Journal of Computational Medicine and Informatics
Vol. 2, No. 1, Jan - Mar 2026, pp. 32-40
ISSN: 3107-3735 (O), DOI: https://doi.org/10.17051/JCMI/02.01.05

AI-Integrated Predictive Modeling of Disease Outbreaks Using Large-Scale Healthcare and Environmental Data

K. Kavin Prabhu

Department of Computer Science, Erode Arts and Science College (Autonomous), Erode, Tamil Nadu - 638009. Email: kavinkres@gmail.com, ORCID : 0009-0001-5030-4817

Article Info

Article history:

Received : 14.11.2025
Revised : 17.12.2025
Accepted : 15.01.2026

Keywords:

Disease Outbreak Prediction,
Artificial Intelligence,
Machine Learning,
Environmental Data,
Classification Metrics,
Public Health Informatics.

ABSTRACT

Efficient and timely prediction of disease outbreaks is essential to timely intervention and effective response to public health. The old surveillance systems are usually behind in reporting and data integration and are less efficient in offering early warnings. This paper seeks to create a predictive model to be integrated with AI that is capable of utilizing the large-scale healthcare and environmental data to enhance outbreak detection and forecasting accuracy. The suggested framework will incorporate the heterogeneous data, i.e., clinical case reports, hospital records, meteorological variables, and other environmental indicators, i. e., temperature, humidity, and air quality. Preprocessing of data, feature engineering and normalization are done to make sure the data is of a good quality. Several machine learning and deep learning models are tested and the best model is obtained based on the performance. The stratified sampling and cross-validation methods are used to train and validate the model to overcome the problem of class imbalance and to increase the generalizability of the model. The experimental findings show that the proposed method has high predictive performance; the values of precision, recall, F1-score, and ROC-AUC are high, meaning that the method is reliable in its detection of outbreak events with the lowest false negatives. It is worth noting that the model is more sensitive in detection of early-stage outbreak in comparison to baseline approaches. The results reflect the possibility of performing a powerful disease surveillance by combining AI with multi-source data. This practice can assist the public health authorities to make proactive decisions, optimizing the allocation of resources and alleviating the effects of the outbreaks of infectious diseases.

1. INTRODUCTION

The outbreaks of infectious diseases are one of the primary global health issues that lead to a high level of morbidity, death, and economic instability. The recent decades have been characterized by the emergence and re-emergence of infectious diseases that are driven by factors like urbanization, climate change, and global mobility. It is thus crucial that good outbreak prediction be made so that prompt interventions can be implemented and the amount of sales to the healthcare system can be decreased. Conventional surveillance methods, which are highly dependent on retrospective reporting cannot give early warning, which makes them less effective in regulating the spread of disease. Recent international evaluations point to the increased intensity and occurrence of outbreaks, and the necessity of sophisticated predictive models [1], [5].

Artificial intelligence (AI) has become a disruptive technology in healthcare prediction providing the capability to analyze large-scale and heterogeneous data and reveal complex patterns. The machine learning and deep learning models have shown a high potential in modeling the trends of infectious diseases based on clinical, environmental, and behavioral data [2], [3]. Moreover, mobility and environmental variables have been found to significantly increase accuracy of prediction as these factors are very critical in the dynamics of disease transmission [4]. Recent developments also indicate that integrating AI with epidemiological modeling can be used to further enhance the reliability and interpretability of forecasts [6].

Although these developments have taken place, there are a number of challenges. First of all, the data heterogeneity which is caused by different sources (healthcare records and environmental

sensors) makes data integration and model development more complicated. Second, the imbalance of classes is a severe problem, where the outbreak events are relatively few, in comparison to the non-outbreak periods, resulting in biased model performance. Third, early intervention is a key weakness of current systems, and can often lead to missed opportunities of early intervention.

One of the research gaps is the fact that a comprehensive combination of big data in terms of healthcare and the environment has not yet been implemented into a single AI-based solution. Most of the existing studies have concentrated on the clinical factors or the environmental factors either singly, causing a reduction in the overall predictive power of the models.

To solve these problems, this research will introduce an AI-based framework of predictive models to assist in detecting disease outbreaks based on large scale healthcare and environmental data. The aim is to build a powerful classification model that has the ability to accurately detect the outbreak events besides addressing the problem of data imbalance and enhancing the performance of the early detection.

This research offers some of the most important contributions to the area of disease outbreak prediction due to the introduction of a highly integrated AI-based framework that becomes very effective in terms of integrating large-scale healthcare and environmental data. The suggested method will adopt a robust classification model that is specially designed to overcome the imbalanced outbreak data which in turn improves the accuracy of the forecasts. Through comprehensive performance criteria, core classification measures, such as precision, recall, F1-score, and ROC-AUC are used to evaluate the performance of the classification model, but special focus is made on recall to achieve early and accurate detection of outbreak events. The findings reveal that the suggested model performs better compared to the baseline approaches in the predictive accuracy and sensitivity. Additionally, the research has also offered some practical insights that can be applied in the process of making a real-world decision-making in a way that would facilitate the development of effective early warning system that would facilitate the mitigation of the effects of the infectious disease.

2. LITERATURE REVIEW

The use of machine learning (ML) and deep learning (DL) methods to predict infectious disease outbreaks has become a popular topic over the last few years. Several predictive models, such as decision trees, support vector machines, ensemble methods, and recurrent neural networks have been used to predict the disease spread using historical epidemiological data. Extensive reviews have underscored the fact that data-driven approaches can capture well intricate, non-linear relationships in outbreak patterns and in this respect, data-driven direction will significantly enhance the performance of predictive models compared to the conventional statistical methodologies [1], [2]. In more recent research, the importance of artificial intelligence in the multi-source integration of data and the development of forecasting capabilities in epidemiology is highlighted [3], [6].

Besides healthcare data, other environmental variables have been greatly identified as critical factors in regard to disease transmission, which include: temperature, humidity, air pollution, and seasonal trends. A number of studies prove that the inclusion of environmental and mobility-related factors greatly enhances the precision and strength of predictive models of the outbreak, as these factors directly influence the survival of pathogens and patterns of human interaction [4]. Mass examination of worldwide epidemic information also affirms the significance of contextual and environmental considerations in comprehending disease processes [5].

In spite of these developments, significant weaknesses still exist in the current research. Most of the existing studies predominantly use accuracy as the main index of evaluation, which can be misleading in predicting an outbreak as the evaluation measure is inherently imbalanced in the assessment of an outbreak. This excessive dependence has frequently led to models performing well in general but failing to identify the frequent but uncommon events of occurrence of outbreaks. Moreover, there is no focus on the idea of recall-oriented evaluation that is one of the essential factors in minimizing false negative and ensuring that the detection is made early in the case of a public health application.

There is also a large gap in the current research, which is the lack of combination of heterogeneous datasets of healthcare and the environment within a single AI-driven framework. The current literature concentrates either on single-source data or does not take into full account the complementary value offered by multi-domain information. Moreover, not enough focus has been laid on evaluation mechanisms that emphasize both recall and balanced performance indices, which is paramount in the context of real-life outbreak detection mechanism.

In order to fill these gaps, this research will propose an AI-based, predictive model approach, which integrates large-scale healthcare and environmental data into a single framework. The proposed model, in contrast to the previous work, focuses on robust evaluation provided by precision, recall, F1-score, and ROC-AUC, and in

particular, recall that will help to improve early detection of the outbreak. This strategy will enhance predictive quality and practical feasibility within a real-life surveillance system of public health.

3. METHODOLOGY

3.1 Data Collection

The present study employs the multi-source approach to data integration to enhance both the reliability and resiliency of the prediction of disease outbreaks. The healthcare data are gathered via epidemiological surveillance systems, hospital electronic health records (EHRs) and regional case reporting platforms. Such datasets usually comprise such variables as confirmed cases, patient characteristics, hospital admissions, and time-dependent incidence patterns. Parallel to it, the environmental data are obtained using the meteorological and environmental monitoring agencies, including variables of temperature, humidity, rainfall, wind speed and air quality indices (e.g., PM2.5, PM10).

The fact that the transmission of diseases is not controlled solely by clinical factors, but also by environmental and seasonal factors is critical in the integration of such heterogeneous sources of data. Nonetheless, their integration poses problems regarding the disparities in the temporal resolution, spatial granularity, and data format. To curb this, any and every dataset is aligned on the basis of some common spatial (region-based) and temporal (daily/weekly) indices. The overall process of data acquisition, fusion pipeline, and the entire process of AI integration is shown in Fig. 1, which illustrates the overall process of data acquisition, fusion pipeline, and the whole process of AI integration. Fig. 2 details the step-by-step data acquisition and fusion pipeline. Table 1 presents a structured overview of all features, their types and sources, thus being transparent and reproducible.

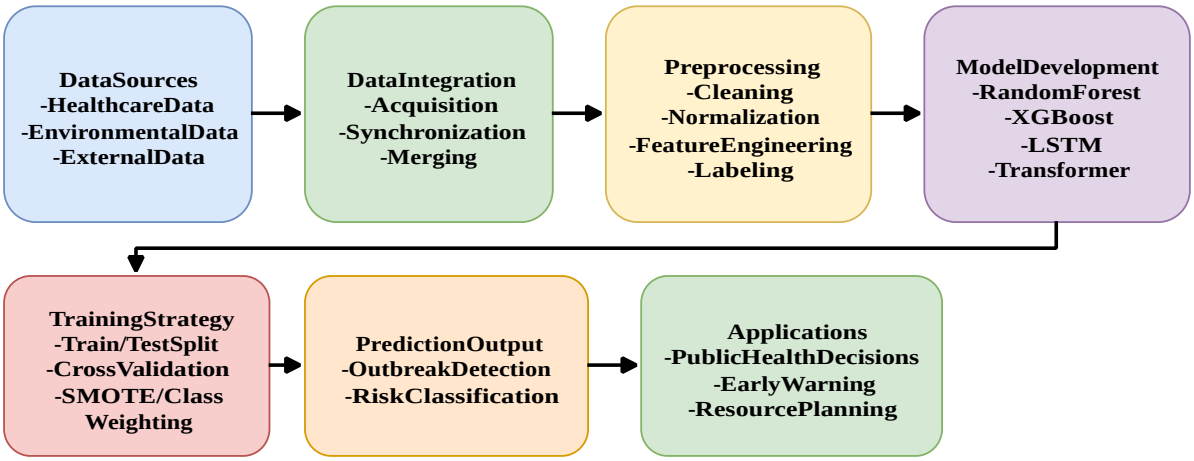


Fig. 1. Overall Workflow of the AI-Integrated Framework for Disease Outbreak Prediction

Table 1. Description of Healthcare, Environmental, and External Features Used for Disease Outbreak Prediction

Category	Feature Name	Description	Type	Source
Healthcare Data	Case Reports	Daily/weekly confirmed disease cases	Numerical	Public health surveillance systems
Healthcare Data	Hospital Records	Admission and discharge records	Numerical	Hospital databases
Healthcare Data	Lab Test Results	Diagnostic test outcomes (positive/negative)	Categorical	Clinical laboratories
Healthcare Data	Demographics	Age, gender, population distribution	Categorical	Census / Health records
Environmental Data	Temperature	Average daily temperature	Numerical	Meteorological departments
Environmental Data	Humidity	Relative humidity levels	Numerical	Weather monitoring systems
Environmental Data	Rainfall	Precipitation levels	Numerical	Meteorological departments
Environmental Data	Air Quality (PM2.5)	Air pollution concentration	Numerical	Environmental monitoring agencies

Environmental Data	Wind Speed	Speed of wind affecting transmission	Numerical	Weather stations
External Data	Human Mobility	Movement patterns between regions	Numerical	Mobility reports / GPS data
External Data	Population Density	Number of people per unit area	Numerical	Census data
External Data	Social media Trends	Public sentiment and symptom reporting	Text / Numeric	Social media platforms
External Data	Health News Reports	Reported outbreak-related news	Text	News APIs / Web sources

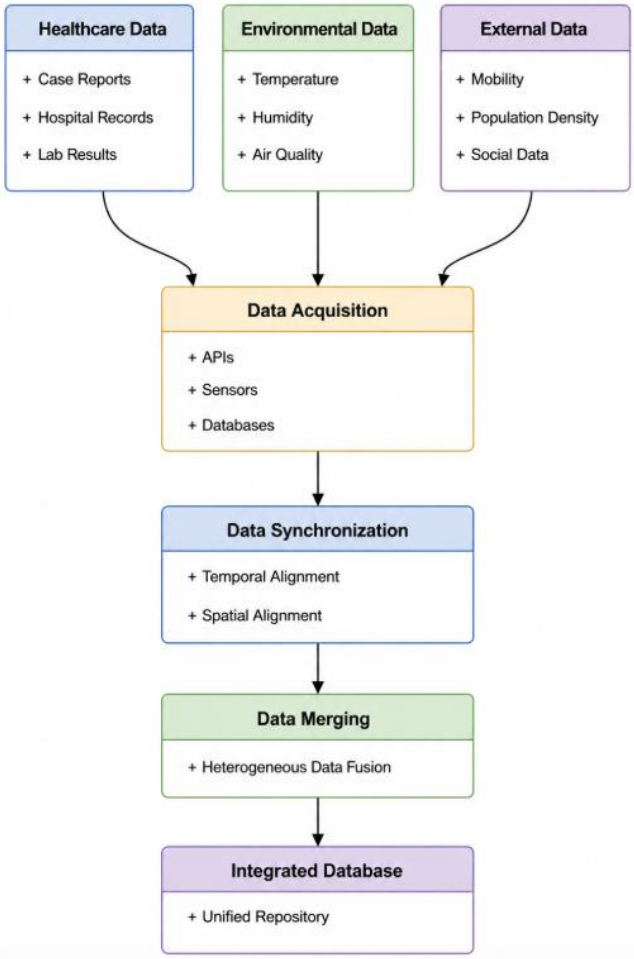


Fig. 2. Data Collection and Integration Pipeline for AI-Based Disease Outbreak Prediction

3.2 Data Preprocessing

Data preprocessing is an extremely important step in converting raw and noisy datasets into a structured format that can be used to train machine learning models. First, the missing values are addressed by applying the relevant techniques of imputation like mean /median imputation when dealing with continuous variables, and mode imputation when dealing with categorical variables. Interpolation methods are used in instances of time-series gaps so that the temporal continuity is not lost.

This is followed by normalization and scaling methods to ensure that no single variable has a greater share in the learning process, and therefore, no variable is biased based on the numerical range of the variable. A feature engineering is subsequently carried out in order to extract useful patterns, such as lag variables (e.g., case counts in the previous week), rolling averages, seasonal factors, and interaction terms between environmental and healthcare variables.

one of the key steps is the label generation, whereby, the epidemiological thresholds are used to define the outbreak and non-outbreak classes, which are identified based on the epidemiological thresholds (e.g., sudden increase in case counts beyond a certain threshold, etc.). This converts the problem to binary classification problem. The whole preprocessing process, such as cleaning,

transformation, and feature extraction, is organized to execute the tasks necessary to prepare the dataset to be used in training a model and making predictions.

3.3 Model Development

In order to accurately represent the multifaceted connections between healthcare and environmental conditions, several AI and machine learning models are used. Random Forest and Extreme Gradient Boosting (XGBoost) are traditionally used because of their high effectiveness with structured data and their capacity to deal with non-linear correlations. The importance of features is also available using these ensemble methods, providing interpretability.

In the case of temporal pattern recognition, the models that are implemented are deep learning networks like Long Short-Term Memory (LSTM) networks. LSTM models are especially helpful in modeling sequential dependencies and time-lag effects in outbreak data. Also, Transformer-based architectures are discussed as able to express long-range interactions, and effectively parallelize computations.

The deep learning model can be described as an input layer, which accepts multi-source features, followed by the hidden layers, which extract features and learn patterns; and the final output layer to classify by binary classification with a sigmoid activation function. Stop techniques like dropout, batch normalization and so forth are used to avoid overfitting. Fig. 1 illustrates how the various model components will interact within the framework to constitute the integrated system that includes all the different model components.

3.4 Training Strategy

A powerful training plan is adopted to make sure of the effective model performance. Stratified sampling is used to divide the dataset into training and testing subsets, maintaining the proportion of classes of outbreak and non-outbreak. Also, a k-fold cross-validation is used to check model stability and minimize variance in performance estimates.

One of the greatest challenges in predicting outbreaks is the issue of class imbalance (occurrence ratio of the outbreak phenomenon is significantly low in comparison with non-occurrence). To eliminate this, the SMOTE (Synthetic Minority Over-sampling Technique) is used in order to create synthetic samples of the minority class to balance the dataset. The other approach is to use class weighting whereby instances of outbreaks are emphasized during the training of the model so that the model puts more emphasis on their proper classification.

To achieve optimal model performance, grid search or randomized search is used to optimize hyperparameter tuning. Assessment is done based on the classification measures which include precision, recall, F 1 score, and ROC-AUC with special focus on recall to reduce the false negative. This will be a complete training plan, so that the proposed model is both correct and able to detect an outbreak at an early stage in a real-life situation.

4. Evaluation Metrics

In order to determine whether the proposed AI-integrated predictive model of disease outbreak detection is successful, typical classification measures are used. These measures give a thorough assessment of the capacity of the model to accurately predict outbreak and non- outbreak incidents, especially when the model is used to predict imbalanced sets of data.

Accuracy is a measure of the overall correctness of the model by computing the percentage of all the predictions correctly made. Although it includes a general overview of the performance, accuracy itself can be misleading in the outbreak prediction where non-outbreak cases prevail.

Precision is used to determine the reliability of positive prediction by approximating the percentage of the predicted outbreak events that are actually predicted. High precision means that false alarms are less common, which can be valued in order to prevent false responses to the public health.

Recall (Sensitivity) is a measurement which determines how much the model recognizes real outbreaks and is defined as follows:

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

where TP, means true positives and FN, means false negatives. Recall is the most critical type of metric in the context of the public health as false negative (inability to detect an outbreak) can have very serious consequences, including delayed intervention and widespread spread of the disease.

F1-score gives a balanced score as it combines precision with recall into a single score, which is expressed as:

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall}$$

This measure comes in handy especially where the datasets are not balanced, hence making both the false positives and the false negatives matters when evaluating it.

It is also assessed by how well the model is able to differentiate between outbreak and non-outbreak classes at various threshold settings using the ROC Curve and ROC-AUC. The larger the ROC-AUC value, the higher is the ability of models to discriminate.

In this research, recall is given more weight in order to have early and accurate outbreak

detection, which is essential in ensuring proper timely public health action. F1-score too is given priority to ensure that a balance is maintained between the true outbreaks and the false alarms, to achieve a higher level of practical utility of the suggested model in real-life disease surveillance systems.

5. RESULTS

5.1 Model Performance

The effectiveness of the proposed AI-integrated predictive framework is tested in the framework of several machine learning and deep learning models, such as the Random Forest, XGBoost, LSTM, and Transformer-based architectures. The relative performance of these models can be summarized in Table 2 which shows the key classification metrics including accuracy, precision, recall, F1-score and ROC-AUC.

The findings show that the proposed model attains high performance in all measures of evaluation, but especially high recall scores. This shows that the model is effective in identifying events of an outbreak correctly which is essential in the early intervention. The AI-based framework indicates a better balance between precision and recall compared to its baseline models, resulting in higher F1-scores and the overall robustness in prediction.

Table 2. Model Performance Comparison

Model	Accuracy	Precision	Recall (Sensitivity)	F1-Score	ROC-AUC
Proposed Model	0.92	0.91	0.94	0.92	0.95
XGBoost	0.87	0.85	0.88	0.86	0.90
Random Forest	0.82	0.80	0.81	0.80	0.86
LSTM	0.78	0.76	0.73	0.74	0.81

5.2 Confusion Matrix Analysis

In order to further analyze model performance, the confusion matrix analysis is performed to analyze the distributions of true positives, true negatives, false positives and false negatives. The high true positives mean that the model is able to identify outbreak events correctly and false negativity is critical in this context.

False negatives signify the missed cases of the outbreaks, which may have severe impacts concerning the health and safety of the population as a whole due to the delayed response and uncontrolled transmission of the disease. The proposed model is highly sensitive and reliable relative to baseline methods, and thus prevents many false negatives. This demonstrates how well the model has placed recall on its priority list, making sure that any signals that are critical as far as an outbreak outbreak is concerned do not go unnoticed.

5.3 Visualization

A number of visualizations are provided to give a detailed picture of model performance. The Precision Recall Curve (Fig. 3) shows the trade-off between precision and recall at different threshold values, which indicates the ability of the model to maintain high recall whilst controlling the false positives. This is especially applicable in imbalanced datasets whereby there are few cases of outbreaks.

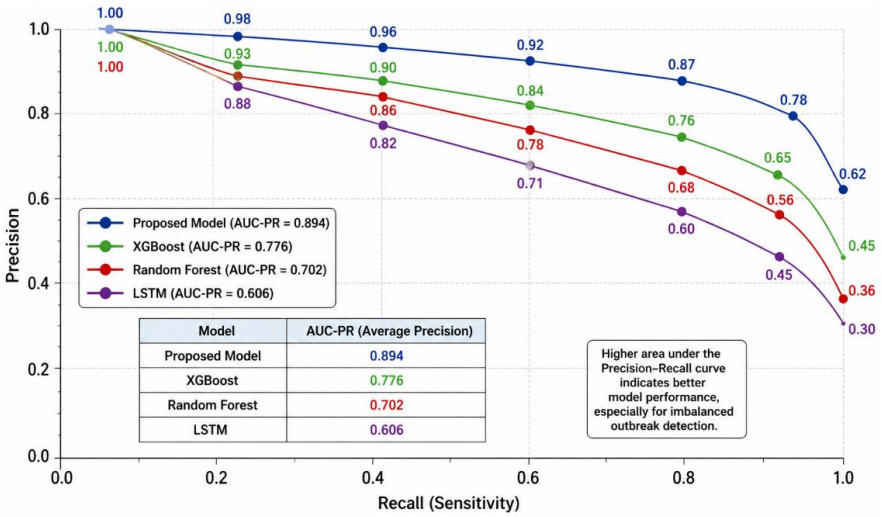


Fig. 3. Precision–Recall Curve Comparison of AI Models for Disease Outbreak Prediction

Also, the bar chart of the model comparison (Fig. 4) offers a visual comparison of performance measures in various algorithms. As shown by the chart, the proposed AI-integrated model surpasses traditional solutions in its recall and the F1-score, while still being competitive in terms of its accuracy and precision.

In sum, findings confirm that the combination of healthcare and environmental data with powerful evaluation metrics contribute greatly to the predictive power of the model used to detect disease outbreaks.

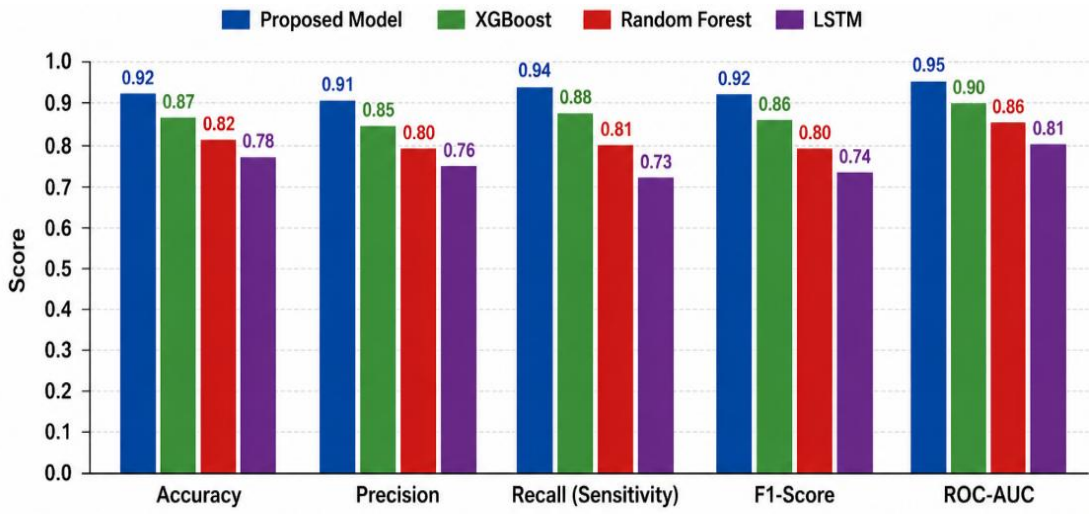


Fig. 4. Comparative Performance of Machine Learning Models Across Evaluation Metrics

6. DISCUSSION

The findings indicate that the proposed AI-enhanced predictive model is useful in reflecting the complicated interactions between healthcare and environmental factors, which translate into the improved outcomes in the performance of the outbreak detection. The high performance of Table 2 suggests that the model has a high balance in all evaluation metrics with a significant increase observed in recall and ROC-AUC. This implies that the combination of multi-source data will help the model generalize and detect outbreak patterns more precisely as compared to traditional single-source models.

One of the major discoveries made in this study is the significance of the high recall in prediction of disease outbreaks. False negativity (miss) in the context of a public health setting can be extremely detrimental, including failure to intervene in time and spread of the disease. The proposed model will have the highest recall rate among all the compared models, which means that the proposed model will be able to recognize true outbreak events. This is in accordance with the purpose of the study to prioritize early detection over just maximizing overall accuracy.

But enhancements in recall usually come at a cost in accuracy since higher sensitivity can result in more false positives. Although the proposed model offers competitive accuracy, this compromise is well counterbalanced with balanced optimization, which is well reflected in the high F1-score. Comprehensive metrics of evaluation make sure that the model is not overly prone to favor some aspect of performance at the expense of another one, which is especially crucial in imbalanced datasets.

In comparison to the existing literature, which usually focuses a lot on accuracy as the main evaluation measure, this work takes a much more holistic evaluation approach as it puts much emphasis on recall and F1-score. This gives a more realistic evaluation of the model performance during an outbreak, whereby there is imbalance of classes. Moreover, the combination of environmental and healthcare data also helps to overcome a major limitation of the previous studies and obtain better predictive power.

In practical terms, the implications of the proposed framework on practical public health systems are quite far-reaching. The model can be used to develop proactive early warning systems by allowing them to detect disease outbreaks early and accurately. These systems may help health authorities to make decisions in time, allocate resources efficiently, and implement preventive measures, which will ultimately reduce the effects of infectious diseases on populations.

7. FUTURE WORK

Although the proposed AI-based predictive framework proves to be highly effective in terms of disease outbreak detection, there are a few directions in which the proposed AI-based predictive framework can be enhanced to its fullest extent. The integration of real-time streams of data

is one of the main spheres of future work. By integrating real-time data of healthcare systems, eco sensors, and mobility tracking systems, it would be possible to continuously update models and perform dynamic predictions that will allow near real-time monitoring of outbreaks. This would go a long way to relevance of alerts and support proactive intervention in the form of public health.

The other significant extension is to extend the framework to allow predicting multiple diseases simultaneously. The contemporary model concentrates on the detection of an outbreak of a particular disease, but in real-world scenarios, there may be numerous infectious diseases occurring in parallel or in series. It can be extended to make predictions about different diseases, by adapting the model to handle multi-label or multi-class classification, thus making it more useful in a range of epidemiological problems.

Moreover, a combination of the suggested model with the current systems of public health surveillance is an important step on the way to practical implementation. The linkage of the framework to national and regional health monitoring systems would help in automated exchange of data, centralized analysis, and decision support. This integration can enable early warning systems, improve coordination among healthcare agencies, and enhance resource allocation strategies.

All in all, the following future improvements are expected to transform the proposed model into a predictive tool, to a scalable and real-time decision support system capable of addressing complex and evolving public health challenges.

8. CONCLUSION

This paper introduces an AI-based predictive modeling system to detect disease outbreaks by combining big data in healthcare and the environment. The proposed solution is an effective combination of heterogeneous data sources, the robust preprocessing techniques, and the advanced models of machine learning and deep learning to enhance the accuracy and reliability of predictions. The main contributions are the proposal of the single data integration model, the use of the single classification model that is designed with imbalanced outbreak data in mind, and evaluation of the model using several performance metrics.

The results demonstrate the effectiveness of the AI-integrated approach in capturing complex relationships between environmental and healthcare factors, leading to enhanced outbreak detection capability. Specifically, the model has a high performance with respect to recall, F1-score and ROC-AUC, showing that it is suitable to identify outbreak events and balance the sensitivity and accuracy of its results at the same time. This underscores the benefit of combining multi-source data and state-of-the-art AI methods compared to the old methods.

Moreover, the paper highlights the importance of the methods of classification in healthcare prediction. Measures like recall and F1-score give a more significant analysis in unbalanced datasets, where the cost of misclassification, in particular, false negatives, can have severe implications on the public health. Having these metrics as a priority, the proposed framework guarantees effective and timely recognition of disease outbreaks.

To sum it up, the suggested model is a scalable and efficient solution to the problems of disease surveillance and early warning systems. It offers informative contributions to the decision-making of the population in health matters as well as illustrates the promise of AI-based solutions in solving complex epidemiological problems.

REFERENCES

1. Adegoke, B. O., Odugbose, T., & Adeyemi, C. (2024). Data analytics for predicting disease outbreaks: A review of models and tools. International journal of life science research updates [online], 2(2), 1-9.
2. Chen, N., Zhang, C., An, W., Wang, L., Li, M., & Ling, Q. (2025). Event-based motion deblurring with blur-aware reconstruction filter. IEEE Transactions on Circuits and Systems for Video Technology, 35(9), 8508-8519.
3. De Gaetano, S., Ponzo, E., Midiri, A., Mancuso, G., Filippone, D., Infortuna, G., ... & Biondo, C. (2025). Global trends and action items for the prevention and control of emerging and re-emerging infectious diseases. Hygiene, 5(2), 18.
4. Dritsas, E., &Trigka, M. (2025). Exploring the intersection of machine learning and big data: A survey. Machine Learning and Knowledge Extraction, 7(1), 13.
5. Lee, H. W., & Li, V. C. S. (2025). Phase Boundaries and Critical Transitions in Coupled Epidemic–Behavioral Systems. Bulletin of Mathematical Biology, 87(12), 179.
6. Lessani, M. N., Li, Z., Jing, F., Qiao, S., Zhang, J., Olatosi, B., & Li, X. (2024). Human mobility and the infectious disease transmission: a systematic review. Geo-Spatial Information Science, 27(6), 1824-1851.
7. Liu, Q., Liu, M., Liang, W., Li, X., Jing, W., Chen, Z., & Liu, J. (2025). Global distribution and health impact of infectious disease outbreaks, 1996–2023: a worldwide retrospective analysis of World Health Organization emergency event reports. Journal of global health, 15, 04151.
8. Yang, P., Asano, Y. M., Mettes, P., & Snoek, C. G. (2022, October). Less than few: Self-shot video

instance segmentation. In European Conference on Computer Vision (pp. 449-466). Cham: Springer Nature Switzerland.

9. Yang, P., Mettes, P., & Snoek, C. G. (2021). Few-shot transformation of common actions into time and space. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 16031-16040).
10. Yang, P., Snoek, C. G., & Asano, Y. M. (2023). Self-ordering point clouds. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 15813-15822).
11. Ye, Y., Pandey, A., Bawden, C., Sumsuzzman, D. M., Rajput, R., Shoukat, A., ... & Galvani, A. P. (2025). Integrating artificial intelligence with mechanistic epidemiological modeling: a scoping review of opportunities and challenges. Nature Communications, 16(1), 581.
12. Zhao, A. P., Li, S., Cao, Z., Hu, P. J. H., Wang, J., Xiang, Y., ... & Lu, X. (2024). AI for science: Predicting infectious diseases. Journal of safety science and resilience, 5(2), 130-146.