Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

Exploratory Data Analysis And Predicttive Insights On Influenza(Flu) Virus Trends Using Machine Learning

Sathish H[1], Sai Kishore G[2], Shruthi B[2], Arya Sree K[2], Sai Ganesh Ch[2]

[1]Assistant Professor, Department of CSE, Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India

[2]UG Students, Department of CSE, Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India

Correspondence

Sathish H

Assistant Professor, Department of CSE, Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India

- Received Date: 25 Jan 2026
- Accepted Date: 14 Mar 2026
- Publication Date: 20 Mar 2026

Keywords

Influenza, Exploratory Data Analysis (EDA), Machine Learning, Random Forest, Multiclass Classification, Predictive Modeling, Disease Surveillance, Public Health.

Abstract

Influenza is a recurring viral infection that poses a significant global health challenge due to its seasonal outbreaks and rapid spread. This study focuses on applying Exploratory Data Analysis (EDA) to examine historical influenza data and identify key patterns, correlations, and seasonal variations that influence the spread of the virus. Understanding these trends is essential for improving early detection and prevention strategies.

Machine learning techniques are employed to enhance prediction accuracy, with a Random Forest classifier used for multiclass classification of influenza risk levels. The system is developed using Python and integrated with a Django-based web application, enabling real-time predictions. Scikit-learn is used for model training, joblib for model persistence, and PostgreSQL for database management. The results demonstrate the effectiveness of machine learning in forecasting influenza outbreaks and highlight its role in supporting data-driven public health decision-making.

Introduction

Influenza, commonly known as flu, is a contagious respiratory illness caused by influenza viruses that affect the nose, throat, and lungs. Despite advancements in medical science and vaccination, influenza continues to cause seasonal outbreaks and remains a major global health concern. Its rapid mutation and transmission make accurate prediction challenging. In recent years, artificial intelligence (AI) and data analytics have emerged as effective tools for analyzing disease patterns and forecasting outbreaks.

This project aims to analyze historical influenza data and develop predictive models using Exploratory Data Analysis (EDA) and machine learning techniques. The Random Forest algorithm is employed for multiclass classification to improve prediction accuracy. The system is implemented using Python, with a Django-based web application for user interaction. PostgreSQL is used for data management, while scikit-learn and joblib support model training and deployment.

The proposed approach provides a data-driven solution for predicting influenza trends, enabling healthcare authorities to make informed decisions. It supports efficient resource allocation, early warning systems, and improved public health preparedness, thereby reducing the impact of seasonal flu outbreaks.

Background

Influenza viruses are highly dynamic in nature and frequently undergo mutations, which makes accurate prediction of outbreaks a complex task. Traditional statistical methods such as time-series models are effective in identifying seasonal trends but often fail to capture nonlinear relationships and hidden patterns present in large healthcare datasets. In addition, conventional influenza monitoring systems rely heavily on historical data analysis and delayed reporting mechanisms, which limit their effectiveness in providing early warnings. With the emergence of Artificial Intelligence (AI) and Machine Learning (ML), more advanced techniques have been developed to analyze complex datasets, identify meaningful patterns, and improve prediction accuracy. Exploratory Data Analysis (EDA) plays a crucial role in this process by enabling a deeper understanding of data distributions, correlations, and seasonal variations before model development. By combining EDA with machine learning, it is possible to build more reliable and efficient systems for influenza prediction and healthcare analytics.

Problem Statement

Despite advancements in healthcare technology, existing influenza prediction systems face several limitations that reduce their effectiveness. Most current systems lack real-time prediction capabilities and depend on

Citation: Sathish H, Sai Kishore G, Shruthi B, Arya Sree K, Sai Ganesh Ch. Exploratory Data Analysis And Predicttive Insights On Influenza(Flu) Virus Trends Using Machine Learning. GJEIIR. 2026;6(3):0168.

manual data collection and delayed reporting, which hinders early detection of outbreaks. Traditional models often exhibit low prediction accuracy due to their inability to handle complex and high-dimensional data. Furthermore, many systems do not provide an integrated platform that combines data analysis, predictive modeling, and user interaction, making them less accessible and less useful for both patients and healthcare professionals. Therefore, there is a need for an intelligent and efficient system that can analyze user symptoms, predict influenza risk using machine learning techniques, and provide real-time results through a user-friendly interface. The proposed system addresses these challenges by integrating Exploratory Data Analysis, Random Forest-based prediction, and a web-based platform to enhance accuracy, accessibility, and overall healthcare decision-making.

Literature Survey

Early research in influenza prediction primarily relied on statistical models such as ARIMA and regression analysis, which were effective in capturing seasonal patterns but struggled with complex and nonlinear relationships in data. With the advancement of machine learning, various algorithms such as Decision Trees, Support Vector Machines, and Random Forest have been applied to disease prediction tasks, showing improved performance and accuracy. Recent studies have also explored the integration of predictive models with web-based platforms to provide real-time healthcare solutions. These systems aim to combine data analysis, prediction, and user interaction in a single framework. However, many existing approaches still lack comprehensive integration of Exploratory Data Analysis, machine learning, and interactive interfaces. This project builds upon previous research by combining EDA and Random Forest classification within a Django-based web application, thereby enhancing prediction accuracy and usability while providing a scalable solution for influenza monitoring.

Proposed system

The proposed system is designed as an integrated influenza prediction platform that utilizes machine learning techniques to analyze user-input symptoms and predict disease risk. The system operates through a structured workflow in which the user enters symptoms through a web interface, and the data is processed by the backend server. The input data is then preprocessed and converted into a suitable format for the machine learning model. The Random Forest classifier analyzes the processed data and predicts the influenza risk level, which is subsequently displayed to the user. The system is built using a multi-layer architecture consisting of frontend, backend, database, and machine learning components. The frontend is developed using Django templates to provide an interactive user interface, while the backend handles request processing and model integration. PostgreSQL is used as the database for storing user and prediction data securely. The system also includes modules such as user management, prediction processing, doctor interaction, and administrative control, all working together to provide a complete healthcare solution. This integrated approach ensures improved prediction accuracy, real-time analysis, and efficient data handling.

Modules Used

User Module:The user module handles registration, login, and authentication processes. It allows users to securely enter symptoms and view prediction results through the web interface.

Patient Module: The patient module allows users to input symptoms and view influenza prediction results. It also supports accessing consultation features and maintaining personal health records.

Prediction Module: The prediction module is the core component where the Random Forest algorithm processes input symptoms. It generates accurate influenza risk predictions based on trained machine learning models.

Doctor Module: The doctor module enables healthcare professionals to access patient prediction records. It supports consultation by allowing doctors to analyze results and provide medical guidance.

Admin Module: The admin module manages system operations such as user control, monitoring activities, and maintaining data integrity. It ensures smooth functioning and security of the platform.

Database Module: The database module stores user data, symptoms, and prediction results using PostgreSQL. It ensures reliable data management and efficient retrieval of information.

TECHNOLOGIES USED

Programming Language: Python

Web Framework: Django

Database: PostgreSQL

Machine Learning Algorithm: Random Forest

Machine Learning Library: Scikit-learn

Frontend Technologies: HTML, CSS, Bootstrap

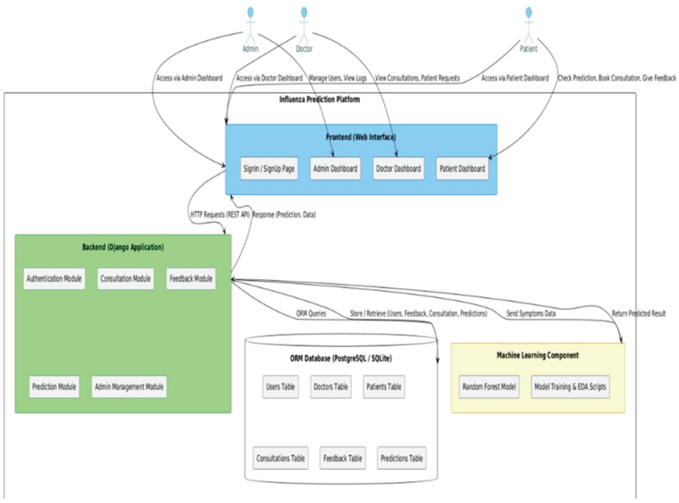
Development Tools: Visual Studio Code

Browser: Google Chrome

Advantages

The proposed influenza prediction system offers several advantages over traditional disease monitoring methods by integrating machine learning and web-based technologies. It provides improved prediction accuracy using the Random Forest algorithm, which effectively handles complex and high-dimensional healthcare data. The system enables real-time prediction, allowing users to receive immediate results based on their symptoms, which supports early detection and timely preventive measures. The user-friendly interface ensures easy interaction for both patients and healthcare professionals, while automation reduces manual effort and minimizes human errors. Additionally, the use of PostgreSQL ensures secure and reliable data storage, maintaining data integrity and consistency. The integration of modules such as patient, doctor, admin, and prediction enhances overall system functionality, making it a complete healthcare solution. Furthermore, the system is scalable and can be extended to support multiple diseases and larger datasets, making it highly useful for future healthcare applications.

System Architecture



Results

The implementation of the proposed system demonstrates significant improvements in influenza prediction accuracy and efficiency. The Random Forest model effectively classifies influenza risk levels based on user-input symptoms, providing reliable and consistent results. The system responds quickly to user inputs, ensuring real-time prediction capabilities. Additionally, the integration of Exploratory Data Analysis helps in understanding underlying data patterns, contributing to improved model performance. The overall system successfully combines data analysis, machine learning, and web technologies to deliver accurate predictions and enhance user experience. These results highlight the effectiveness of the proposed approach in supporting healthcare monitoring and early disease detection.

Conclusion

This project presents a machine learning-based approach for analyzing and predicting influenza trends using Exploratory Data Analysis and the Random Forest algorithm. The integration of predictive modeling with a web-based platform provides a practical and user-friendly solution for influenza monitoring. The system enables users to receive real-time predictions based on symptoms, thereby supporting early detection and preventive healthcare measures. By combining data analytics, machine learning, and web technologies, the project demonstrates how modern computational tools can enhance healthcare decision-making and disease surveillance. The proposed system is scalable, efficient, and capable of supporting future advancements in predictive healthcare applications.

Future scope

The proposed system can be further enhanced by integrating real-time healthcare data from hospitals and government health organizations to improve prediction accuracy and enable dynamic monitoring of influenza outbreaks. Advanced machine learning techniques, including deep learning models, can be incorporated to enhance predictive performance. The development of a mobile application would improve accessibility and user engagement by allowing users to access the system through smartphones. Additionally, the system can be extended to support prediction of multiple diseases such as dengue, malaria, and COVID-19. Cloud deployment can also be considered to improve scalability, performance, and availability for large-scale usage. These enhancements would further strengthen the system and expand its applicability in real-world healthcare scenarios.

References

1. L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
2. F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
3. F. Chollet, Deep Learning with Python, 1st ed., Manning Publications, 2018.
4. World Health Organization, “Influenza (Seasonal),” 2023. [Online]. Available: https://www.who.int
5. Centers for Disease Control and Prevention, “Influenza Symptoms and Diagnosis,” 2023. [Online]. Available: https://www.cdc.gov
6. T. M. Mitchell, Machine Learning, New York: McGraw-Hill, 1997.
7. I. Sommerville, Software Engineering, 10th ed., Pearson Education, 2016.
8. R. S. Pressman and B. R. Maxim, Software Engineering: A Practitioner’s Approach, 8th ed., McGraw-Hill Education, 2015.
9. Django Software Foundation, “Django Documentation,” 2024. [Online]. Available: https://docs.djangoproject.com
10. PostgreSQL Global Development Group, “PostgreSQL Documentation,” 2024. [Online]. Available: https://www.postgresql.org/docs/
11. S. Santillana et al., “Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance,” PLoS Computational Biology, vol. 11, no. 10, 2015.
12. Y. Yang, M. Lipsitch, and J. Shaman, “Inference of Seasonal Influenza Transmission Dynamics,” Proceedings of the National Academy of Sciences, vol. 112, no. 9, pp. 2723–2728, 2015.