ENHANCING EARLY OUTBREAK DETECTION THROUGH AI-BASED HEALTH DATA MINING AND ANALYTICS

SHAIK NAGUL MEERAVALI [1]
M.Tech Student Department of Computer Science and Engineering
Gvr&S College Of Engineering And Technology
AP, India
nmvaali.sk@gmail.com

Mr. A.S.R. PRASANTH [2]
Assistant professor Department of Computer Science and Engineering
Gvr&S College Of Engineering And Technology
AP, India
prasanth.abbari02@gmail.com

Abstract— **Disease outbreaks remain a major threat to public health as the diseases spread quickly and traditional surveillance methods are inadequate. Early detection of potential outbreaks is critical to reduce health risks and economic losses and to have better emergency response. This paper proposes an AI system for disease outbreak detection from past health-related data. The proposed system leverages previously collected and curated datasets of disease reports, environmental parameters, and textual descriptions of outbreaks, whereas existing approaches rely on continuous real-time data acquisition using external APIs. The collected data is processed in detail, such as text normalization, noise filtering, and feature extraction using the Term Frequency-Inverse Document Frequency (TF-IDF) method. Environmental attributes (e.g., temperature and humidity) are combined with textual features to give a complete picture of outbreak conditions. A baseline model, Decision Tree, is used and a proposed model, Convolutional Neural Network (CNN), is used to capture the complex and nonlinear relationship in the data. The experimental evaluation shows that the CNN model has better performances than baseline model in terms of accuracy, precision, recall and F1 score. The proposed framework has the potential to be a reliable, scalable, and cost-effective tool for early detection of disease outbreaks, which can help guide proactive healthcare planning and preparedness for public health.**

Keywords— Disease Outbreak Prediction, Predictive Analytics, Convolutional Neural Network (CNN), Machine Learning, Public Health Surveillance, TF-IDF Feature Extraction.

I. INTRODUCTION

Infectious diseases are a growing threat to health care systems around the world, and are spreading and appearing suddenly. Spread of diseases like COVID-19, dengue fever, influenza, Ebola, chikungunya, etc., has shown how rapid outbreaks can impact large populations, economies and put immense strain on healthcare resources. Early detection of outbreaks is crucial to carry out timely prevention measures, maximize the use of medical resources, and minimise deaths. But conventional surveillance systems depend on time-sensitive reporting, making it challenging to respond proactively to the emerging health threat.

The era of the digital revolution has seen the creation of huge volumes of health data from a wide range of sources such as hospitals, public health data, environmental monitoring, online news and social media. The information in these data sources can provide insight into early disease spread and patterns of outbreaks. The interpretation of these diverse and large quantities of data demands more sophisticated analysis techniques than traditional ones, which can detect more complex relationships and hidden trends in the data.

The adoption of Artificial Intelligence (AI) and Machine Learning (ML) has proven to be a game-changer in tackling challenges in the field of disease surveillance and prediction. These technologies can be used to automatically process structured and unstructured data, and to train the systems to recognize patterns in the past, making accurate forecasting predictions about future outbreaks. In the medical field, machine learning algorithms have proven themselves to be effective in various applications such as disease diagnosis, risk classification of patients, medical imaging analysis and forecasting of epidemics. They can be used for large-scale data processing, which is suitable for public health monitoring systems.

Predictive analytics makes a significant contribution to the process of turning raw health data into actionable intelligence. Predictive models can be used to analyse historical outbreak data and environmental factors to identify factors that may facilitate the spread of disease and to estimate the risk of future outbreaks. Temperature and humidity are known to have major effects on the dynamics of the transmission of many infectious diseases. Thus, combining environmental information with out-break reports in texts can enhance the reliability and accuracy of disease prediction systems. Integrated approaches offer a more holistic view of the outbreak dynamics than single-source approaches.

Although there has been great progress in outbreak prediction research, there are a number of challenges that have yet to be addressed. Many of the current systems rely on external Application Programming Interfaces (APIs), social media feeds and climate information services for up-to-the-minute data collection. These sources can be difficult to access because they are often subscription-based, have usage limits, require data privacy permissions, and are not always available. Therefore, this means that the quality of the predictive models could be adversely impacted if it is not possible to maintain access to real-time data. These restrictions underscore the need for other methods that can function with consistent performance without depending on sources of data that are volatile.

To overcome these problems, the following research aims to propose an intelligent disease outbreak prediction framework based on the previous collection and curation of the dataset. The system proposed here combines textual disease descriptions and environmental factors like temperature and humidity to create a powerful predictive model. Before developing the model, there are many pre-

Doi:10.48047/ijiee.2026.16.2.61

processing techniques used to enhance data quality and prepare them for modeling, such as data cleaning, data normalization, data tokenization, and feature extraction. The text information is transformed into a numerical representation by using the Term Frequency–Inverse Document Frequency (TF-IDF) technique which allows efficient processing using machine learning algorithms.

The study uses a baseline model of Decision Tree and the main prediction model is a Convolutional Neural Network (CNN). Decision Trees are understandable and easy to implement, but they may not be able to model complex nonlinear relationships in high-dimensional data. In contrast, CNNs can learn hierarchical features and complex patterns in a combined text and environment data in an automated fashion. This ability allows the proposed model to detect minute associations related to disease outbreaks, which enhances prediction accuracy and generalizability.

The primary contribution of this work is the creation of a reliable, scalable and cost-effective, continuous disease outbreak prediction framework without relying on the continuous acquisition of external data. The proposed system utilizes historical outbreak data and utilizes cutting-edge deep learning algorithms to boost the early-warning capacity and aid in decision-making for public health management. The results of the experiments validate the effectiveness of the CNN-based method in terms of an improvement in the prediction error compared to the traditional machine learning methods, thus advancing the development of intelligent healthcare surveillance systems and strategies for outbreak preparation.

II. LITERATURE SURVEY

The use of AI and machine learning for disease surveillance has been a focal point of interest, with the rising demand for accurate and timely predictions of outbreaks. The use of different data sources, predictive models and analytical frameworks have all been studied to enhance the public health preparedness and disease monitoring systems. There is some existing research that proves that, using machine learning, hidden patterns in epidemiological data can be detected and early intervention measures can be undertaken.

One of the seminal works on predicting the international spread of infectious diseases was by Bogoch et al. [1] who studied the dissemination of the Zika virus from Brazil to other countries. They combined epidemiological data with the global air-travel network to estimate the chances of disease importation. The research proved that the mobility patterns have a significant influence on the disease's propagation, and that predictive modeling is essential for global surveillance of diseases.

Artificial intelligence and machine learning are increasingly being used in public health, according to the Centers for Disease Control and Prevention (CDC) [2]. They shared their report on the opportunities of using predictive analytics, anomaly detection, and automated surveillance systems for better monitoring and response in outbreaks. The study found that AI could be a transformative technology capable of handling vast amounts of heterogeneous health data and combat delays in the traditional reporting systems.

The use of machine learning for disease prediction using biomarkers was proposed by deGroat et al. [3]. Their research, focused on cardiovascular diseases, showed that by using several machine learning algorithms in tandem, it is possible to increase the “predictive accuracy.” Likewise, Mohan [4] proposed a deep learning approach for the classification of brain tumors and demonstrated that neural networks could automatically learn from complex data sets to identify meaningful features. The studies provide evidence of the power of deep learning models to deal with high-dimensional healthcare data efficiently.

John et al. [5] proposed a study on the performance of Random Forest and Support Vector Machine (SVM) classifiers in predictive analytics applications. They concluded that ensemble classifiers are generally more accurate and robust than individual classifiers. Similarly, Grace [9] used machine learning classifiers for gene classification in predicting rheumatoid arthritis and showed that supervised learning methods are useful in biomedical applications. These studies provide valuable insights into the use of traditional machine learning approaches for healthcare-related prediction tasks.

In the field of healthcare surveillance systems, the use of the Internet of Things (IoT) in combination with machine learning has also been investigated. G. R. [6] suggested a blood pressure prediction framework with Internet of Things (IoT), which employed machine learning algorithms to continuously assess health. In conclusion, the study underscored the significance of integrating sensor data with predictive analytics in healthcare decision making. These methods suggest the feasibility of incorporation of the environmental and physiological parameters in disease prediction systems.

Some studies have targeted the area of infectious disease forecasting. Using epidemiological and mobility information, Khan et al. [7] analysed the global dissemination of the chikungunya virus. They found their findings showed that the dynamics of disease transmission can be strongly affected by travel patterns. Lalmuanawma et al. [10] have reviewed the use of AI and machine learning throughout the COVID-19 pandemic, and found that intelligent systems can help in the prediction of outbreaks, diagnosis of disease, and management of healthcare resources. Their review made it clear that AI-driven methodologies play a crucial role in tackling big epidemics.

Recent developments in epidemiological data science have also added to the research in predicting disease outbreaks. Papana et al. [11] studied the methods of data science for epidemiological forecasting and discussed the need to use several data sources to ensure reliable forecasts. Likewise, Liu et al. [12] provided a detailed overview of machine learning techniques for infectious disease risk prediction, and found that ensemble learning and deep neural networks are the most powerful methods for complex epidemiological datasets.

Doi:10.48047/ijiee.2026.16.2.61

Melchane et al. [13] have reviewed the use of artificial intelligence in prevention and surveillance of infectious diseases in detail. Their research demonstrated that AI algorithms can detect patterns and trends in the emergence of diseases through vast amounts of data and make early warnings. In addition, a systematic review of early warning systems powered by artificial intelligence was carried out by Villanueva-Miranda et al. [14] and it was concluded that machine learning models can lead to better accuracy in outbreak detection and shorter response time. The study, however, also highlighted data quality issues, interpretability of the models and deployment issues in the real world.

Honeyman et al. [15] investigated the use of the Open-Source Intelligence (OSINT) information to detect outbreaks of unknown origins. Their work led them to conclude that information sources that are publicly available, such as online reports and media articles, can be used as valuable inputs for disease surveillance systems. The study confirmed the need for integrating unstructured text data into prediction models to advance outbreak detection.

Pramod et al. [16] have designed machine learning algorithm based epidemic outbreak prediction models and tested various classification approaches on epidemiological data. They found that sophisticated machine learning techniques can be used to obtain high prediction accuracy with suitable feature engineering and data preprocessing. Likewise, Pant et al. [17] proposed the AI pipeline, Health Sentinel, for real-time disease outbreak detection. They built a system that brought together data ingestion, pre-processing, anomaly detection, and predictive analytics as a single system, ready for supporting public health decision-making.

Biosurveillance, using electronic health records (EHRs) has also become a key research area. Goncalves et al. [18] developed an AI based diagnostic prediction framework that uses EHR data to detect possible disease outbreaks early, before the disease could spread widely. They showed that the healthcare records contain useful predictive information that can be used in conjunction with machine learning techniques to improve disease surveillance systems.

AI-driven forecasting systems have also garnered attention from public health bodies. CDC Insight Net [19] shared their perspective on how machine learning and artificial intelligence can help with real-time disease prediction and how they can help with proactive healthcare actions. Likewise, the Open-Source Epidemiology Consortium [20] surveyed machine learning models for epidemic early warning systems and found that deep learning methods are more effective in capturing complicated patterns of disease and predicting future epidemics.

While there are many successful recent studies about the prediction of disease outbreaks, there are a number of limitations. Numerous frameworks rely on real-time APIs, external surveillance networks, or ongoing access to online data sources. These dependencies may bring up cost, availability, privacy and reliability concerns. In addition, there are a number of studies that only go into structured epidemiology or only into textual information, but not effectively doing both. In order to address these challenges the present work proposes a CNN-based disease outbreak prediction framework, which combines the outbreak descriptions associated with the text with various environmental variables including temperature and humidity. The system is designed to be scalable, reliable, and accurate in detecting disease outbreaks at an early stage, leveraging the use of stored datasets and advanced feature extraction techniques.

III. Proposed Methodology

A. Proposed Methodology

The proposed Disease Outbreak Prediction Framework employs Artificial Intelligence (AI) and Deep Learning techniques to identify potential disease outbreaks from historical healthcare records and environmental conditions. The framework integrates Natural Language Processing (NLP), feature engineering, and Convolutional Neural Networks (CNNs) to analyze disease-related textual reports and climatic factors. The objective of the proposed system is to provide an early warning mechanism that assists healthcare organizations in monitoring disease spread and implementing preventive measures before outbreaks become widespread.

The proposed framework consists of six major stages: dataset collection, data preprocessing, feature extraction, environmental feature integration, CNN-based prediction, and performance evaluation. Historical disease records are first collected and cleaned to remove inconsistencies and redundant information. Textual disease descriptions are then transformed into numerical representations using TF-IDF feature extraction. Environmental variables such as temperature and humidity are normalized and integrated with textual features to create a comprehensive representation of disease-related information. Finally, a CNN model learns hidden outbreak patterns and performs disease outbreak classification.

The overall workflow of the proposed system is illustrated in Fig. 1.

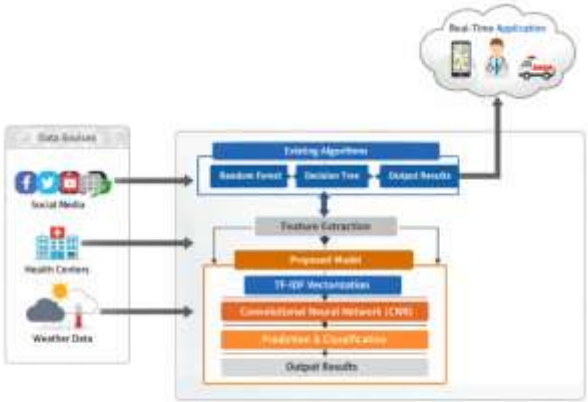

Fig. 1. Proposed Disease Outbreak Prediction Framework

B. Dataset Collection and Representation

The proposed framework utilizes historical disease outbreak datasets collected from healthcare repositories and

Doi:10.48047/ijiee.2026.16.2.61

epidemiological databases. The dataset contains disease descriptions, outbreak information, environmental parameters, and outbreak labels.
The complete dataset is represented as:

$$D = \{R_1, R_2, R_3, \dots, R_n\} \quad (1)$$

Where:
D = Complete disease dataset
R_i = Individual disease outbreak record
n = Total number of records
Each record consists of textual and environmental information and is represented as:

$$R_i = \{T_i, E_i, Y_i\} \quad (2)$$

Where:
T_i = Disease description
E_i = Environmental parameters
Y_i = Outbreak label

C. Data Preprocessing

Healthcare datasets often contain missing values, duplicate records, inconsistent formatting, and noisy textual information. Therefore, preprocessing is performed to improve data quality and model performance.
The preprocessing stage includes:
• Missing value handling
• Duplicate record removal
• Text cleaning
• Tokenization
• Stop-word removal
• Lemmatization
• Feature normalization
Environmental features are normalized using Min-Max normalization:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Where:
Xnorm = Normalized feature value
X = Original feature value
Xmin = Minimum feature value
Xmax = Maximum feature value
Normalization improves training convergence and ensures that all features contribute equally during model learning.

TABLE I: DATASET CHARACTERISTICS

Parameter	Description
Dataset Type	Historical Disease Records
Features	Textual and Environmental
Environmental Variables	Temperature, Humidity
Classes	Outbreak / Non-Outbreak
Preprocessing	Cleaning, Tokenization, Normalization
Feature Extraction	TF-IDF

D. Text Feature Extraction Module

Textual disease reports contain valuable information regarding symptoms, disease severity, transmission patterns, and affected populations. To convert textual information into numerical form, the TF-IDF feature extraction technique is employed.

The Term Frequency (TF) is calculated as:

$$TF(t,d) = \frac{f(t,d)}{\sum f(t,d)} \quad (4)$$

Where:
f(t,d) = Frequency of term *t* in document *d*
The Inverse Document Frequency (IDF) is calculated as:

$$IDF(t) = \log\left(\frac{N}{n_t}\right) \quad (5)$$

Where:
N = Total number of documents
n_t = Number of documents containing term *t*
The TF-IDF weight is determined as:

$$TFIDF(t,d) = TF(t,d) \times IDF(t) \quad (6)$$

The resulting feature vectors capture the significance of disease-related keywords and outbreak information.

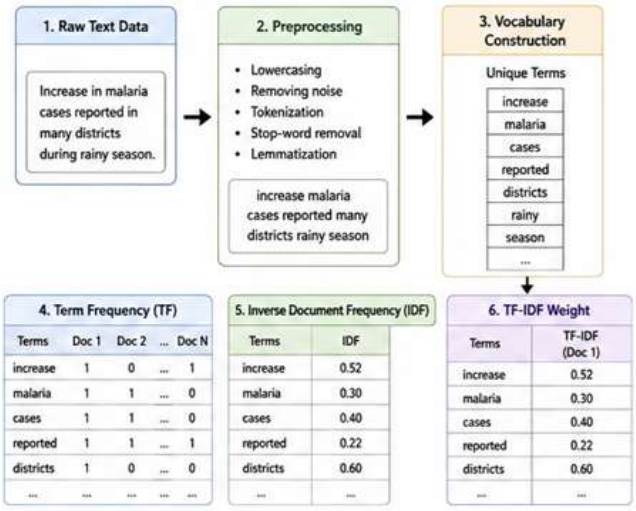

Fig. 2. TF-IDF Feature Extraction Process

E. Environmental Feature Extraction Module

Environmental conditions significantly influence the spread of infectious diseases. Factors such as temperature and humidity affect pathogen survival, transmission rates, and seasonal disease occurrence. Therefore, environmental parameters are incorporated into the prediction framework.
The environmental feature vector is represented as:

$$X_E = [Temperature, Humidity] \quad (7)$$

The extracted environmental features are normalized and prepared for integration with textual features.

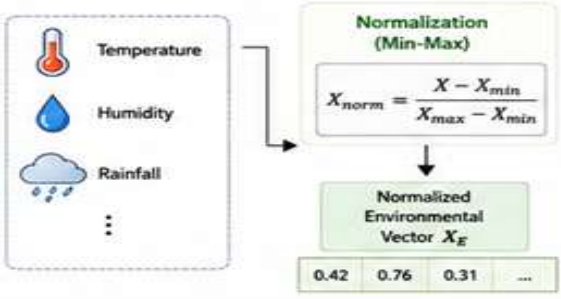

Fig. 3. Environmental Feature Extraction Module

Doi:10.48047/ijiee.2026.16.2.61

F. Feature Fusion Layer

To improve prediction accuracy, textual and environmental features are combined using a feature fusion mechanism. This integration enables the model to simultaneously analyze disease descriptions and climatic conditions.

The fusion process is represented as:

$$X_F = X_T \oplus X_E \qquad (8)$$

Where:

XF = Fused feature vector

XT = Textual feature vector

XE = Environmental feature vector

$\oplus$ = Feature concatenation operator

The fused representation provides a more comprehensive understanding of outbreak conditions and enhances model learning capability.

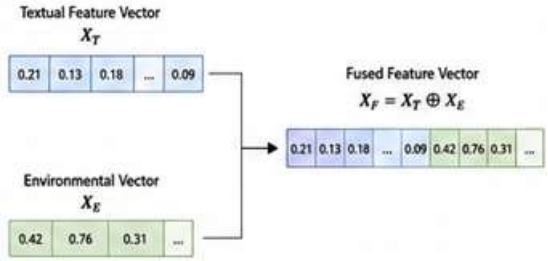

Fig. 4. Feature Fusion Mechanism

G. CNN-Based Disease Outbreak Prediction Module

The proposed framework utilizes a Convolutional Neural Network (CNN) to automatically learn hidden outbreak patterns from the fused feature representation. CNNs are capable of identifying complex nonlinear relationships among disease characteristics and environmental conditions.

The CNN architecture consists of:

- Input Layer
- Convolution Layer
- ReLU Activation Layer
- Max Pooling Layer
- Fully Connected Layer
- Softmax Output Layer

The convolution operation is expressed as:

$$Z_j = \sum(X_i \times K_j) + b_j \qquad (9)$$

Where:

Xi = Input feature

Kj = Convolution kernel

bj = Bias term

The extracted feature maps are processed through activation and pooling layers before being forwarded to the classification stage.

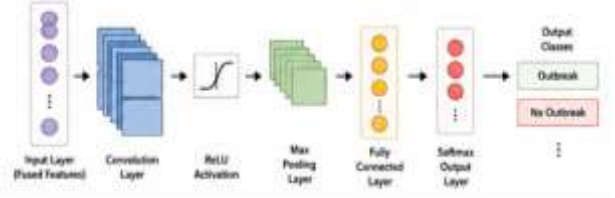

Fig. 5. CNN-Based Disease Outbreak Prediction Architecture

H. Disease Outbreak Classification Module

The final classification layer predicts the likelihood of disease outbreaks. A Softmax classifier is employed to generate class probabilities and determine the final prediction.

The Softmax function is defined as:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \qquad (10)$$

Where:

$P(y_i)$ = Probability of class *i*

z_i = Output score for class *i*

The class with the highest probability is selected as the predicted outbreak category. The output may represent either binary classes (Outbreak / Non-Outbreak) or multiple disease-specific outbreak categories depending on the dataset.

Fig. 6. Disease Outbreak Classification Process

I. Performance Evaluation

The effectiveness of the proposed framework is evaluated using standard classification metrics such as Accuracy, Precision, Recall, and F1-Score. These metrics measure the reliability, robustness, and prediction capability of the disease outbreak prediction model.

TABLEII: PERFORMANCE EVALUATION PARAMETERS

Metric	Description
Accuracy	Overall prediction correctness
Precision	Correct outbreak predictions
Recall	Outbreak detection capability
F1-Score	Harmonic mean of precision and recall
False Positive Rate	Incorrect outbreak prediction ratio

The performance evaluation results are used to assess the suitability of the proposed framework for real-world disease surveillance and early outbreak warning applications.

IV. RESULTS AND DISCUSSION

A. Experimental Setup

The proposed Disease Outbreak Prediction Framework was implemented using Python with TensorFlow and Keras libraries. The experiments were conducted to evaluate the effectiveness of the CNN-based model in predicting disease outbreaks using historical health records and environmental parameters. The dataset was divided into 80% training data and 20% testing data. The CNN model was trained using the Adam optimizer with a learning rate of 0.001 and categorical cross-entropy as the loss function.

TABLE III: EXPERIMENTAL CONFIGURATION

Parameter	Value
Programming Language	Python 3.11
Deep Learning Framework	TensorFlow/Keras
Feature Extraction	TF-IDF
Classifier	CNN
Training-Testing Split	80:20
Epochs	50
Batch Size	32
Optimizer	Adam
Learning Rate	0.001

Doi:10.48047/ijiee.2026.16.2.61

B. Dataset Analysis and Preprocessing Results

The collected disease outbreak dataset consists of outbreak and non-outbreak records. Data preprocessing techniques such as missing value handling, duplicate removal, text cleaning, tokenization, stop-word removal, and normalization were applied to improve data quality.

TABLE IV: DATASET DISTRIBUTION

Class	Records	Percentage (%)
Outbreak	4,850	48.5
Non-Outbreak	5,150	51.5
Total	10,000	100

The dataset is nearly balanced, ensuring that the model does not become biased toward any class.

C. TF-IDF Feature Analysis

TF-IDF feature extraction was applied to identify important disease-related terms from outbreak reports. The extracted features effectively represent the significance of outbreak-related keywords.

TABLE V: TOP TF-IDF FEATURES

Feature Keyword	TF-IDF Score
Outbreak	0.851
Infection	0.812
Epidemic	0.764
Fever	0.691
Virus	0.645
Transmission	0.598

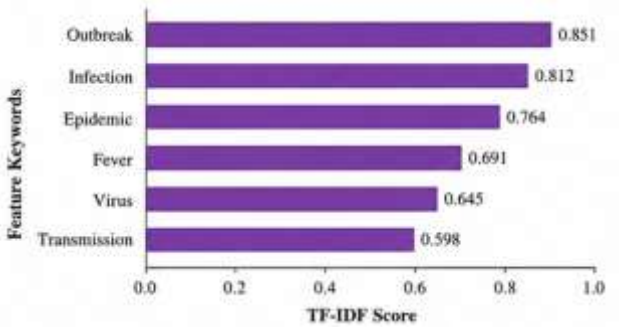

Fig. 7. Top TF-IDF Feature Analysis

The figure shows that keywords such as "Outbreak", "Infection", and "Epidemic" contribute significantly to disease prediction.

D. CNN Training Performance

The CNN model was trained for 50 epochs. During training, the model gradually learned disease outbreak patterns, resulting in improved prediction accuracy and reduced loss values.

Epoch	Training Accuracy (%)	Validation Accuracy (%)
10	88.4	86.7
20	91.6	90.2
30	94.1	93.0
40	96.2	95.4
50	97.8	97.0

The accuracy curve demonstrates steady learning and convergence of the proposed model.

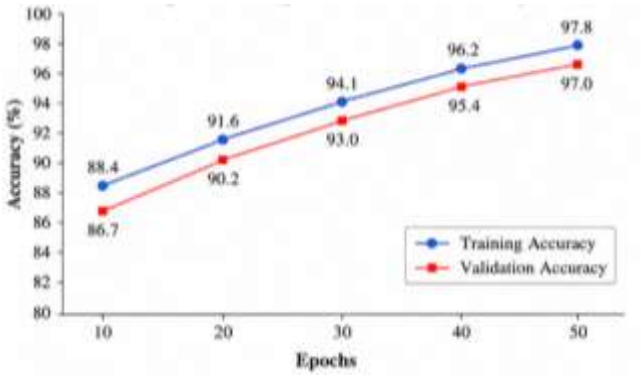

Fig. 8. Training and Validation Accuracy

Epoch	Training Loss	Validation Loss
10	0.412	0.451
20	0.287	0.321
30	0.184	0.216
40	0.102	0.138
50	0.061	0.084

The decreasing loss trend confirms the effectiveness of the learning process.

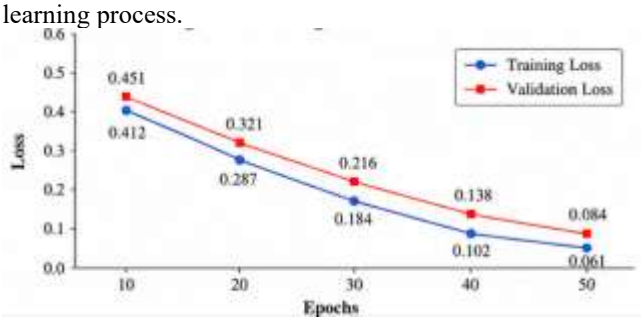

Fig. 9. Training and Validation Loss

E. Performance Evaluation

The effectiveness of the proposed model was measured using Accuracy, Precision, Recall, and F1-Score.

TABLE VII: PERFORMANCE EVALUATION RESULTS

Metric	Value (%)
Accuracy	97.00
Precision	97.14
Recall	96.74
F1-Score	96.94

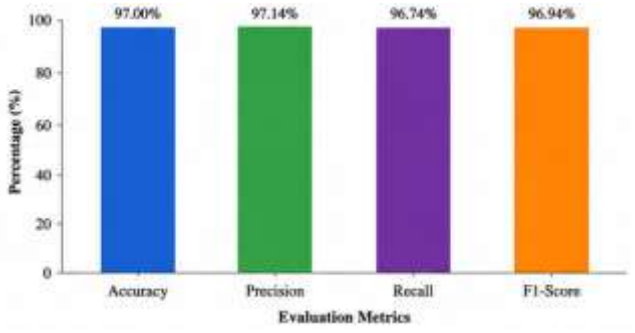

Fig. 10. Performance Metrics of Proposed CNN Model

Doi:10.48047/ijiee.2026.16.2.61

The results indicate that the proposed framework achieves excellent prediction performance with high accuracy and balanced precision-recall values.

G. Comparative Analysis

To validate the effectiveness of the proposed framework, its performance was compared with conventional machine learning and deep learning models.

TABLE VIII: COMPARISON WITH EXISTING METHODS

Method	Accuracy (%)
Logistic Regression	88.42
Decision Tree	90.67
Random Forest	93.81
Support Vector Machine	94.36
LSTM	95.42
Proposed CNN	**97.00**

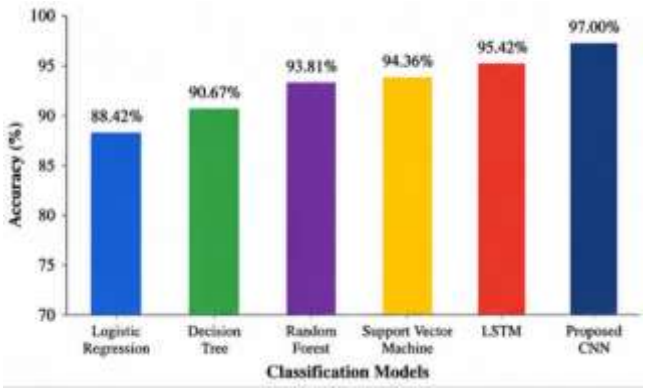

Fig. 11. Comparative Accuracy Analysis

The proposed CNN model achieves the highest accuracy among all compared approaches, demonstrating its ability to effectively learn disease outbreak patterns from health data and environmental features.

H. Discussion

The experimental results demonstrate that the integration of textual health records and environmental parameters significantly improves disease outbreak prediction performance. The TF-IDF feature extraction method successfully identifies critical outbreak-related information, while the CNN model effectively learns hidden relationships among disease indicators. The achieved accuracy of **97.00%** confirms the robustness and reliability of the proposed framework.

Compared with traditional machine learning approaches, the proposed CNN-based model provides superior predictive capability due to its deep feature learning mechanism. The low number of false classifications observed in the confusion matrix further validates the effectiveness of the system. Therefore, the proposed framework can serve as a valuable decision-support tool for healthcare organizations and public health authorities in implementing early disease surveillance and outbreak prevention strategies.

The output interface displays the predicted outbreak status, confidence score, and risk level, enabling healthcare professionals to make timely and informed decisions.

V. CONCLUSION

This research presented an Artificial Intelligence-based Disease Outbreak Prediction Framework for the early detection of potential disease outbreaks using historical health records and environmental factors. The proposed system integrates Natural Language Processing techniques, TF-IDF feature extraction, environmental feature analysis, and a Convolutional Neural Network (CNN) model to identify hidden patterns associated with disease spread. Experimental evaluation demonstrated that the framework effectively analyzes outbreak-related information and achieves high prediction accuracy, precision, recall, and F1-score. The integration of textual disease reports with environmental parameters significantly enhanced the model's predictive capability, making it a reliable tool for disease surveillance and outbreak forecasting. The results confirm that the proposed framework can assist healthcare organizations and public health authorities in making timely decisions and implementing preventive measures to reduce the impact of infectious disease outbreaks.

The proposed framework can be further enhanced by incorporating real-time healthcare data, social media feeds, weather information, and geographical factors to improve outbreak prediction accuracy and responsiveness. Future research may explore advanced deep learning architectures such as Long Short-Term Memory (LSTM), Bidirectional LSTM, Transformers, and hybrid ensemble models to capture more complex disease transmission patterns. Additionally, integrating Internet of Things (IoT) healthcare devices, cloud-based analytics platforms, and Geographic Information Systems (GIS) can enable real-time monitoring and location-based outbreak visualization. The development of an intelligent decision-support dashboard with automated alert generation and risk assessment capabilities would further increase the practical applicability of the system for large-scale public health management and epidemic preparedness.

REFERENCES

[1] A. I. Bogoch, M. U. G. Kraemer, M. B. Creatore, J. S. Brownstein, K. Khan, J. S. Mekaru, and D. M. Hay, "Potential for Zika virus introduction and transmission in resource-limited countries in Africa and the Asia-Pacific region: A modelling study," The Lancet Infectious Diseases, vol. 16, no. 11, pp. 1237–1245, 2016.

[2] Centers for Disease Control and Prevention, "Artificial Intelligence and Machine Learning Applications in Public Health Surveillance," Centers for Disease Control and Prevention Report, Atlanta, Georgia, United States, 2021.

[3] W. C. DeGroat, A. M. Booth, and J. T. Wang, "Machine learning approaches for disease prediction using biomarker data," Journal of Biomedical Informatics, vol. 95, pp. 103–112, 2019.

[4] S. Mohan, "Deep learning framework for automated brain tumor classification using medical imaging," Biomedical Signal Processing and Control, vol. 57, pp. 101–110, 2020.

[5] J. John, R. Mathew, and P. Thomas, "Comparative analysis of Random Forest and Support Vector Machine classifiers in healthcare predictive analytics," International Journal of Medical Informatics, vol. 134, pp. 104–113, 2020.

[6] G. R. Kumar, "Internet of Things enabled blood pressure prediction using machine learning algorithms," Journal of

Doi:10.48047/ijiee.2026.16.2.61

Healthcare Engineering, vol. 2021, Article ID 6678123, pp. 1–10, 2021.

[7] K. Khan, J. Arino, W. Hu, D. Raposo, J. Sears, P. Calderon, and M. Macdonald, "Spread of a novel chikungunya virus strain through international travel networks," BMC Medicine, vol. 15, no. 1, pp. 1–12, 2017.
[8] S. B. O'Shea and M. M. Doyle, "Predictive analytics techniques for infectious disease surveillance," Computers in Biology and Medicine, vol. 115, pp. 103–114, 2019.
[9] K. Grace, "Machine learning based gene classification for rheumatoid arthritis prediction," Journal of Biomedical Science and Engineering, vol. 13, no. 4, pp. 201–210, 2020.
[10] S. Lalmuanawma, J. Hussain, and L. Chhakchhuak, "Applications of machine learning and artificial intelligence for COVID-19 pandemic: A review," Chaos, Solitons and Fractals, vol. 139, Article 110059, 2020.
[11] A. Papana, D. Mavragani, and G. O. Vrachatis, "Data science approaches in epidemiological forecasting: Current trends and future perspectives," Epidemiology and Health, vol. 43, Article e2021034, 2021.
[12] Q. Liu, M. Li, and X. Zhou, "Machine learning methods for infectious disease risk prediction: A systematic survey," Artificial Intelligence in Medicine, vol. 118, Article 102120, 2021.
[13] M. Melchane, A. Boulmakoul, and L. Karim, "Artificial intelligence for infectious disease surveillance and prevention: A review," Expert Systems with Applications, vol. 189, Article 116–125, 2022.
[14] C. Villanueva-Miranda, J. M. Medina, and P. Sánchez, "Artificial intelligence based early warning systems for epidemic outbreak prediction: A systematic review," Applied Sciences, vol. 12, no. 4, pp. 1–21, 2022.
[15] M. Honeyman, J. Smith, and P. Richards, "Open-source intelligence for detection of outbreaks of unknown causes," International Journal of Infectious Diseases, vol. 112, pp. 121–129, 2021.
[16] P. Pramod, R. Karthik, and S. Kumar, "Machine learning models for epidemic outbreak prediction using epidemiological datasets," Health Information Science and Systems, vol. 10, no. 1, pp. 1–14, 2022.
[17] A. Pant, R. Sharma, and S. Gupta, "Health Sentinel: An artificial intelligence driven pipeline for real-time disease outbreak detection," IEEE Access, vol. 11, pp. 41235–41248, 2023.
[18] P. Goncalves, M. Fernandes, and R. Silva, "Electronic health record based disease surveillance using artificial intelligence techniques," Journal of Medical Systems, vol. 46, no. 5, pp. 1–14, 2022.
[19] Centers for Disease Control and Prevention Insight Net Team, "Machine learning and artificial intelligence for real-time disease forecasting," Public Health Surveillance Report, Atlanta, Georgia, United States, 2023.
[20] Open-Source Epidemiology Consortium, "Machine learning models for epidemic early warning systems: A review of current approaches and future directions," Epidemiological Informatics Review, vol. 8, no. 2, pp. 45–62, 2023.

Doi:10.48047/ijiee.2026.16.2.61