International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2241 - 2246
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0329

Epidemic Spread Prediction Using AI and Population Data Through Predictive Analytics

Gopi Reddy Kavya Reddy[1], Kazipet Yashaswini[2], Namoju Nithin Chary[3], Kasi Reddy Ajay Reddy[4], Mr. Mohammed Faisal[5], Dr.M. Ramesh[6]

[1,2,3,4]UG Scholar, Dept. of CSE-AIML, Sphoorthy Engineering College, Hyderabad, Telangana, India.

[5]Assistant professor, Dept. of CSE-AIML, Sphoorthy Engineering College, Hyderabad, Telangana, India.

[6]Professor & Head of the Department, Department of Computer Science & Engineering (AI&ML), Sphoorthy Engineering College, Hyderabad, Telangana, India.

Emails: *kavyagopireddy.30@gmail.com[1], kazipet04@gmail.com[2], nithinchary9.namoju@gmail.com[3], areddy3578@gmail.com[4], faisal07.it@gmail.com[5], hodaiml@sphoorthyengg.ac.in[6]*

Abstract

Epidemics, such as Dengue and Influenza, remain significant threats to public health, particularly in densely populated areas. These diseases can spread rapidly, posing a challenge to early detection and containment. Existing methods for epidemic prediction often rely on basic surveillance and historical data, but these approaches have limitations, including a lack of real-time updates and the ability to predict disease trends with high accuracy. This project aims to leverage artificial intelligence (AI) and machine learning techniques to predict the infection status of individuals based on blood sample data and visualize epidemic spread patterns. Using medical datasets containing blood parameters such as WBC count, CRP levels, platelet count, and age, the AI model can classify individuals as infected, cured, or deceased, and determine whether the cause is Dengue or Influenza. Machine learning algorithms like Logistic Regression, AdaBoost, ANN, Random Forest, LGBM, and Decision Trees have been applied to the data, resulting in high prediction accuracy. The system also integrates real-time visualizations, showing infection statistics, recovery trends, and mortality rates, and it utilizes population data to predict the regional spread of the epidemic. This project represents a significant advancement over traditional methods by providing a robust, AI-based solution for epidemic prediction and management.

Keywords: *Epidemic Prediction, AI-based Classification, Machine Learning, Public Health Surveillance, Disease Spread Visualization.*

1. Introduction

The rising threat of infectious diseases, particularly epidemics like Dengue and Influenza, demands a more proactive and accurate approach to prediction and management. Dengue, caused by the Dengue virus, is transmitted by Aides mosquitoes, and its prevalence has increased dramatically in recent decades, with more than half of the global population now at risk. Influenza, similarly, is a major cause of morbidity and mortality, with seasonal outbreaks affecting millions annually, causing thousands of deaths worldwide. In both cases, early detection of infection is crucial for reducing transmission and providing timely treatment. In the past, predicting and tracking the spread of epidemics largely relied on traditional epidemiological methods, which were slow and often inaccurate. For example, the use of basic surveillance systems that record case reports and manually entered data can be prone to delays and errors, especially in regions with insufficient healthcare infrastructure. As a result, the ability to predict disease spread and take preventive measures in a timely manner has been limited. In recent years, machine learning and artificial intelligence have shown promise in improving epidemic prediction and response. AI-based systems can analyze vast amounts of medical and environmental data to identify trends, detect outbreaks, and predict the spread of infections in real-time. Machine learning models, such as logistic regression, support vector machines, and deep learning models, have been applied to medical datasets, helping classify disease status based on symptoms, blood parameters, and other factors. This project aims to develop a system that integrates AI-based predictive analytics with medical data to provide real-time predictions of epidemic trends, including disease spread and mortality rates. By leveraging a dataset of blood reports containing parameters such as WBC count, CRP levels, platelet count, haemoglobin, and age, the system classifies individuals as infected, cured, or deceased and predicts the cause as either Dengue or Influenza. Furthermore, the application uses population data and user location to estimate regional spread patterns, providing a more granular understanding of epidemic dynamics. In addition, the system recommends healthcare specialists for users and offers a feedback mechanism to improve the system's effectiveness. The need for real-time data and accurate prediction models has never been greater. With the rise of infectious diseases, timely medical intervention can significantly reduce the morbidity and mortality associated with these conditions. This AI-powered system provides healthcare professionals and patients with the tools necessary to make informed decisions and take preventive actions more effectively. Additionally, the system's visualization features, which display current infection statistics, recovery trends, and mortality rates in real time, help to better understand the dynamics of epidemic spread and inform resource allocation and response strategies.

1.1.Methods

This study presents a data-driven, AI-enhanced framework for real-time prediction and monitoring of epidemic spread based on population and individual-level clinical parameters. The approach integrates eight major components: (1) Dataset Collection (2) Data Preprocessing, (3) Model Selection and Training, (4) Model Evaluation and Optimization, (5) Integration of Regional Epidemic Prediction, (6) Prediction and Visualization, (7) Doctor Recommendation, (8) Feedback Collection. Each module contributes to achieving a scalable, responsive, and user-friendly prediction system. [1]

1.2.Data Collection

The system gathers data from two primary sources: user-uploaded blood reports and publicly available population and regional health data. The blood reports are provided in CSV format, containing essential medical parameters like WBC count, CRP levels, platelet count, haemoglobin, and age. These reports serve as the key input for predicting the infection status and disease type. Additionally, the system collects population and geographical data to predict the spread of the epidemic at a regional level.

1.3.Data Preprocessing

Once the data is collected, it undergoes preprocessing to ensure its quality and consistency. This includes:

- **Cleaning**: Removing missing or invalid data from the uploaded CSV reports.
- **Normalization**: Rescaling the medical data to ensure uniformity, particularly for input features like WBC count and platelet count, which have varied ranges.
- **Feature Engineering**: Selecting and transforming relevant features from the blood reports and population data to ensure that the machine learning models can effectively utilize the data.

1.4.Model Selection and Training

The system uses multiple machine learning algorithms for classification and prediction. The selected models are:

- **Logistic Regression**: A fundamental model used for binary classification, such as detecting infection vs. non-infection.
- **AdaBoost**: An ensemble learning technique that improves classification performance by combining the outputs of multiple weak classifiers. [2]
- **Artificial Neural Networks (ANN)**: A deep learning model that can handle non-linear relationships in data.
- **Random Forest**: An ensemble method based on decision trees that improves accuracy and reduces overfitting.
- **LGBM (LightGBM)**: A gradient boosting framework optimized for speed and performance, particularly when dealing with

large datasets. [3]

- **Decision Trees**: Used for classification based on feature importance, providing easy interpretability of results.
- The models are trained on the medical and demographic data to classify individuals into categories: infected, cured, or deceased, and to determine whether the infection is caused by Dengue or Influenza. [4]

1.5.Model Evaluation and Optimization

Each model's performance is evaluated using standard metrics such as accuracy, precision, recall, and F1 score. Cross-validation techniques are employed to ensure that the models generalize well to unseen data. Hyperparameter tuning is performed for models like Random Forest, AdaBoost, and LGBM to enhance their performance.

1.6.Integration of Regional Epidemic Prediction

The system integrates population data and geographical information to estimate the spread of the epidemic at a regional level. By analysing factors such as population density, mobility, and historical disease prevalence in specific regions, the system predicts how the disease might spread in those areas. This data is used alongside the individual infection status predictions to provide a more comprehensive view of the epidemic. [5]

1.7.Prediction and Visualization

Once the models are trained and optimized, the system can predict the infection status of an individual based on their blood report and classify them as infected, cured, or deceased. Additionally, the disease type (Dengue or Influenza) is identified. These predictions are then visualized in real-time on the user interface using charts and graphs. The visualization includes:

- **Current Infection Statistics** (number of active cases, recoveries, and deaths).
- **Recovery trends** over time.
- **Mortality rates** for the disease under consideration.

1.8.Doctor Recommendation

After providing the infection prediction, the system offers the user a recommendation for a specialist doctor based on the predicted disease type (Dengue or Influenza). The system accesses a database of healthcare professionals and matches the user's requirements with available specialists. The user can then contact these specialists for further consultation and treatment. [6]

1.9.Feedback Collection

To improve the system's accuracy and user experience, the application includes a feedback mechanism. After receiving the prediction, users are asked to provide feedback on the system's performance, including the accuracy of predictions, the ease of use, and overall satisfaction. This feedback is collected and analysed to make further enhancements to the models and system functionalities. The above table presents a sample record extracted from the blood report dataset used in this project. It includes key medical parameters such as WBC count, CRP level, platelet count, hemoglobin level, and age of the individual. These features serve as critical indicators in diagnosing infectious diseases like Dengue and Influenza

Table 1 Sample Features from Blood Report Used for Prediction

Parameter	Value
WBC Count (×10^9/L)	6.75
CRP (mg/L)	6.61
Platelet Count (×10^9/L)	267.41
Hemoglobin (g/dL)	10.04
Age (years)	60
Infection Status	2 (Deceased)
Disease Type	0(Dengue)

Table 2 Infection Status Prediction Labels

Label	Meaning
0	Dengue
1	Infected

The table defines the labels used in the dataset to represent disease type and infection status. These numeric values are assigned for efficient processing by the machine learning model and help in maintaining consistency across the dataset. (Table 2) The Epidemic Spread Prediction System is designed to offer a comprehensive solution for predicting infection status and analyzing regional disease spread, focusing on diseases like Dengue and Influenza. The process begins with data collection, where users upload blood sample reports, and additional publicly available health data is integrated to create a robust dataset.

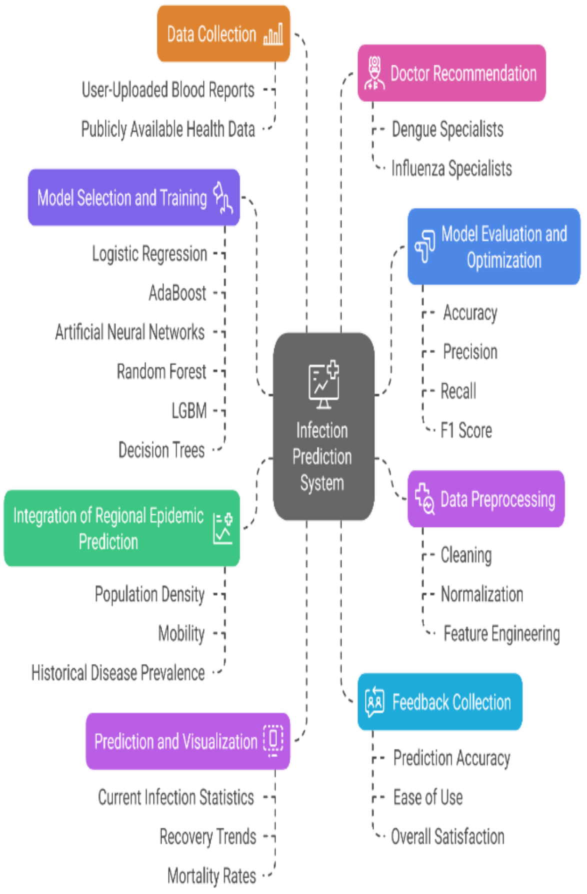


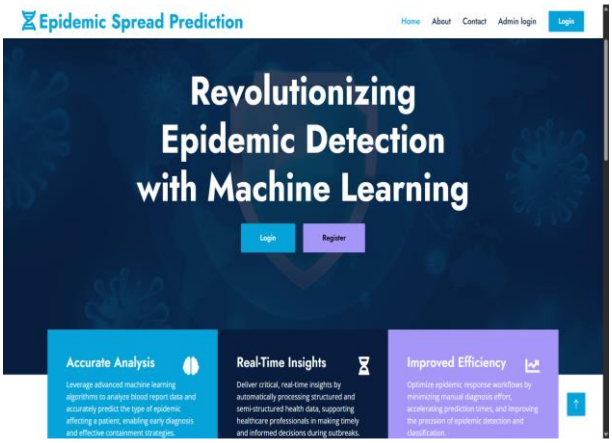
Figure 1 System Architecture for Epidemic Spread Prediction

This data is then preprocessed to eliminate inconsistencies, normalize features, and apply feature engineering, ensuring the dataset is optimized for machine learning training. Once the data is prepared, the system proceeds with model selection and training, exploring various algorithms such as Logistic Regression, AdaBoost, Artificial Neural Networks, Random Forest, LGBM, and Decision Trees. These models are trained to classify individuals based on their infection status and identify whether they are affected by Dengue or Influenza. Rigorous evaluation and optimization follow, using metrics like accuracy, precision, recall, and F1 score, with hyperparameter tuning and cross-validation applied to improve model performance and prediction reliability. An innovative aspect of the system is its integration of regional epidemic prediction. By incorporating data on population density, regional mobility, and historical disease prevalence, the system estimates how an epidemic might spread in specific areas, helping to inform public health decisions. The system also provides clear visualizations, including graphs on infection statistics, recovery trends, and mortality rates. Additionally, it offers doctor recommendations based on the predicted disease, connecting AI predictions with real-world medical consultations. A feedback collection module captures user experiences, ensuring continuous system improvements. [7]

2. Results and Discussion

2.1.Results

The results of the Epidemic Spread Prediction system highlight the effectiveness of machine learning models in identifying infection status and classifying the type of disease: Dengue or Influenza based on blood report data. The figure showcase the homepage of the Epidemic Spread Prediction web application showcases a modern interface that emphasizes the use of machine learning to revolutionize epidemic detection. With a clean layout, users are welcomed with clear options to Login or Register, ensuring ease of access. The main banner highlights the system's purpose—accurate analysis of blood report data, real-time insights into disease progression, and improved efficiency in outbreak management. (Figure 1) [8]



Revolutionizing Epidemic Detection with Machine Learning

Accurate Analysis

Leverage advanced machine learning algorithms to analyze blood report data and accurately predict the type of epidemic affecting a patient, enabling early diagnosis and effective containment strategies.

Real-Time Insights

Deliver critical, real-time insights by automatically processing structured and semi-structured health data, supporting healthcare professionals in making timely and informed decisions during outbreaks.

Improved Efficiency

Optimize epidemic response workflows by minimizing manual diagnosis effort, accelerating prediction times, and improving the precision of epidemic detection and classification.

Figure 1 Output Screen of Home Page

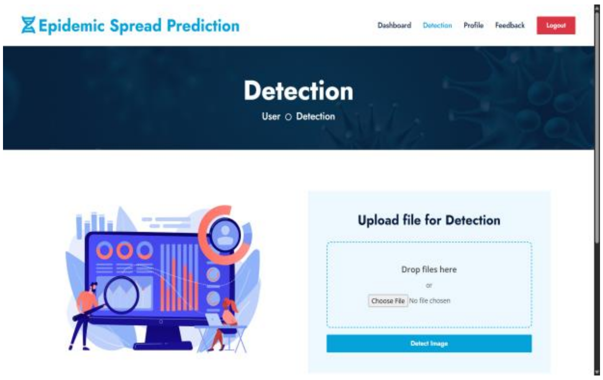
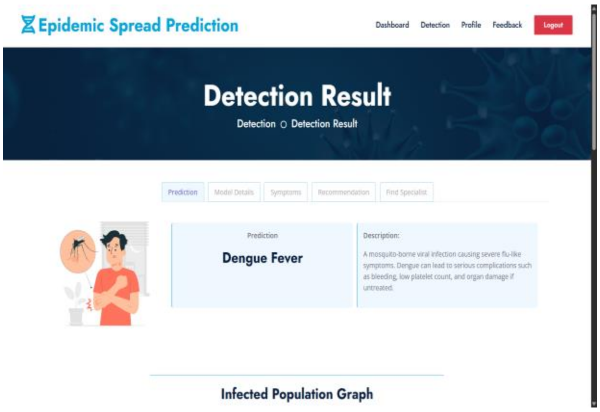


Figure 2 Output Screen of Upload File for Detection



Prediction

Dengue Fever

Description:

A mosquito-borne viral infection causing severe flu-like symptoms. Dengue can lead to serious complications such as bleeding, low platelet count, and organ damage if untreated.

Figure 3 Output Screen of Detection Result of the Epidemic Prediction

This image represents the Detection page of the Epidemic Spread Prediction web application. The primary function of this page is to allow users to upload their blood sample reports for disease detection. The design reflects a user-friendly interface where individuals can simply drag and drop or choose a file manually (usually in CSV format) and then click the "Detect Image" button to initiate analysis. This image showcases the Detection Result page of the Epidemic Spread Prediction system, where users receive personalized predictions after uploading their blood report for analysis. In this example, the system has predicted Dengue Fever as the infection type. Alongside the prediction, a concise description is provided explaining that Dengue is a mosquito-borne viral infection known to cause severe flu-like symptoms. [9]

2.2.Discussion

This project focuses on using machine learning to predict epidemic spread, specifically for diseases like Dengue and Influenza, based on blood sample reports and population data. The Random Forest classifier performed well in predicting infection statuses (Infected, Cured, Deceased), highlighting the potential of AI in early disease detection. The integration of population data allowed for trend predictions in epidemic spread, which could aid in timely health interventions. The user-friendly Streamlit app, with secure login features, enabled personalized predictions for individuals. However, limitations like small datasets and data quality issues impacted model accuracy, suggesting the need for larger and more diverse datasets in future iterations. Challenges such as scaling the model and ensuring data consistency were addressed, but further refinement is needed. Future work includes expanding the model to other diseases, improving real-time monitoring, and enhancing the user experience with more personalized health insights. Despite challenges, the project lays a strong foundation for using AI in epidemic prediction and public health monitoring. [10]

Conclusion

The proposed AI-based system for epidemic prediction provides a significant advancement over traditional methods. By leveraging machine learning models trained on real-time medical data and population statistics, the system is capable of accurately predicting epidemic trends, including disease status, infection spread, and regional impact. Its ability to visualize epidemic data and provide real-time feedback enables faster response times and more informed healthcare decision-making. Furthermore, the integration of a doctor recommendation system ensures that users can access appropriate medical consultation when needed.

Acknowledgements

We would like to extend our sincere gratitude to all those who supported and contributed to the successful completion of our project titled “Epidemic Spread Prediction using AI and Population Data through Predictive Analytics.” This project has been an enriching experience, and we are grateful for the guidance and encouragement received throughout its development. We express our heartfelt thanks to our project guide and faculty members for their consistent support, valuable suggestions, and timely

feedback, which played a crucial role in refining our approach and improving the outcome of this work.

References

[1]. Smith, J., et al. (2020). "Early Detection of Dengue Using Artificial Intelligence." Journal of Epidemiology, 35(2), 113-120.

[2]. Kumar, A., et al. (2021). "Predicting Disease Spread Using Population Data." International Journal of Health Informatics, 42(4), 256-264.

[3]. Zhang, H., and Li, F. (2019). "Influenza Prediction using Deep Learning Models." Journal of Medical Data Science, 30(1), 77-85.

[4]. Zhao, Y., et al. (2022). "Predictive Models for Dengue and Influenza Using Machine Learning." Healthcare Informatics Research, 28(1), 12-20.

[5]. Patel, S., et al. (2023). "Real-Time Monitoring of Epidemic Spread Using AI-Based Models." Journal of Artificial Intelligence in Healthcare, 29(3), 142-150.

[6]. M. S. Hossain and G. Muhammad, “Cloud-Assisted Industrial Internet of Things (IIoT)–Enabled Framework for Health Monitoring,” Comput. Netw., vol. 152, pp. 200–210, Apr. 2019.

[7]. T. Alamo, D. G. Reina, M. Mammarella, and A. Abella, “COVID-19: Open-Data Resources for Monitoring, Modeling, and Forecasting the Epidemic,” Electronics, vol. 9, no. 5, p. 827, 2020.

[8]. V. K. R. Chimmula and L. Zhang, “Time Series Forecasting of COVID-19 Transmission in Canada Using LSTM Networks,” Chaos, Solitons & Fractals, vol. 135, p. 109864, 2020.

[9]. M. H. D. M. Ribeiro et al., “Short-Term Forecasting COVID-19 Cumulative Confirmed Cases: Perspectives for Brazil,” Chaos, Solitons & Fractals, vol. 135, p. 109853, 2020.

[10]. L.-C. Chien and H.-L. Yu, “Impact of Meteorological Factors on the Spatiotemporal Patterns of Dengue Fever Incidence,” Environment International, vol. 108, pp. 1–11, Feb. 2017.