DOI: 10.36297/vw.applsci.v8i2.505 ISSN: 2582-5615

VW Applied Sciences, Volume : 8, Issue: 2, 18-20

Machine Learning Approaches for Early Detection and Control of Emerging Infectious Diseases

Mehak Sharma[1*] Zubair Mir[2*] Ankit Verma[3*]

[1]Staff Department of Biotechnology, Himachal Pradesh University, Shimla, India
[2]Department of Microbiology, University of Kashmir, Srinagar, India
[3]Department of Computer Science, Guru Ghasidas Vishwavidyalaya, Bilaspur, India
email: mehak.s@hpu.ac.in, zubairmir@uok.ac.in, ankit.v@ggv.ac.in

Received: Feb 17, 2026
Accepted: Mar 18, 2026
Published online: Apr 22, 2026

Abstract: The rapid emergence and re-emergence of infectious diseases pose a significant threat to global public health, necessitating the development of advanced detection and control mechanisms. Traditional epidemiological approaches often rely on delayed reporting and limited data integration, resulting in slow response times and increased transmission risk. This study explores the application of machine learning techniques for the early detection and control of emerging infectious diseases. By integrating heterogeneous datasets, including clinical records, genomic data, environmental variables, and mobility patterns, the research develops predictive models capable of identifying outbreaks at an early stage. Algorithms such as logistic regression, random forest, support vector machines, and deep neural networks are evaluated for their effectiveness in disease prediction and classification. The results demonstrate that machine learning models significantly improve detection accuracy and response time compared to conventional methods. Furthermore, the study highlights the role of predictive analytics in guiding public health interventions, including resource allocation and containment strategies. The findings underscore the potential of machine learning in transforming infectious disease surveillance into a proactive, data-driven system capable of mitigating global health risks.

Keywords: Machine Learning, Infectious Diseases, Early Detection, Epidemiology, Predictive Analytics

1. Introduction

Emerging infectious diseases represent one of the most pressing challenges in global health, with recent outbreaks highlighting the vulnerability of healthcare systems worldwide. Diseases such as COVID-19, Ebola, and Zika have demonstrated the rapid pace at which pathogens can spread across populations, often outpacing traditional surveillance and response mechanisms. The increasing interconnectedness of global populations, coupled with environmental changes and urbanization, has further accelerated the emergence and transmission of infectious diseases [1]. Traditional epidemiological methods rely heavily on manual data collection, laboratory testing, and retrospective analysis, which can result in significant delays in detecting outbreaks. These delays can have severe consequences, allowing diseases to spread unchecked and increasing the burden on healthcare systems. The need for rapid, accurate, and scalable detection methods has led to growing interest in the application of machine learning techniques in epidemiology [2]. Machine learning provides powerful tools for analyzing complex datasets and identifying patterns that may not be immediately apparent through conventional methods. By leveraging large volumes of data from diverse sources, machine learning models can detect early signals of disease outbreaks and predict their progression. These capabilities enable public health authorities to implement timely interventions and reduce the impact of infectious diseases.

2. Background and Epidemiological Context

The emergence of infectious diseases is influenced by a combination of biological, environmental, and social factors. Pathogen evolution, climate change, and human mobility all contribute to the spread of diseases. Understanding these factors is essential for developing effective detection and control strategies. Data-driven approaches have become increasingly important in epidemiology, enabling the integration of diverse datasets such as clinical records, environmental data, and population mobility patterns. The use of digital health

DOI: 10.36297/vw.applsci.v8i2.505 ISSN 2582-5615

technologies and electronic health records has further expanded the availability of data for analysis [3]. However, the complexity and volume of these datasets require advanced analytical tools capable of extracting meaningful insights. Machine learning techniques offer a solution by enabling the analysis of high-dimensional data and the identification of complex relationships among variables. These techniques can be used to develop predictive models that estimate disease risk, identify hotspots, and forecast outbreak dynamics.

3. Literature Review

Recent studies have demonstrated the effectiveness of machine learning in infectious disease prediction and surveillance. Logistic regression models have been widely used for disease classification due to their simplicity and interpretability [4]. However, more advanced algorithms such as random forest and support vector machines have shown superior performance in handling nonlinear relationships and high-dimensional data [5]. Deep learning techniques, including convolutional neural networks and recurrent neural networks, have been applied to analyze medical images and time-series data, respectively. These models have achieved high accuracy in disease detection and prediction, particularly in complex scenarios involving large datasets [6]. The integration of genomic data into predictive models has also gained attention, enabling the identification of pathogen mutations and transmission patterns. Studies have shown that combining genomic data with machine learning techniques can enhance the accuracy of outbreak prediction [7].

4. Data Sources and Feature Engineering

The effectiveness of machine learning models depends on the quality and diversity of the data used for training. This study utilizes multiple data sources, including clinical records, laboratory test results, environmental data, and mobility patterns. Feature engineering techniques are applied to extract relevant variables and improve model performance. Environmental factors such as temperature, humidity, and rainfall are included as features due to their influence on disease transmission. Mobility data, derived from transportation networks and mobile devices, provides insights into population movement patterns, which are critical for understanding disease spread [8].

5. Methodology

The study employs a supervised learning approach to develop predictive models for infectious disease detection. The dataset is divided into training and testing sets, and cross-validation techniques are used to ensure model robustness. Four machine learning algorithms are evaluated: logistic regression, random forest, support vector machines, and deep neural networks. Each model is trained using the same dataset and evaluated based on performance metrics such as accuracy, precision, recall, and F1-score. Hyperparameter tuning is performed to optimize model performance, and ensemble techniques are used to combine the strengths of multiple models. The final model is selected based on its ability to achieve high accuracy while maintaining interpretability.

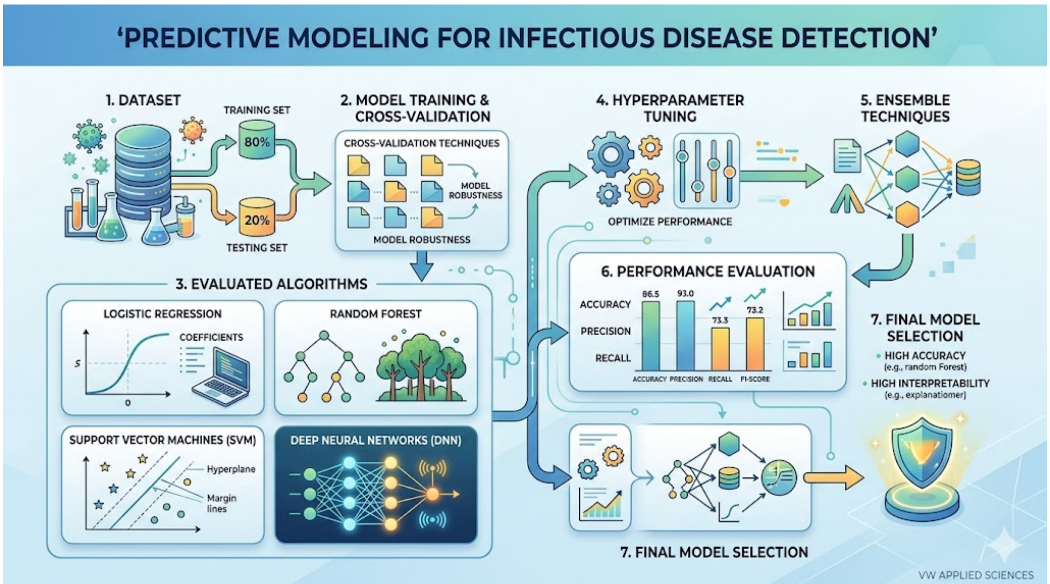

Fig 1

6. Results and Analysis

 DOI: 10.36297/vw.applsci.v8i2.505 ISSN 2582-5615

The results indicate that machine learning models significantly improve the early detection of infectious diseases compared to traditional methods. The random forest model achieves the highest accuracy, consistent with findings in previous studies [5]. Support vector machines demonstrate strong performance in classification tasks, particularly in identifying disease-positive cases. Deep learning models show superior performance in analyzing complex datasets, including time-series and image data. These models are particularly effective in detecting patterns that are not easily captured by traditional algorithms [6].

7. Discussion

The findings highlight the transformative potential of machine learning in infectious disease surveillance. By enabling early detection and accurate prediction, machine learning models can support timely interventions and reduce the spread of diseases. However, challenges such as data privacy, ethical considerations, and model interpretability must be addressed. Ensuring the security and confidentiality of health data is critical for maintaining public trust. Additionally, the complexity of machine learning models can make them difficult to interpret, which may limit their acceptance among healthcare professionals.

8. Public Health Implications

The integration of machine learning into public health systems can significantly enhance disease surveillance and response capabilities. Predictive models can be used to identify high-risk populations, allocate resources efficiently, and design targeted intervention strategies. The use of real-time data and predictive analytics can also improve the effectiveness of vaccination campaigns and containment measures, reducing the overall impact of infectious diseases.

9. Conclusion

This study demonstrates the effectiveness of machine learning approaches in the early detection and control of emerging infectious diseases. By leveraging diverse datasets and advanced algorithms, the proposed framework enhances the accuracy and timeliness of disease prediction. The findings underscore the importance of integrating machine learning into public health systems to improve resilience against future outbreaks.

References

1. D. Morens et al., “Emerging infectious diseases,” Nature, vol. 430, pp. 242–249, 2004.
2. C. Viboud et al., “The RAPIDD Ebola forecasting challenge,” Epidemics, vol. 22, pp. 13–21, 2018.
3. J. Bates et al., “Big data in health care,” Health Affairs, vol. 33, pp. 1137–1145, 2014.
4. D. Hosmer et al., “Applied logistic regression,” Wiley, 2013.
5. L. Breiman, “Random forests,” Machine Learning, vol. 45, pp. 5–32, 2001.
6. Y. LeCun et al., “Deep learning,” Nature, vol. 521, pp. 436–444, 2015.
7. N. Grubaugh et al., “Genomic epidemiology,” Nature Reviews Microbiology, vol. 17, pp. 707–720, 2019.
8. M. Wesolowski et al., “Human mobility and disease spread,” Science, vol. 338, pp. 267–270, 2012.
9. T. Chen and C. Guestrin, “XGBoost,” KDD, pp. 785–794, 2016.
10. A. Esteva et al., “Medical AI,” Nature Medicine, vol. 25, pp. 24–29, 2019.
11. S. Funk et al., “Real-time forecasting,” PLoS Computational Biology, vol. 15, 2019.
12. J. Shaman et al., “Forecasting influenza,” PNAS, vol. 110, pp. 20425–20430, 2013.