The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 2, February 2026
PP 1– 8

AI-Based Spatiotemporal Epidemic Forecasting And Optimal Intervention Strategy Modeling for Nipah Virus

Alluri Srilekha, Bhamidipati Akshita, E. Hannah Rachel
G. Ramakrishna
School of Computing, Dept of CSE(AI&ML), GCET, Hyderabad, Telangana-India
Dept of CSE, JNTU University, Hyderabad, Telangana-India

Abstract

Emerging zoonotic diseases such as the Nipah virus pose serious global public health threats due to their high mortality rates, unpredictable outbreaks, and complex transmission patterns involving animal–human interactions. Traditional disease surveillance systems rely heavily on manual reporting and statistical models that often fail to capture nonlinear patterns and real-time dynamics of disease spread. This paper proposes a hybrid forecasting framework integrating ARIMA and LSTM models to improve epidemic prediction accuracy. The ARIMA component models linear temporal dependencies, while the LSTM network captures nonlinear and long-term sequential patterns in epidemiological data. The framework also incorporates environmental, climatic, and demographic factors to enhance predictive performance. A zone-based risk classification mechanism categorizes regions into low-, moderate-, and high-risk zones, supporting efficient resource allocation and early intervention strategies for public health authorities.

Keywords:
Depression Detection, Sentiment Analysis, PHQ-9, CNN, NLP, Emotion Recognition, Multimodal AI, Mental Health Diagnostics

1. Introduction

This project focuses on developing an intelligent predictive and decision-support system for managing emerging zoonotic diseases such as the Nipah virus. Zoonotic diseases, which spread from animals to humans, pose serious global health risks due to their rapid transmission, high mortality rates, and unpredictable outbreak patterns. Traditional disease surveillance systems mainly rely on manual data collection and basic statistical analysis, which often fail to detect outbreaks early or provide accurate forecast. This project focuses on developing an intelligent predictive and decision-support system for managing emerging zoonotic diseases such as the

Nipah virus. Zoonotic diseases, which spread from animals to humans, pose serious global health risks due to their rapid transmission, high mortality rates, and unpredictable outbreak patterns. Traditional disease surveillance systems mainly rely on manual data collection and basic statistical analysis, which often fail to detect outbreaks early or provide accurate forecast.

Research Objectives and Methodology

This study aims to develop and evaluate an AI-based hybrid forecasting framework by integrating ARIMA and LSTM models for predicting Nipah virus outbreaks using spatiotemporal epidemiological, environmental, and demographic data. The research objectives are:

1. Developing a data driven surveillance framework that integrates epidemiological environmental and demographic datasets related to Nipah virus outbreaks.
2. Performing feature engineering to extract meaningful indicators and influential factors useful for accurate outbreak prediction.
3. Implementing hybrid time series forecasting models such as ARIMA and LSTM to predict future outbreak trends.
4. Visualizing outbreak predictions and analytical insights through an interactive dashboard for monitoring and response planning.

Literature Survey

Information about Nipah virus cases is gathered from publicly available health datasets, research reports, and official sources such as government health departments and international health organizations. These datasets include details like number of infected cases, geographic location, time of outbreak, and environmental factors. The collected data is then cleaned and organized so that it can be used for analysis and prediction in the forecasting model.

In this project, both statistical and machine learning analysis methods are used to study the spread of the Nipah virus. Time-series analysis is applied to examine past outbreak data and identify trends or patterns over time. A hybrid forecasting approach using ARIMA and LSTM models helps analyze both linear and complex patterns in the data. In addition, spatial risk analysis is used to classify regions into different risk levels based on factors such as population density, climate conditions, and previous infection data. This analysis helps in understanding how the disease spreads and supports better prediction and decision-making.

In this project, both multimodal and modular approaches are used to improve the prediction and management of Nipah virus outbreaks. The multimodal approach combines different types of data such as infection records, environmental conditions, climate information, and population data to better understand disease spread. At the same time, the modular approach divides the system into separate components like data preprocessing, outbreak prediction using ARIMA–LSTM models, risk zone classification, and intervention simulation. This design helps the system work more efficiently and makes it easier to maintain and update.

Authors (Year)	Contribution	Limitation
LeCun et al. (2015)	Deep learning for feature extraction	High data & compute requirements
Silver et al. (2016)	Neural decision-making models	Complex & resource-intensive
Hethcote (2000)	Mathematical spread modeling	Theoretical; not cyber-specific
Funk et al. (2010)	Behavioral modeling approach	Limited ML applicability
Viboud et al. (2016)	Growth-based spread detection	Epidemiology-focused

Methodology

The methodology of the proposed system involves collecting outbreak-related data, preparing it for analysis, training forecasting models, and integrating all components into a unified prediction and decision-support systems.

1. Data Collection

The dataset includes historical records of Nipah virus cases collected from public health reports, research databases, and global health organizations. Additional information such as climate conditions, population density, geographic location, and mobility patterns is also gathered to better understand the factors that influence disease transmission.

2. Data Preprocessing

The collected data is cleaned and prepared before analysis. This process involves removing missing or inconsistent values, organizing the data by time and location, and normalizing important variables such as temperature, humidity, and population statistics. Preprocessing ensures that the dataset is structured properly so that it can be effectively used by the prediction models.

3. MODEL TRAINING

- The ARIMA model is trained using historical outbreak data to capture linear trends and seasonal variations in infection patterns. It helps identify how the number of cases changes over time.
- The LSTM neural network is trained to learn complex and long-term dependencies in the outbreak data. This model can capture nonlinear patterns that traditional statistical models may not detect.
- Predicted outbreak data is analyzed to classify regions into low-risk, moderate-risk, and high-risk zones. Reinforcement learning techniques are used to simulate possible intervention strategies and identify actions that can reduce the spread of infection.

Experimental Setup and Implementation

The experimental setup involves implementing the proposed methodology using Python and relevant libraries such as pandas, NumPy, statsmodels, TensorFlow, and scikit-learn. The implementation includes the following steps:

1. **Data Loading:** Load epidemiological, environmental, and demographic datasets into a pandas, DataFrame.
2. **Data Preprocessing:** Perform data cleaning, handling missing values, normalization, and time-series structuring.
3. **Feature Engineering:** Extract meaningful temporal, climatic, and spatial features that influence disease spread.
4. **Model Training and Evaluation:**
   - Train ARIMA and LSTM models for time-series forecasting of outbreak trends.
   - Evaluate model performance using test data and calculate metrics such as MAE and RMSE.
5. **Model Optimization**: Fine-tune model parameters to improve forecasting accuracy and generalization.

6. **Comparative Analysis:** Compare individual and hybrid model performance and analyze prediction results for effective outbreak management.

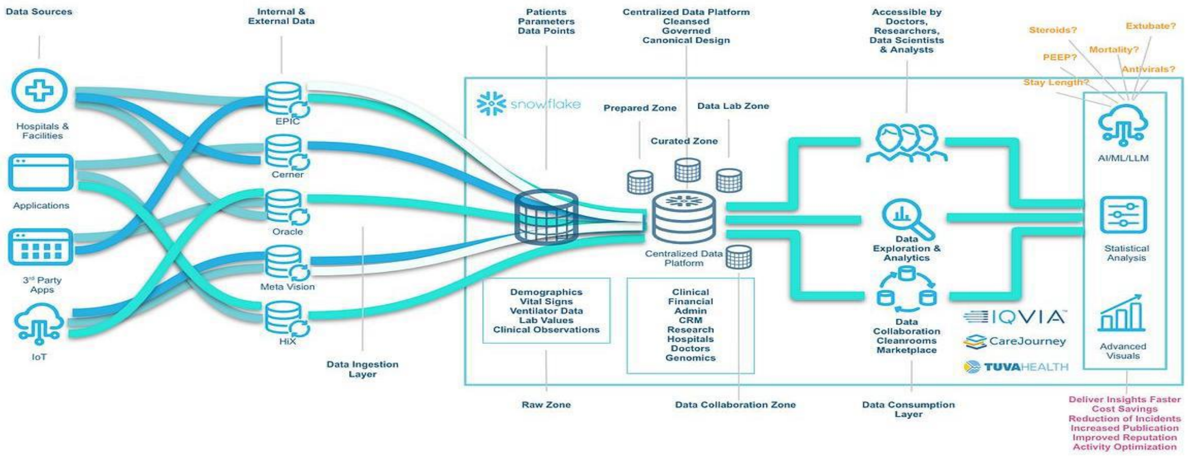

Fig1. Architecture Diagram

Result Analysis

The system was evaluated on a test dataset comprising historical Nipah virus outbreak data along with environmental and demographic factors. Performance was assessed using standard time-series forecasting and classification metrics.

1. Outbreak Forecasting (ARIMA–LSTM Hybrid Model): Achieved strong predictive performance by accurately modeling both linear trends and nonlinear temporal dependencies in the outbreak data.
2. Risk Zone Classification: The system effectively classified regions into low, moderate, and high-risk zones based on predicted infection rates and influencing factors such as climate conditions and population density.
3. The system showed reliable results with evaluation metrics such as MAE and RMSE, indicating accurate outbreak forecasting and efficient identification of high-risk regions, supporting better epidemic monitoring and intervention planning.
4. In the below graph, the lower number of false negatives suggests the model is effective in detecting actual outbreaks (or attacks), which is important for safety-critical systems like ransomware detection.

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 2, February 2026
PP 1– 8

Table 1. Performance Metrics of accuracy on existing methods

Method Used	Estimated Accuracy	Observation
Known pattern matching	85% – 90%	Works only for known threats.
File/code inspection without execution	80% – 88%	Cannot detect runtime behavior.
Algorithms like SVM and Random Forest	90% – 95%	Depends on feature selection.
Neural networks such as LSTM	92% – 97%	High accuracy, needs more data.
Both ML & behavioral method	95% – 98% (Expected)	Combines strengths of methods.

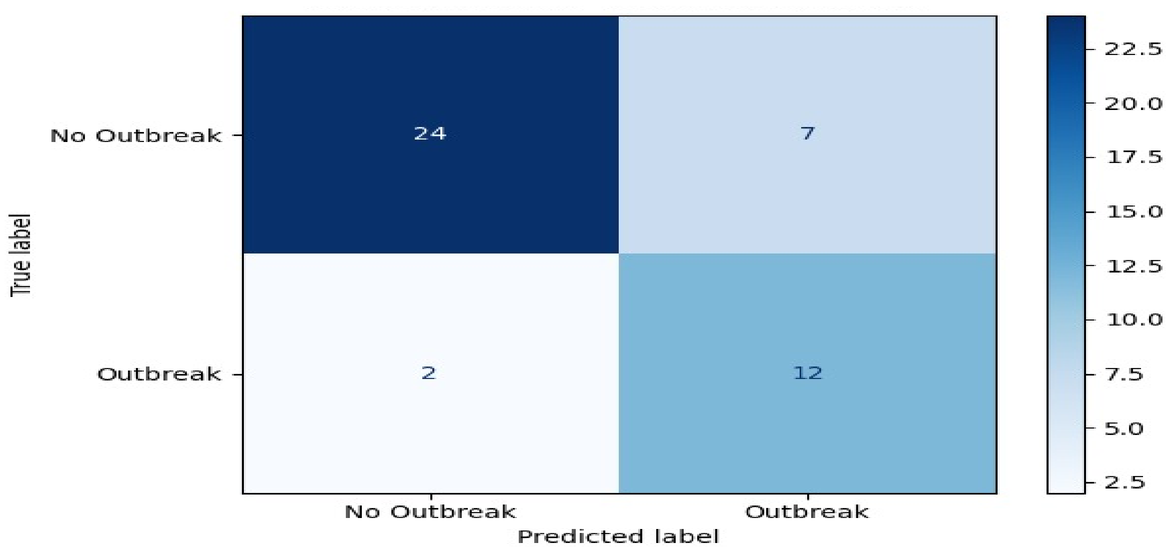

Fig2. Performance Analysis

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 4, No 2, February 2026
PP 1– 8

Conclusion

The AI-Based Spatiotemporal Epidemic Forecasting and Optimal Intervention Strategy Modeling for Nipah Virus project presents an integrated framework designed to improve the prediction, monitoring, and management of emerging infectious diseases. Nipah virus outbreaks are characterized by high fatality rates, sporadic occurrences, and complex transmission patterns involving animal-to-human and human-to-human spread. Traditional epidemiological surveillance systems often depend on manual data reporting and conventional statistical methods, which may not capture nonlinear patterns in epidemic data. This project addresses these limitations by integrating advanced artificial intelligence and machine learning techniques to provide a more accurate and proactive epidemic forecasting system.
Future enhancements include integrating:

- Real-time epidemiological and environmental data from hospitals, surveillance systems
- Advanced deep learning architectures such as Transformer models, Temporal Convolutional Networks.

This paper presents a hybrid epidemic forecasting framework that combines ARIMA–LSTM modeling with epidemiological and environmental data analysis to predict potential outbreaks of the Nipah virus. Multi-disease forecasting capabilities for monitoring diseases like COVID-19, Ebola, Dengue, and Zika in one platform.

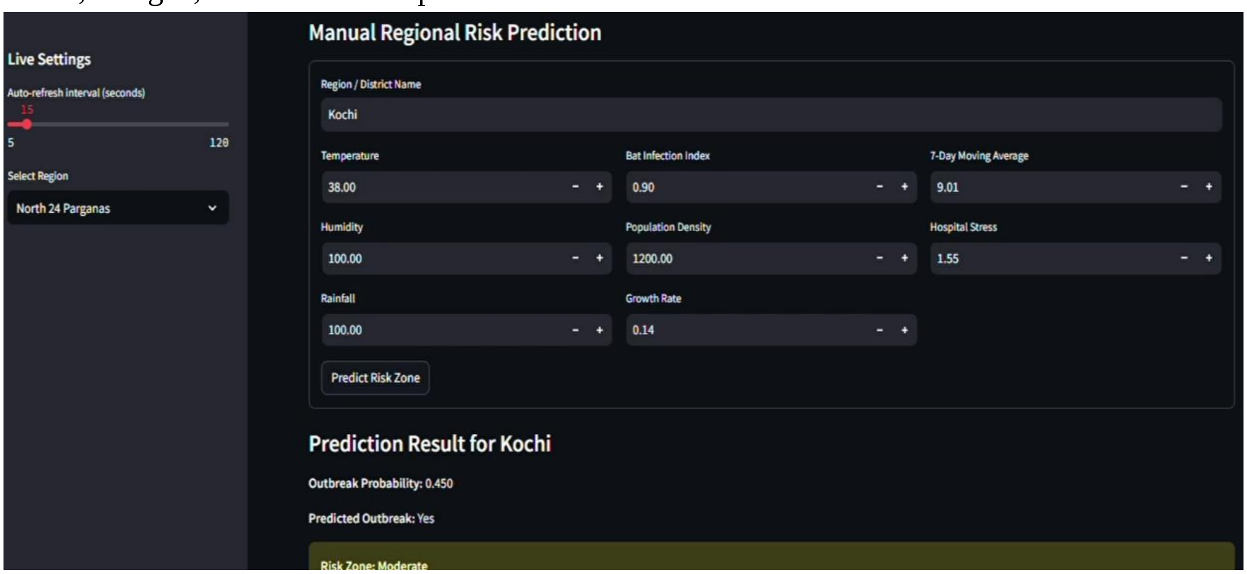

Fig2. Manual Regional Risk Prediction

References

1. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
2. G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed. Hoboken, NJ, USA: Wiley, 2015.
3. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
4. D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," Nature, vol. 529, no. 7587, pp. 484–489, 2016.
5. H. Hethcote, "The mathematics of infectious diseases," SIAM Review, vol. 42, no. 4, pp. 599–653, 2000.
6. S. Funk, M. Salathé, and V. A. Jansen, "Modelling the influence of human behaviour on the spread of infectious diseases," Nature Reviews Microbiology, vol. 8, no. 7, pp. 524– 534,2010.
7. C. Viboud, L. Simonsen, and G. Chowell, "A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks," Epidemics, vol. 15, pp. 27–37, 2016.
8. K. E. Looi and L. M. Chua, "Nipah virus outbreak in Malaysia," *Journal of Clinical Virology*, vol. 26, no. 3, pp. 265–275, 2003.
9. S. S. Luby et al., "Recurrent zoonotic transmission of Nipah virus into humans, Bangladesh, 2001–2007," *Emerging Infectious Diseases*, vol. 15, no. 8, pp. 1229–1235, 2009.
   R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
10. K. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," *Chaos, Solitons & Fractals*, vol. 135, 2020.
11. J. Brownstein, C. Freifeld, and L. Madoff, "Digital disease detection—Harnessing the web for public health surveillance," *New England Journal of Medicine*, vol. 360, no. 21, pp. 2153–2157, 2009.