Early Dengue Identification Using CBC Data and Artificial Intelligence

K. Pavani[1], Ch. Satyanarayana Reddy[2], M.Sumalatha[3]

#1 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.
#2 Assistant Professor in the Department of MCA, SRK Institute of Technology, Vijayawada.

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada.

Abstract: an artificial intelligence-based technique for early dengue detection using complete blood count (CBC) data. The comprehensive study included a wide range of machine learning (ML) and deep learning (DL) models, including Logistic Regression, SVM, Naive Bayes, Random Forest, AdaBoost, XGBoost, MLP, LightGBM, and several ensemble techniques including Stacking and Voting Classifiers. The Stacking Classifier, which combines MLP, XGBoost, and Logistic Regression with LightGBM, was utilized to increase prediction accuracy. Additionally, a Voting Classifier that integrated predictions from many models had an outstanding accuracy and F1 score of 98%. The models' performance was evaluated using accuracy, precision, recall, F1 score, and area under the curve (AUC), and comparison graphs were utilized to demonstrate each model's level of success. The findings hold out a lot of potential for using ensemble techniques in dengue prediction, which might lead to the development of more accurate diagnostic tools. Additionally, the project recommends using the Flask framework to develop a user-friendly interface that would enable secure authentication and efficient user testing.

Index terms - *Dengue detection, artificial intelligence, complete blood count (CBC), machine learning, deep learning, ensemble learning, stacking classifier, voting classifier, early diagnosis, medical data analysis.*

1. INTRODUCTION

The human body has a defense mechanism in place to ward off external microbial dangers because of its inherent sensitivity. Despite this, germs and viruses frequently infect individuals, which can result in highly fatal diseases. For instance, the Aedes mosquito is the primary vector of the virus that causes dengue disease. Every year, millions of people worldwide have dengue fever, and countless more experience its aftereffects. According to estimates from the World Health Organization (WHO) and the European Union, dengue fever affected about six million people in over 92 countries in 2023. In Bangladesh alone, there were around 0.31 million cases of hemorrhagic fever and over 1,600 fatalities. Dengue is most prevalent in urban or peri-urban settings in tropical and subtropical parts of the world. Uncontrolled urbanization, chaotic development, and inadequate sanitation are the main causes of this.

Dengue fever is most prevalent in countries in Southeast Asia, the Western Pacific, and Africa, according to the WHO's most current data. Bangladesh was the country in Southeast Asia with the most dengue cases between June and October. The highest number of dengue-related fatalities and illnesses in recent memory occurred in 2023.

2. LITERATURE SURVEY

2.1 A predictive analytics model using machine learning algorithms to estimate the risk of shock development among dengue patients:

ABSTRACT: Dengue is a common viral disease in tropical and subtropical countries. There are several clinical symptoms of dengue, ranging from asymptomatic seroconversion to severe dengue disease. Certain symptoms, such as severe bleeding, severe organ failure, and severe plasma leakage that causes shock or fluid accumulation with breathing difficulties, are indicative of severe dengue. Examining the evolution of shock while including the patients' physiological information and biochemical traits might help comprehend the progression of the illness and identify shock early. This study collects physiological patient data from dengue patients using the electronic medical records of a University Malaya Medical Center. A prediction model that is learned from the evaluation of a patient's physiological data is the foundation for both effective medication and preventing the onset of shock in critically ill patients. This work presents the predictive performance of machine learning algorithms to assess the risk of shock development in dengue patients. Logistic regression, decision trees, neural networks, and support vector machines are evaluated. Lastly, the weak learner is also exposed to ensemble learnings of bagging and boosting in order to maximize performance. The testing data shows that the bagging algorithm outperforms other competing methods, outperforming the individual decision tree by 14.5%. Day 2's full blood count (FBC), in particular haemoglobin (Hb), is a reliable indicator of the likelihood of severe dengue.

2.2 A Comparative Study between Time Series and Machine Learning Technique to Predict Dengue Fever in Dhaka City:

ABSTRACT: Dengue is the most dangerous virus that mosquitoes can spread to people. Despite government attempts, dengue outbreaks are becoming increasingly common in Bangladesh. A significant part of public health interventions is knowledge, attitude, and practice (KAP) research. The primary goal of this work is to use machine learning and time series analysis approaches to forecast the occurrence of dengue sickness in Dhaka. The model with the lowest MAPE will then be identified by comparing the models. Between January 2016 and July 2021, the Directorate General of Health and Services (DGHS) and WHO provided monthly data for this study. The study's findings demonstrate that neural networks outperform time series analysis in terms of prediction accuracy. Model 04, which had five hidden layers and produced the least error model with an error value of 0.003032557, was found to be the best-fit neural network (NN) model for the prediction of dengue illness in Dhaka city. The MAPE and RMSE values are 1.15273 and 7.588889e−06, respectively. However, the baseline dengue data used in the time series study is not steady. Use the augmented Dickey–Fuller (ADF) test in addition to the unit root test to make the difference stationary. The dengue data series is shown to be stable at the first-order difference by the ACF and PACF, which do not show a noticeable spike in the first-order difference. The ARIMA (6, 1, 1) model has the lowest AIC = -251.8, RMSE = 0.0310797, and MAPE = 15.2892, making it the best model for predicting the dengue fatality rate. Therefore, the NN model offers better prediction performance since it has the lowest MAPE value of these two models. Thus, in terms of prediction performance, neural networks do better than time series analysis. The NN model's 12-month dengue fever fatality rates show that the disease's death rate is decreasing every month. This study is more inventive than any other since its methodology differs from other research strategies. The model selection criterion is based on the most effective performance metrics, MAPE, which provide the best prediction performance and the lowest error. In light of this, the author draws the conclusion that machine learning performs better in terms of prediction than time series analysis.

2.3 Land use and meteorological influences on dengue transmission dynamics in Dhaka city, Bangladesh:

ABSTRACT: Background: Dengue fever, a virus mostly spread by Aedes mosquitoes, is a major public health problem in Dhaka, Bangladesh. Population density and other climatic and sociodemographic factors have an impact on the dengue outbreak in Dhaka. The greatest dengue indices are seen in the residential mixed zone. Urbanization, decreased vegetation, and population increase are linked to dengue prevalence, while mosquito larvae have been linked to a number of environmental parameters, such as temperature, relative humidity, rainfall, and the air pollution index. Methods: We utilize spatial and temporal analytic tools to investigate the association between dengue illness incidence, climatic parameters, and land use features using a huge dataset that spans several years. This study examines the built environment and climate as independent factors to better understand dengue dynamics. ArcPy is a Python program that facilitates the processing of geographic data, while ArcMap 10.7 is used to visualize geographical data. Results: The results of our study demonstrate that the spatial patterns of dengue incidence in Dhaka are significantly influenced by land use. Dengue hotspots have been identified in Badda, Jatrabari, Kadamtali, Mirpur, Mohammadpur, Sobujbagh, Shyampur, Tejgoan, Dhanmondi, and Uttara. All of these areas use more residential land and have denser populations than the other Thana of Dhaka. There is a substantial correlation between temperature (0.25), precipitation (.803), specific humidity (0.74), relative humidity (0.76), wind speed (0.4), and dengue incidence patterns. This study highlights the structural utilization and climatic relationship in dengue epidemics, with climatic conditions being a primary contributor to these oscillations. Results This study highlights the complex relationship between weather, land use, and the spread of dengue in Dhaka. The results of this study are extremely pertinent to several areas, including urban planning, public health campaigns, and vector control tactics. A comprehensive understanding of the geographical and temporal patterns of dengue transmission can assist build precise and effective prevention strategies to lessen the effects of dengue in metropolitan places like Dhaka.

2.4 The recent burden of dengue infection in Bangladesh: A serious public health issue:

ABSTRACT: Dengue is a rapidly growing vector-borne disease in South and Southeast Asia. This article provides an overview of the present dengue situation in Bangladesh, highlighting the major public health problems caused by this infectious disease. Between January and September 2023, 203406 people were infected; 989 of them passed away, resulting in a case fatality rate of 0.49%. Between July and September, 95.2% of these deaths and 96.1% of infections occurred. Infection and mortality rates were positively and strongly associated with both population density and the air quality index. Other environmental and social factors may have an impact on the dengue infection burden. These include methods for managing waste and storing water, rainfall, humidity, temperature, and unplanned urbanization. If we want to reduce transmission and death rates, it is imperative that we prioritize the early treatment of dengue patients and take steps to address the risk factors associated with dengue infection in the country.

2.5 The Epidemiologic and Clinical Characteristics of the 2023 Dengue Outbreak in Bangladesh:

ABSTRACT: The current dengue outbreak in Bangladesh in 2023 is the worst that has ever been documented. There is a dearth of epidemiological studies on the outbreak. The t test with two tails was employed. Multivariable logistic regression was investigated. In 2023, the dengue outbreak caused about 277,801 illnesses and 1393 deaths. Nearly 52% of the cases originated outside of Dhaka. The proportion of males to women was around 3:2. The greatest frequency of cases was found in those aged 19 to 29 (28.7%, 79 673 of 277 801; P =.001). The total case fatality rate (CFR) was 0.5%. The highest CFR (12%) was found among children aged 0 to 10. Fever was the most prevalent symptom (99%), followed by joint pain (86%). We found significantly higher risks of hospitalizations (aOR, 3.26; 95% CI, 3.11–4.04; P =.006), cases (aOR, 3.85; 95% CI, 3.25–4.12; P =.001), and fatalities (adjusted odds ratio [aOR], 4.21; 95% CI, 3.93–4.74; P =.05). This is one of the earliest descriptions of the current dengue outbreak's clinical and epidemiological characteristics.

3. METHODOLOGY

i) Proposed Work:

By using cutting-edge ensemble learning techniques, the improved dengue detection system improves upon the underlying model and greatly increases prediction accuracy and dependability. The system makes use of many machine learning and deep learning techniques, including Logistic Regression, SVM, Naive Bayes, Random Forest, AdaBoost, XGBoost, MLP, and LightGBM, using complete blood count (CBC) data as input. By combining predictions, ensemble techniques like Voting Classifier and Stacking Classifier (MLP, XGBoost, Logistic Regression, and LightGBM) improve the model's output. A hybrid CNN-LSTM model was also assessed in order to investigate the possibilities of deep learning, which helped to increase performance measures including accuracy, precision, recall, F1 score, and AUC. When these methods are used together, dengue cases may be identified early and with high accuracy, facilitating prompt treatment decisions.

The Flask web framework was used to create the system's user-friendly interface, which enables real-time interaction and diagnostic testing. Only authorized users, such medical professionals, may access patient data and prediction results thanks to our front end's strong user authentication. Healthcare professionals with little technological expertise may utilize the system efficiently because to its user-friendly design. The suggested system is a reliable, scalable, and secure tool for early dengue identification and preventive healthcare management because of its capacity to process data effectively and produce consistent findings across a variety of patient profiles.

ii) System Architecture:

The suggested dengue detection system's architecture is a layered AI-based framework that effectively analyzes CBC data to produce precise forecasts. Raw CBC data is gathered at the data layer and preprocessed for feature extraction, normalization, and noise reduction. The model layer, which contains a variety of machine learning and deep learning models including Logistic Regression, SVM, Random Forest, XGBoost, MLP, and hybrid deep learning models like CNN-LSTM, receives this enhanced data. Ensemble learning methods like Stacking and Voting Classifiers use the results of several models to improve prediction reliability. Performance criteria like as accuracy, recall, precision, and F1 score are used to assess these combined findings. A user interface created with the Flask framework that gives healthcare professionals authorized access is the last output layer. The system is both technically effective and user-friendly thanks to this interface, which enables users to enter patient CBC data, obtain prediction results in real-time, and facilitate early diagnosis and treatment.

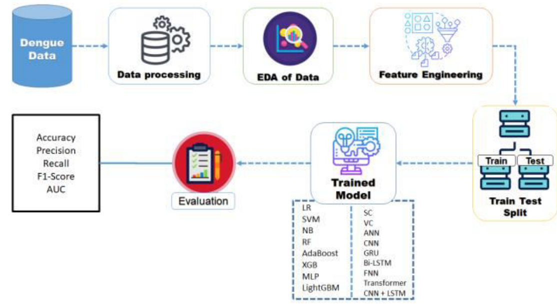

Fig.1. Proposed Architecture

iii) MODULES:

- **Data loading:** We will import the dataset using this module.
- **Data Preprocessing:** Data processing for the dengue detection system includes label encoding to translate categorical variables into numerical values, cleaning the dataset by eliminating unnecessary or missing items, and eliminating duplicate records to maintain data integrity. These procedures guarantee that the data is consistent, clean, and prepared for model training.
- **EDA:** Finding trends, patterns, and outliers in the CBC dataset is the goal of exploratory data analysis, or EDA. In order to better grasp data aspects prior to model training, it involves visualising distributions, correlations, and sample outcomes.
- **Feature Engineering:** In feature engineering, the target variable (y) and pertinent features (X) are chosen from the CBC dataset. In order to improve model accuracy and performance during training, SMOTE sampling is used to correct class imbalances, and ExtraTree is then used for feature selection.
- **Splitting data into train & test:** Data will be separated into train and test using this module.

- **Model generation:** Model building: PC, RF RFE with SelectkBest, Chi2 FS, ExtraTree FS LR, SVM, NB, RF, AdaBoost, XGB, MLP, LightGBM, SC (XGB + LR+MLP with LightGBM), VC (Boosted DT + ExtraTree), ANN, CNN, GRU, Bi-LSTM, FNN, Transformer, CNN + LSTM.
- **Admin login:** This module allows the administrator to log in.
- **Drug Side Effect Prediction:** This module allows users to upload test data.
- **Prediction:** The final prediction was shown.

iv) ALGORITHMS:

- ✓ **Logistic Regression:** used for fast implementation and interpretability in binary classification to forecast the probability of dengue based on CBC data.
- ✓ **SVM (Support Vector Machine):** By identifying the best hyperplane to divide data into groups of infected and non-infected individuals, it offers reliable dengue prediction.
- ✓ **Naive Bayes:** a probabilistic approach that uses conditional independence to classify dengue quickly and accurately, especially when working with limited samples.
- ✓ **Random Forest:** An ensemble approach, perfect for managing intricate relationships in the CBC data, increases accuracy by constructing numerous decision trees.
- ✓ **AdaBoost:** improves dengue detection accuracy by repeatedly modifying model weights, which helps weak learners perform better.
- ✓ **XGBoost:** An effective gradient boosting approach that reduces mistakes in earlier models to produce dengue forecasts that are extremely accurate.
- ✓ **MLP (Multi-Layer Perceptron):** a deep learning algorithm for dengue diagnosis that discovers intricate, non-linear correlations in CBC data.
- ✓ **LightGBM:** Fast and scalable dengue predictions are possible with this gradient boosting technique, particularly for big datasets with intricate characteristics.
- ✓ **Stacking Classifier:** use a variety of prediction strengths to increase accuracy by combining MLP, XGBoost, and Logistic Regression.
- ✓ **Voting Classifier:** combines predictions from ExtraTree and Boosted Decision Trees to produce a dengue detection result that is more accurate and dependable.
- ✓ **`ANN (Artificial Neural Network):** a deep learning method that successfully extracts intricate patterns from the CBC data for dengue prediction by simulating the way the human brain works.
- ✓ **CNN (Convolutional Neural Network):** For accurate dengue prediction, it extracts spatial hierarchies from CBC data; it is very useful for feature extraction.
- ✓ **`GRU (Gated Recurrent Unit):** GRU, a form of a recurrent neural network, is helpful for time-series-based dengue predictions because it can identify sequential relationships in CBC data.
- ✓ **Bi-LSTM (Bidirectional Long Short-Term Memory):** identifies sequential patterns in the CBC data to improve dengue prediction by capturing both historical and future data sequences.
- ✓ **FNN (Feedforward Neural Network):** CBC data is processed in a single direction for quick prediction that doesn't rely on historical data,

making it appropriate for simple categorisation applications.

- ✓ `**Transformer:** makes use of attention processes to handle sequential CBC data effectively, resulting in dengue forecasts that are extremely accurate.
- ✓ `**CNN + LSTM:** enhances dengue prediction accuracy by combining CNN for feature extraction with LSTM for sequential data processing, which captures both temporal and spatial patterns.

4. EXPERIMENTAL RESULTS

The experimental findings showed that in terms of dengue detection accuracy, ensemble models performed noticeably better than individual algorithms. With an accuracy and F1 score of 98%, the Voting Classifier that incorporated ExtraTree and Boosted Decision Trees outperformed all other evaluated models, demonstrating remarkable predictive abilities. The resilience of ensemble techniques was further shown by the Stacking Classifier, which combined MLP, XGBoost, and Logistic Regression with LightGBM to provide extremely accurate predictions. By identifying spatial and sequential patterns in CBC data, deep learning models such as CNN, Bi-LSTM, and Transformer significantly improved model accuracy. Metrics including accuracy, precision, recall, F1 score, and AUC were used to thoroughly assess the system. The findings were displayed using comparison graphs, demonstrating the model's capacity to accurately and effectively identify dengue, even in its early stages.

Accuracy: A test's accuracy is determined by its capacity to distinguish between healthy and ill cases. To gauge the accuracy of the test, find the percentage of examined instances that had true positives and true negatives. According to the computations:

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Precision is the number of affirmative cases or the classification's accuracy rate. The following formula is applied to assess accuracy:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: A model's ability to recognise every instance of a pertinent machine learning class is measured by its recall. The ratio of accurately predicted positive observations to the total number of positives indicates how well a model can identify class instances.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: Mean Average Precision is one ranking quality metric (MAP). It considers the number of relevant recommendations and their position on the list. MAP at K is calculated as the arithmetic mean of the Average Precision (AP) at K for each user or query.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

F1-Score: An accurate machine learning model is indicated by a high F1 score. combining precision and recall to increase model correctness. The accuracy statistic quantifies the frequency with which a model correctly predicts a dataset.

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

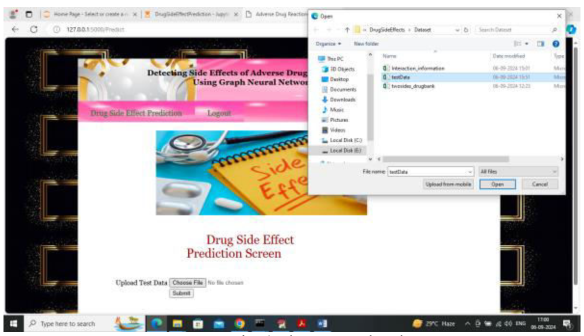

Fig.4. dataset upload

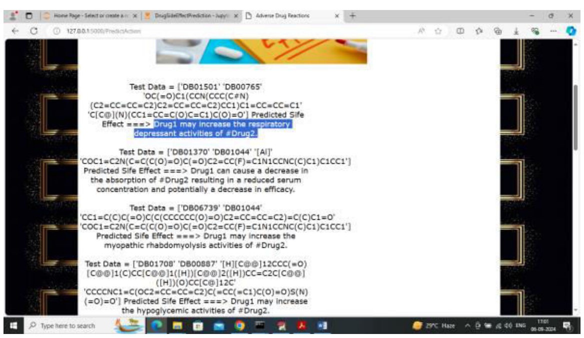

Fig.5. results

	Algorithm Name	Accuracy	Precision	Recall	FSCORE
0	Existing KNN	91.560102	91.920962	91.305660	91.323465
1	Existing Decision TRee	95.140665	95.339149	95.618899	95.454287
2	Propose GNN	97.698210	97.889953	97.702843	97.722705
3	Extension CNN2D	99.872123	99.903101	99.888143	99.895292

Fig.6.accuracy table results

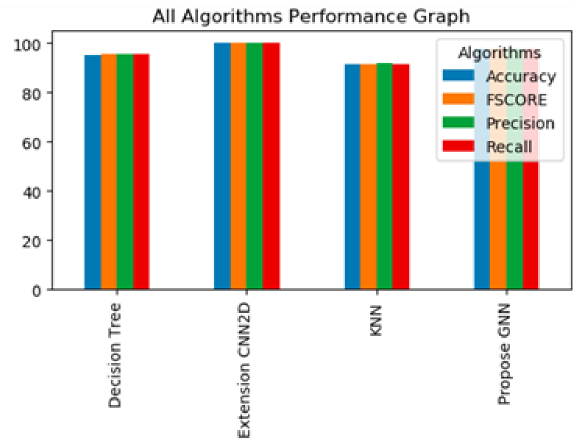

Fig.6. graphical representation

5. CONCLUSION

In summary, an AI-based approach for CBC data-based early dengue detection has a lot of potential to improve healthcare outcomes by facilitating accurate and timely diagnosis. A variety of machine learning and deep learning algorithms, including Logistic Regression, SVM, Naive Bayes, Random Forest, AdaBoost, XGBoost, MLP, LightGBM, and complex ensemble techniques like the Stacking and Voting Classifiers, are used by the system to provide accurate predictions. The ability of deep learning models, such as CNN, Bi-LSTM, and Transformer, to recognize complex patterns in the data is enhanced. SMOTE sampling is used to correct class imbalances, while ExtraTree feature selection maximizes model performance. With a Voting Classifier accuracy rate of up to 98%, the technique is quite reliable for dengue detection. Additionally, a user-friendly interface created using Flask ensures usability, while secure user authentication safeguards user privacy. Better real-time diagnostics and dengue identification are made possible by this approach, which may decrease the disease's transmission and improve patient outcomes by facilitating early treatment. By looking at ensemble methodologies and refining the model, the system's capabilities may be further enhanced.

6. FUTURE SCOPE

The future scope of this system involves integrating real-time data from health monitoring devices to enhance early detection capabilities. Adding more diverse datasets may also improve the model's generalizability to different populations. Accuracy and robustness might be further enhanced by looking at complex ensemble techniques and hybrid models. By utilizing cloud-based solutions, healthcare professionals may get more accessibility and scalability. Last but not least, continuous model modifications and enhancements based on new research findings will keep the system at the forefront of dengue detection technology.

REFERENCES

[1] J. K. Chaw, S. H. Chaw, C. H. Quah, S. Sahrani, M. C. Ang, Y. Zhao, and T. T. Ting, ''A predictive analytics model using machine learning algorithms to estimate the risk of shock development among dengue patients,'' Healthcare Anal., vol. 5, pp. 1–17, Jul. 2024.

[2] T. Akter, M. T. Islam, M. F. Hossain, and M. S. Ullah, ''A comparative study between time series and machine learning technique to predict dengue fever in Dhaka city,'' Discrete Dyn. Nature Soc., vol. 2024, pp. 1–12, May 2024.

[3] S. Roy, A. Biswas, M. T. A. Shawon, S. Akter, and M. M. Rahman, ''Land use and meteorological influences on dengue transmission dynamics in Dhaka city, Bangladesh,'' Bull. Nat. Res. Centre, vol. 48, no. 1, pp. 1–24, Mar. 2024, doi: 10.1186/S42269-024-01188-0.

[4] N. Ali, ''The recent burden of dengue infection in bangladesh: A serious public health issue,'' J. Infection Public Health, vol. 17, no. 2, pp. 226–228, Feb. 2024.

[5] N. Sharif, N. Sharif, A. Khan, and S. K. Dey, ''The epidemiologic and clinical characteristics of the 2023 dengue outbreak in Bangladesh,'' Open Forum Infectious Diseases, vol. 11, no. 2, pp. 1–29, Feb. 2024, doi: 10.1093/OFID/OFAE066.

[6] D. C. Kajeguka, F. M. Mponela, E. Mkumbo, A. N. Kaaya, D. Lasway, R. D. Kaaya, M. Alifrangis, E. Elanga-Ndille, B. T. Mmbaga, and R. Kavishe, ''Prevalence and associated factors of dengue virus circulation in the rural community, Handeni district in Tanga, Tanzania,'' J. Tropical Med., vol. 2023, pp. 1–9, Nov. 2023.

[7] M. A. Kabir, H. Zilouchian, M. A. Younas, and W. Asghar, ''Dengue detection: Advances in diagnostic tools from conventional technology to point of care,'' Biosensors, vol. 11, no. 7, p. 206, Jun. 2021.

[8] C. Davi, A. Pastor, T. Oliveira, F. B. d. L. Neto, U. Braga-Neto, A. W. Bigham, M. Bamshad, E. T. A. Marques, and B. Acioli-Santos, ''Severe dengue prognosis using human genome data and machine learning,'' IEEE Trans. Biomed. Eng., vol. 66, no. 10, pp. 2861–2868, Oct. 2019.

[9] D. Sarma, S. Hossain, T. Mittra, Md. A. M. Bhuiya, I. Saha, and R. Chakma, ''Dengue prediction using machine learning algorithms,'' in Proc. IEEE 8th R10

Humanitarian Technol. Conf., Dec. 2020, pp. 1–6.

[10] E. Fernández, M. Smieja, S. D. Walter, and M. Loeb, ''A predictive model to differentiate dengue from other febrile illness,'' BMC Infectious Diseases, vol. 16, no. 1, pp. 1–7, Dec. 2016.

[11] H. Mayrose, G. M. Bairy, N. Sampathila, S. Belurkar, and K. Saravu, ''Machine learning-based detection of dengue from blood smear images utilizing platelet and lymphocyte characteristics,'' Diagnostics, vol. 13, no. 2, p. 220, Jan. 2023.

[12] S. Sabrina Prome, T. Basak, T. Islam Plabon, and R. Khan, ''Prediction of dengue cases in Bangladesh using explainable machine learning approach,'' in Proc. Int. Conf. Inventive Comput. Technol. (ICICT), Apr. 2024, pp. 1–5.

[13] J. D. Mello-Román, J. C. Mello-Román, S. Gómez-Guerrero, and M. García-Torres, ''Predictive models for the medical diagnosis of dengue: A case study in Paraguay,'' Comput. Math. Methods Med., vol. 2019, pp. 1–7, Jul. 2019.

[14] S. K. Dey, M. M. Rahman, A. Howlader, U. R. Siddiqi, K. M. M. Uddin, R. Borhan, and E. U. Rahman, ''Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh: A machine learning approach,'' PLoS One, vol. 17, no. 7, Jul. 2022, Art. no. e0270933.

[15] B. Abdualgalil, S. Abraham, and W. M. Ismael, ''Early diagnosis for dengue disease prediction using efficient machine learning techniques based on clinical data,'' J. Robot. Control (JRC), vol. 3, no. 3, pp. 257–268, May 2022.

[16] S. Q. Ong, P. Isawasan, A. M. M. Ngesom, H. Shahar, A. M. Lasim, and G. Nair, ''Predicting dengue transmission rates by comparing different machine learning models with vector indices and meteorological data,'' Sci. Rep., vol. 13, no. 1, pp. 1–10, Nov. 2023.

[17] M. E. H. Kayesh, I. Khalil, M. Kohara, and K. Tsukiyama-Kohara, ''Increasing dengue burden and severe dengue risk in Bangladesh: An overview,'' Tropical Med. Infectious Disease, vol. 8, no. 1, p. 32, Jan. 2023.

[18] M. T. Sarwar and M. A. Mamun, ''Prediction of dengue using machine learning algorithms: Case study Dhaka,'' in Proc. 4th Int. Conf. Electr., Comput. Telecommun. Eng. (ICECTE), Dec. 2022, pp. 1–6.

[19] M. B. Khan, Z.-S. Yang, C.-Y. Lin, M.-C. Hsu, A. N. Urbina, W. Assavalapsakul, W.-H. Wang, Y.-H. Chen, and S.-F. Wang, ''Dengue overview: An updated systemic review,'' J. Infection Public Health, vol. 16, no. 10, pp. 1625–1642, Oct. 2023.

[20] M. A. Majeed, H. Z. M. Shafri, Z. Zulkafli, and A. Wayayok, ''A deep learning approach for dengue fever prediction in Malaysia using LSTM with spatial attention,'' Int. J. Environ. Res. Public Health, vol. 20, no. 5, p. 4130, Feb. 2023.

[21] R. Real, A. M. Barbosa, and J. M. Vargas, ''Obtaining environmental favourability functions from logistic

regression,'' Environ. Ecol. Statist., vol. 13, no. 2, pp. 237–245, Jun. 2006.

[22] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, ''Applications of support vector machine (SVM) learning in cancer genomics,'' Cancer Genomics Proteomics, vol. 15, pp. 41–51, Jul. 2018.

[23] A. Liaw and M. Wiener, ''Classification and regression by randomforest,'' R News, vol. 2, pp. 18–22, Jul. 2002.

[24] D. J. Hand and K. Yu, ''Idiot's bayes— Not so stupid after all?'' Int. Stat. Rev., vol. 69, no. 3, pp. 385–398, Dec. 2001.

[25] A. Yulianto, P. Sukarno, and N. A. Suwastika, ''Improving AdaBoost-based intrusion detection system (IDS) performance on CIC IDS 2017 dataset,'' J. Phys., Conf. Ser., vol. 1192, Mar. 2019, Art. no. 012018.

[26] T. Chen and C. Guestrin, ''XGBoost: A scalable tree boosting system,'' in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2016, pp. 785–794.

[27] J. Zhang, D. Mucs, U. Norinder, and F. Svensson, ''LightGBM: An effective and scalable algorithm for prediction of chemical toxicity– application to the Tox21 and mutagenicity data sets,'' J. Chem. Inf. Model., vol. 59, no. 10, pp. 4150–4158, 2019.

[28] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[29] N. Hollmann, S. Muller, K. Eggensperger, and F. Hutter, ''TabPFN: A transformer that solves small tabular classification problems in a second,'' 2022, arXiv:2207.01848.

[30] T. Kiranbhai Vyas, ''Deep learning with tabular data: A self-supervised approach,'' 2024, arXiv:2401.15238.

Author Profiles

Ms. K. Pavani Working as Assistant & Head of Department of MCA, in SRK Institute of technology in Vijayawada. She done with MCA, M. Tech in Computer Science. Her area of interest includes Machine Learning with Python and DBMS.

Mr. Ch.Satyanarayana Reddy Completed his MCA, BCA. He has System Administrator, Networking Administrator and Oracle Administrator. He also a web developer and python developer, currently working has an Assistant Professor in the department of MCA at SRK Institute of Technology, Enikepadu, NTR District. His area of interest includes Artificial Intelligence and Machine Learning.

Ms.M.Sumalatha is MCA Student in the Department of Computer Applications at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. He has Completed Degree in B.Sc. (Mathematics,Physics,Computer science) from A.V.R Degree college Avanigadda. His area of interest are DBMS and Machine Learning with Python.