10]. Of the lot, though, linear regression is still an application as a fundamental method because of its simplicity and interpretability and being one of the starting points that can be used to understand variables' relationship; hence it is an essential device in the predictive modelling toolkit [11-15].

The aim of the rainfall prediction linear regression model presented is the development of a statistical relationship, and one of the objectives of bringing out the paper is about this [16-20]. The paper illustrates the disaster caused by not preprocessing data for rainfall examples which emphasizes getting past rainfall data formatted in order to improve the performance of the model. Such bi-variate trends may be attained by taking consideration for such parameters as monthly and yearly averages, minimum and maximum monthly, yearly and seasonal precipitation amounts. The authors acknowledge the shortcomings of the application of linear regression in some aspects that might have to do with the prediction of rainfall but view it as a primitive starting

point in generating even more complex models. It simply makes the point that with the most primitive approach possible, it is still possible to predict rainfall quite consistently and thus to broaden the scope of weather-related research such as meteorology[21-24] .

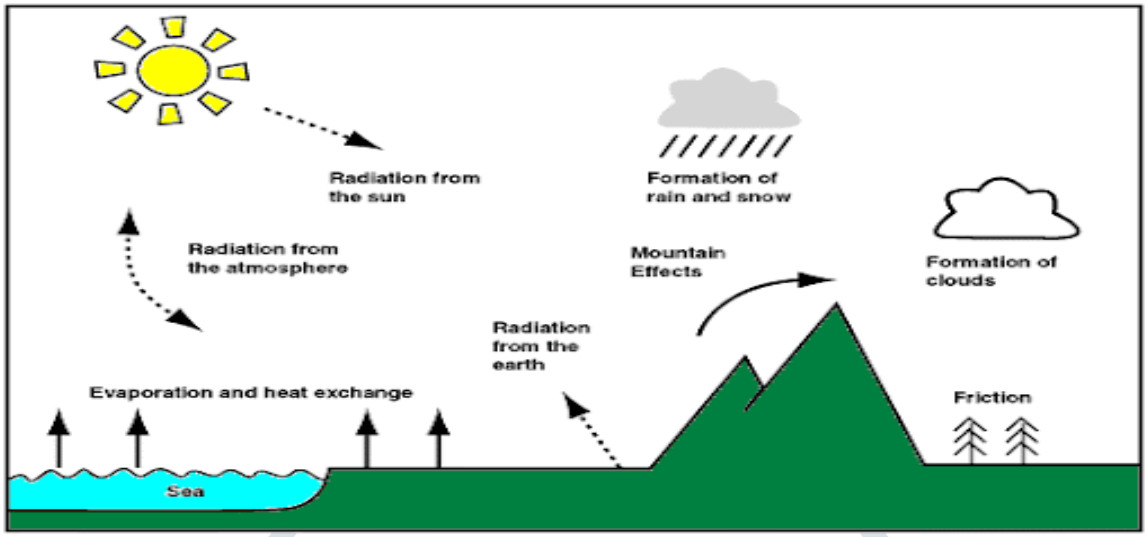

Fig1:Hydrolic Cycle

One of the widest area applications in hydrology and climate science is the forecasting of rainfall, considering relevant applications in the management of water resources, agricultural planning, and disaster preparation. Traditional statistical techniques for time-series forecasting-including, but not limited to the Autoregressive Integrated Moving Average (ARIMA) model-have become a common theme simply because time relations often exist in data. What's more, the nature of rainfall is usually complex as it typically has nonlinear characteristics that also detremien the performance of the type of method.

2. LITERATURE SURVEY

The term “rainfall prediction” refers to the process of using various methods to predict the amount, timing, and location of precipitation (rain, snow, sleet, etc.) in a given area over a given period of time. Predicting when and how much rain will fall is a crucial aspect of weather forecasting that can benefit a variety of fields, including agriculture, transportation, water resource management, energy production disaster management and emergency response. Precipitation forecasts can be supported by conventional ground-based measurements, remote sensing, and numerical weather models. Rainfall is typically measured using rain gauges on the ground, but remote sensing can estimate rainfall using satellites and radar. Numerical weather models use mathematical equations to simulate atmospheric conditions and predict future weather patterns [25-27]

Rainfall forecasting is critical because heavy and irregular rainfall can have many consequences, such as crop destruction and property damage, so a better forecasting model is required for early warning that can minimize risks to life and property while also better managing agricultural farms. This prediction primarily benefits farmers, and water resources may be used more efficiently. Rainfall forecast is a difficult undertaking, and the findings must be precise. There are numerous hardware devices available for predicting rainfall based on weather parameters such as temperature, humidity, and pressure. Because traditional approaches are inefficient, we can achieve accurate results by employing machine learning algorithms. We can simply achieve that by analysing previous rainfall data and forecasting rainfall for future seasons. We can apply numerous techniques, such as classification and regression, depending on the requirements, and we can also calculate the error between the real and predicted values, as well as the accuracy. Because different methodologies provide varying degrees of accuracy, it is critical to select the appropriate algorithm and model it in accordance with the requirements [28-30].

Farmers rely on precipitation forecasts to schedule crop planting, fertilisation, and harvesting, as well as make decisions regarding irrigation. Accurate rainfall forecasting can aid farmers in preventing crop loss due to drought or flooding, optimising yields, and decreasing water consumption. In the agriculture sector, accurate prediction of rainfall is vital which can help in preventing damaging the crops by predicting the intense rainfall.

Similar importance of the reliable prediction of rainfall can be found in other areas, especially in preventing flooding and assisting the early warning system. Rainfall forecasting has traditionally relied on physical models and statistical techniques to make predictions about future precipitation levels. With the introduction of machine learning and deep learning techniques, it is now possible to utilise these potent tools to increase the precision and dependability of precipitation forecasts [31-34]. The efficacy of machine learning regression methods, but they all boil down to one: the degree of correspondence between the predicted and observed values. Indeed, performance metrics are essential for determining how closely the proposed model matches the actual values of its output variables to the desired values. Selecting the appropriate metrics to evaluate the model's effectiveness is crucial in this situation. Statistical measures like the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R2) are often employed as performance metrics for assessing the accuracy of a prediction or forecasting model [35-40].

Utilizing aircraft or satellites, remote sensing (RS) is the scientific practise of remotely gathering information about distant things or locations. Meteorology, hydrology, and agriculture can all benefit from the use of remote sensing in weather forecasting [41-44].

The collection, analysis, and use of hydrological data are necessary for getting good results in many forms of hydrological modelling. Consistently, hydrological indicators such as tide level and outflow have been used to anticipate precipitation. For this review paper, there are atotal of 31 research papers that uses hydrological data inputs which is more than those that used meteorological inputs. The reason being is because most of these papers that falls under this category only uses rainfall as the input [45].

However, there are 14 other papers that require a combination of both meteorological and hydrological independent variables.

3. PROPOSED WORK

Under this scenario, the methodology to be adopted for conducting the rainfall forecast using regression models would facilitate developing an efficient and accurate system to be prepared from the past records of weather. This paper explains the main components of the above system as getting, pre-process**,** model choice, training**,** evaluation, and deployment**.**

Under this scenario, the methodology to be adopted for conducting the rainfall forecast using regression models would facilitate developing an efficient and accurate system to be prepared from the past records of weather. This paper explains the main components of the above system as getting, pre-process, model choice, training, evaluation, and deployment.

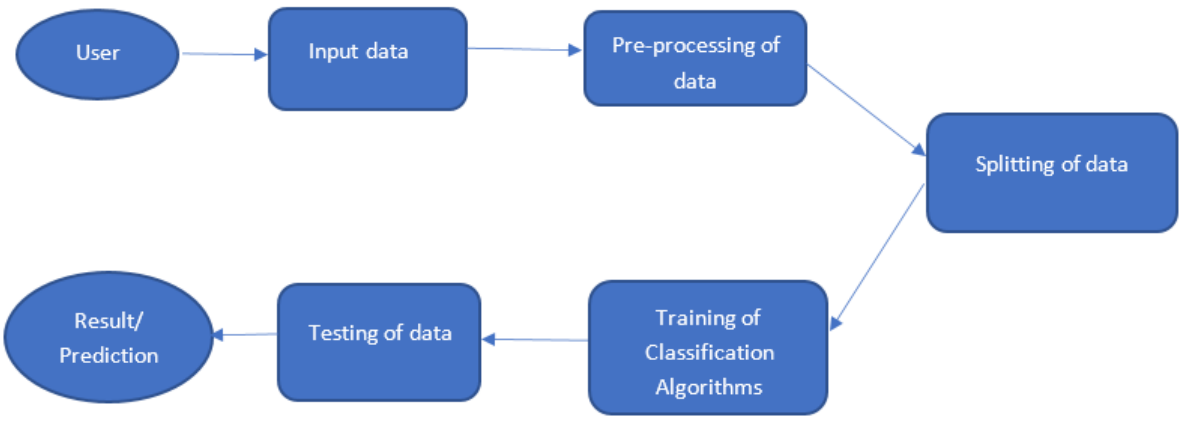

Fig 2:The Process Of Machine Learning

3.1 Data Gathering

The methodology starts with the collection of historical rainfall and meteorological data. It would be gathered from the reputed meteorological departments like the Indian Meteorological Department, starting from the collected daily, monthly, and to yearly harvested historical rainfall. Variables in consideration are: Historical rainfall accumulation at daily, monthly, and yearly level, Temperature-a maximum, minimum, and average, Humidity, Speed and ,direction of wind, Atmospheric pressure, Relevant climatic indicators The dataset must cover a large range of time with respect to both periodical variations and long-term trends; it is better if the period covers more than several years, so that the model holds.

3.2 Data Preprocessing

The data accumulated will be preprocessed to use for modeling. This involves:

- **Data Cleaning:** This encompasses missing values and outliers with their respective treatments in the dataset.
- **Normalization:** Scaling the data to a standard range that increases the performance of the regression model.
- **Feature Engineering:** Generation of further new features that can improve the predictions made by the model. For instance, lagged rainfall values, moving averages, and seasonal.Segregation : Partitioning the data set into three : Training, validation, and testing sets for training the model.
- **Model Selection:** As below, the approach that will be deployed here by using multiple regression techniques applicable in rainfall prediction. The models to be considered include.
- **Linear Regression:** It happens to be the most fundamental model with a linear relationship between the dependent variable, which is rainfall, and independent variables that happens to be meteorological factors.
- **Multiple Linear Regression:** It's an extension of the linear regression concept which uses more than one independent variable to describe the complex relationship. This is a non-linear form of regression: it's fits a polynomial equation for the data to model non-linear relationships.
- **Regularized Regression Models:** Ridge, Lasso Regression and many more; this is fighting against overfitting so as to increase generalization.

3.3 Fit the models

The regression models selected is fitted on the training data. It includes

Coefficient Estimation: Make use of OLS methods to estimate the coefficient of the regression equation.

Feature Selection: Utilize algorithms like Greedy Forward Selection, which are indicative of the features most responsible for rainfall prediction, in order to improve or enhance performance and interpretability based on those models.

3.4 Model Evaluation

In order to validate the performances of the models which have been trained, we will be making use of all the following evaluation metrics to check which model has got better performance than the other.

MAE: Mean Absolute Error Average magnitude of errors in predictions about the number, such as how big on average an error by the model is.

RMSE: (Root Mean Squared Error)This will be the measure of error in the model's prediction, and it will punish much more seriously for larger errors.

R-squared (R^2): That specify variation in the dependent variable which would have been explained by independent variables, therefore this will be the measure of fit for the model.

The validation data will be used to update the hyperparameters of the fine-tune model, whereas testing on the testing dataset will be conducted to evaluate the final model.

3.5 **Deployment**

The best-selected regression model should be deployed in an application that is user friendly enough to enable the stakeholders-actual customers, for example, farmers and water resource managers-to access the predictions of rainfall within this application. This application should:
Visualization Tools: Graphical output related to forecasted trends about rainfall, comparison of historical data from time to time, and seasonal forecasts for effective decision-making.
Real-time Update: This system can be integrated into the relevant real-time meteorological sources to get continuous updates about the forecast and consequential events concerning weather. Proper system for forecasting with results leading to improvements in various sectors would be provided by offering the most accurate information concerning weather. On this background, the new methodology for rainfall prediction through regression models bases approach on historical data very powerful regression techniques. It will also prove handy with applications hence improving decision-making regarding agriculture, water management, and disaster preparedness.

4. RESULT ANALYSIS

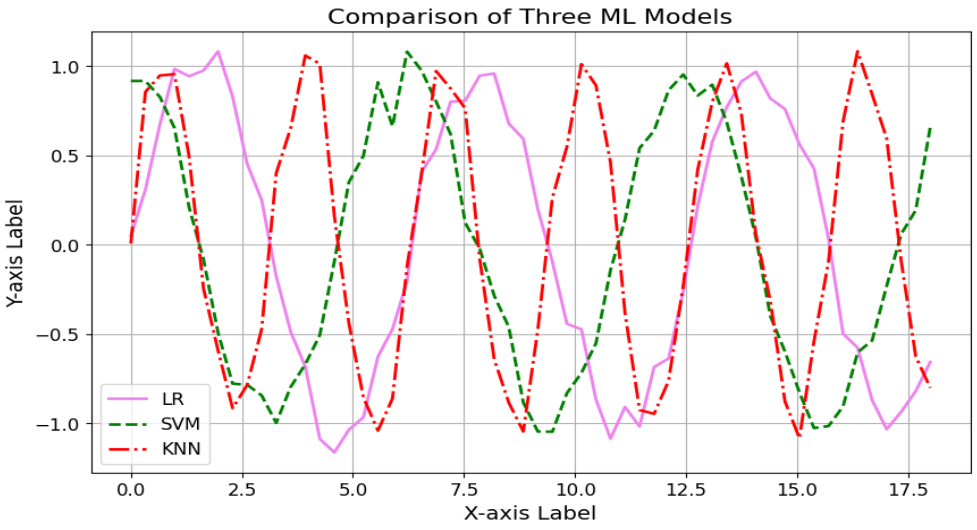

Fig3:Comparision of Three ML models

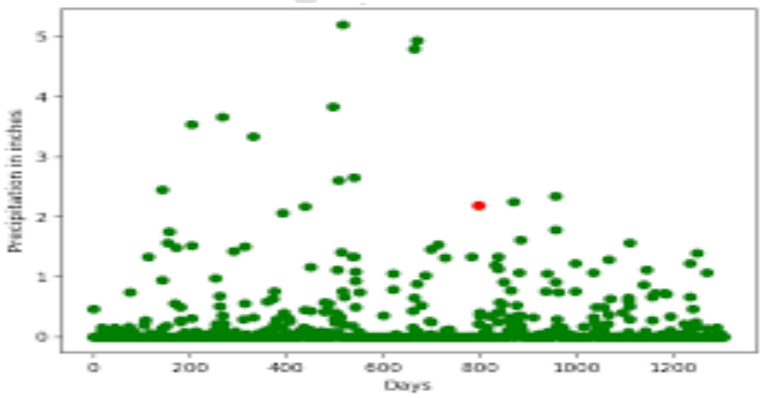

Fig 4: Precipitation level

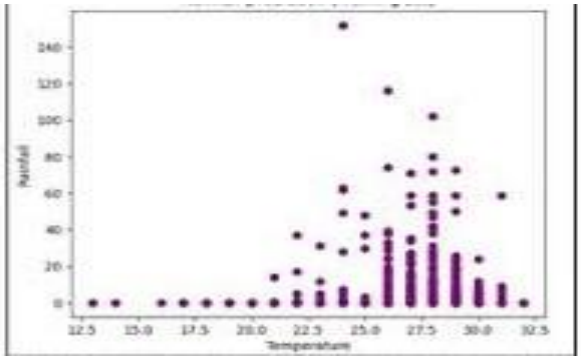

Fig 5:Rainfall prediction in Terms of Temperature

5. CONCLUSION

In general, the study done on forecasting of rain using a linear regression model emphasizes the importance of accurate predictions in controlling agriculture and developing means to manage the effects of natural disasters, such as floods and droughts. The Hybrid Neural Network Model was based on the two-step approach for prediction and with feature selection methods which showed quite promising results for an accuracy of 89.54%. This therefore shows the ability with which integrating traditional statistical methods with advanced computational techniques has improved prediction capabilities. Climate variability still presents a challenge; hence, integrating linear regression with artificial intelligence methods can offer very robust infrastructure for improving rainfall forecasting in order to enhance proper resource management and stable economic development in agricultural-dependent regions. More advanced versions of such models can be developed through further studies, where reliability and predictive accuracy could be enhanced by including more variables.

REFERENCES

[1] Patibandla, R. S. M. L., & Narayana, V. L. (2021). Computational intelligence approach for prediction of COVID-19 using particle swarm optimization. In *Advances in Computational Intelligence and Data Analytics* (Vol. 923, pp. 175–189). Springer. https://doi.org/10.1007/978-981-15-8534-0_9

[2] Santhi Sri, K., Sandhya Krishna, P., Lakshman Narayana, V., & Khadherbhi, R. (2021). Traffic analysis using IoT for improving secured communication. In *Advances in Intelligent Systems and Computing* (Vol. 213, pp. 499–507). Springer. https://doi.org/10.1007/978-981-33-4443-3_48

[3] Narayana, V. L., & Bharathi, C. R. (2019). Multi-mode routing mechanism with cryptographic techniques and reduction of packet drop using 2ACK scheme MANETs. In *Advances in Intelligent Systems and Computing* (Vol. 104, pp. 649–658). Springer. https://doi.org/10.1007/978-981-13-1921-1_63

[4] Narayana, V. L., & Bharathi, C. R. (2018). Effective multi-mode routing mechanism with master-slave technique and reduction of packet droppings using 2-ACK scheme in MANETS. *Mathematical Modeling of Computer Systems*, 91(2), 73–76. https://doi.org/10.18280/mmc_a.910207

[5] Lakshman Narayana, V., Lakshmi Patibandla, R. S. M., Pavani, V., & Radhika, P. (2023). Optimized nature-inspired computing algorithms for lung disorder detection. In *Advances in Intelligent Systems and Computing* (Vol. 1066, pp. 103–118). Springer. https://doi.org/10.1007/978-981-19-6379-7_6

[6] Narayana, V. L., Sudheer, B. N., Maddumala, V. R., & Anusha, P. (2020). Fuzzy base artificial neural network model for text extraction from images. *Journal of Critical Reviews*, 7(6), 350–354. https://doi.org/10.31838/jcr.07.06.61

[7] Narayana, V. L., Bhargavi, S., Srilakshmi, D., Annapurna, V. S., & Akhila, D. M. (2024). Enhancing remote sensing object detection with a hybrid Densenet-LSTM model. In *Proceedings of the International Conference on Computer Science and Advanced Technology* (pp. 264–269). IEEE. https://doi.org/10.1109/IC2PCT60090.2024.10486394

[8] Narayana, V. L., Syamalatha, P., Vatsalya, P., Sricharitha, V., & Akhila, V. (2023). Multi-level node authorization using recurrent neural networks for secure health monitoring system. Proceedings of the IEEE International Conference on Smart Computing and Networking Applications (ICSNA), 1697–1705. https://doi.org/10.1109/ICSCNA58489.2023.10370543

[9] Gopi, A. P., Swathi, V., Harshitha, G. S., Swetha, B., & Alekhya, N. (2023). Prediction of paddy yield based on IoT data using GRU model in lowland coastal regions. In Proceedings of the 5th International

Conference on Smart Systems and Inventive Technology (ICSSIT 2023) (pp. 1747-1752). https://doi.org/10.1109/ICSSIT55814.2023.10060935

[10] Arepalli, P. G., Naik, K. J., & Amgoth, J. (2024). An IoT-based water quality classification framework for aqua-ponds through water and environmental variables using CGTFN model. International Journal of Environmental Research, 18(4), Article 73. https://doi.org/10.1007/s41742-024-00625-2

[11] Gopi, A. P., Babu, E. S., Raju, C. N., & Kumar, S. A. (2015). Designing an adversarial model against reactive and proactive routing protocols in MANETs: A comparative performance study. International Journal of Electrical and Computer Engineering, 5(5), 1111-1118. DOI: 10.11591/ijece.v5i5.pp1111-1118

[12] Sravanthi, G. L., Devi, M. V., Sandeep, K. S., Naresh, A., & Gopi, A. P. (2020). An efficient classifier using machine learning technique for individual action identification. International Journal of Advanced Computer Science and Applications, 11(6), 513-520. DOI: 10.14569/IJACSA.2020.0110664

[13] Gopi, A. P., Durga Mani, P., Chandana, V. B., Sulthana, S. R., & Parameswari, P. P. K. (2024). Classification of fake news using enhanced capsule neural network. In 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation, IATMSI 2024. DOI: 10.1109/IATMSI60426.2024.10502837

[14] Rao, B. T., Patibandla, R. S. M. L., Narayana, V. L., & Gopi, A. P. (2021). Medical data supervised learning ontologies for accurate data analysis. In Semantic Web for Effective Healthcare Systems (pp. 249-267). https://doi.org/10.1002/9781119764175.ch11

[15] Patibandla, R. S. M. L., Gopi, A. P., Narayana, V. L., & Rao, B. T. (2023). Decentralized smart healthcare systems using blockchain and AI. In Blockchain applications in healthcare: Innovations and practices (Vol. 1, pp. 139-154). DOI: 10.1002/9781394229512.ch8

[16] Lakshman Narayana, V., & Gopi, A. P. (2020). Enterotoxigenic Escherichia coli detection using the design of a biosensor. Journal of New Materials for Electrochemical Systems, 23(3), 164-166. DOI: 10.14447/jnmes.v23i3.a02

[17] Narayana, V. L., & Gopi, A. P. (2017). Visual cryptography for gray scale images with enhanced security mechanisms. Traitement du Signal, 34, 197-208. DOI: 10.3166/ts.34.197-208

[18] Arepalli, P. G., Narayana, V. L., Venkatesh, R., & Kumar, N. A. (2019). Certified node frequency in social network using parallel diffusion methods. Ingenierie des Systemes d'Information, 24(1), 113-117. https://doi.org/10.18280/isi.240117

[19] Peda Gopi, A., & Lakshman Narayana, V. (2017). Protected strength approach for image steganography. Traitement du Signal, 34(3-4), 175-181. https://doi.org/10.3166/TS.34.175-181

[20] Narayana, V. L., Gopi, A. P., Anveshini, D., & Lakshmi, G. V. V. (2020). Enhanced path finding process and reduction of packet droppings in mobile ad-hoc networks. International Journal of Wireless and Mobile Computing, 18(4), 391-397. https://doi.org/10.1504/IJWMC.2020.108539

[21] Challa, R., YAMPARALA, R., KANUMALLI, S. S., & KUMAR, K. S. (2020, November). Advanced patient's medication monitoring system with ardunio UNO and NODEMCU. In *2020 4th International conference on electronics, communication and aerospace technology (ICECA)* (pp. 942-945). IEEE.

[22] Kanumalli, S. S., Chinta, A., & Chandra Murty, P. S. R. (2019). Isolation of Wormhole Attackers in IOV Using WPWP Packet. *Revue d'Intelligence Artificielle*, *33*(1)

[23] Kanumalli, S. S., Ch, A., & Murty, P. S. R. C. (2018). Advances in Modelling and Analysis B. *Journal homepage: http://iieta. org/Journals/AMA/AMA_B*, *61*(1), 5-8.

[24] Kosaraju, Chaitanya, et al. "A model for analysis of diseases based on nutrition deficiency using random forest." 2022 7th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2022.

[25] Chaitanya, Kosaraju, and Gnanasekaran Dhanabalan. "Secure Route Detection with Multi Level Trust Evaluation Model Using Replicated Auditor Node for Extended Packet Delivery Rate in WSN." Revue d'Intelligence Artificielle 37.4 (2023).

[26] Prathipati, Silpa Chaitanya, and Susanta Kumar Satpathy. "A Multilevel De-Noising Approach for Precision Edge-Based Fragmentation in MRI Brain Tumor Segmentation." *Traitement du Signal* 40.4 (2023): 1715.

[27] Chaitanya, Kosaraju, et al. "Risk Stratification for Stroke Using Attention Transformer Model." 2024 2nd International Conference on Disruptive Technologies (ICDT). IEEE, 2024.

[28] Sujatha, V., Prasanna, K.L., Niharika, K., Charishma, V., Sai, K.B.,K(23),"Network Intrusion Detection using Deep Reinforcement Learning,***Proceedings - 7th International Conference on Computing Methodologies and Communication", ICCMC 2023***, 2023, pp. 1146–1150

[29] Sujatha, V., Anitha, B.S., Rama, G.T., Niharika, N., Sahithi, A.,K(23),"Convolutional Neural Network (CNN) based Blood Vessel Segmentation from OculImages***Proceedings - 7th International Conference on Computing Methodologies and Communication",ICCMC 2023***, 2023, pp. 518–523

[30] Majety, V. D., & Murali, G. (2018). Remote health watchdog framework for seizure patient using electronic sensors. *International Journal of Engineering and Technology(UAE*, *7*, 783–785. https://doi.org/10.14419/ijet.v7i3.12132

[31] Alapati, N., Anusha, N., Joharika, P., Jerusha, N.J., Tanuja, P.(2023)Prediction of Parkinson's Disease using Machine Learning in 2023 2 nd *International Conference on Electronics and Renewable Systems(ICEARS)(pp.1357–1361).IEEE*

[32] *Naresh, A., Reddy, B.A., Reddy, G.P., Kumari, K.R., Vaishnavi, M.S.(2023)Melanocytic Pigmented Skin Lesion Detection and Classification using Hybrid Deep Features based on Fully Convolutional Network* in 2023 2 nd *International Conference on Electronics and Renewable Systems(ICEARS)(pp.1011–1018).IEEE*

[33] A. Rahman, S. Abbas, M. Gollapalli, R. Ahmed, S. Aftab, M. Ahmad, M.A. Khan, A. Mosavi, Rainfall Prediction System Using Machine Learning Fusion for Smart Cities, Sensors (2022), https://doi.org/10.3390/s22093504.

[34] Y. Kim, S. Hong, Very Short-Term Rainfall Prediction Using Ground Radar Observations and Conditional Generative Adversarial Networks, IEEE Trans. Geosci. Remote Sens. 60 (2021), https://doi.org/10.1109/TGRS.2021.3108812.

[35] A. Perera, S.D. Mudannayake, H.M. Azamathulla, U. Rathnayake, Recent climatic trends in Trinidad and Tobago, West Indies, Asia-Pacific, J. Sci. Technol. (2020) 1–11. https://so01.tci-thaijo.org/index.php/APST/article/view/205343.

[36] D.J. Mehta, S.m., Yadav, Long-term trend analysis of climate variables for arid and semi-arid regions of an Indian State Rajasthan, Int. J. Hydrol. Sci. Technol. (2022), https://doi.org/10.1504/IJHST.2022.120639.

[37] M. Mohammed, R. Kolapalli, N. Golla, S.S. Maturi, Prediction of rainfall using machine learning techniques, Int. J. Sci. Technol. Res. 9 (2020) 3236–3240.

[38] B. Chaplot, Prediction of rainfall time series using soft computing techniques, Environ. Monit. Assess. (2021), https://doi.org/10.1007/s10661-021-09388-1.

[39] D.A. An-Vo, S. Mushtaq, K. Reardon-Smith, L. Kouadio, S. Attard, D. Cobon, R. Stone, Value of seasonal forecasting for sugarcane farm irrigation planning, Eur. J. Agron. 104 (2019) 37–48, https://doi.org/10.1016/j.eja.2019.01.005.

[40] S. Miller, V. Mishra, W.L. Ellenburg, E. Adams, J. Roberts, A. Limaye, R. Griffin, Analysis of a short-term and a seasonal precipitation forecast over kenya, Atmosphere (basel). 12 (2021) 1–14, https://doi.org/10.3390/atmos12111371.

[41] M.A. Boucher, D. Tremblay, L. Delorme, L. Perreault, F. Anctil, Hydro-economic assessment of hydrological forecasting systems, J. Hydrol. 416–417 (2012) 133–144,

[42] https://doi.org/10.1016/j.jhydrol.2011.11.042.

[43] R. Fredyan, G.P. Kusuma, Spatiotemporal Convolutional Lstm With Attention Mechanism for Monthly Rainfall Prediction, Commun. Math. Biol. Neurosci. 2022 (2022). https://doi.org/10.28919/cmbn/7761.

[44] S. Aftab, M. Ahmad, N. Hameed, M.S. Bashir, I. Ali, Z. Nawaz, Rainfall Prediction in Lahore City using Data Mining Techniques, Int. J. Adv. Comput. Sci. Appl. 9 (2018). https://doi.org/10.14569/IJACSA.2018.090439.

[45] J.L. Chu, C.C. Chiang, L.H. Hsu, L.R. Hwang, Y.C. Yu, K.L. Lin, C.J. Wang, S.H. Su, T.S. Yo, A strategy to optimize the implementation of a machine-learning scheme .