Review

Not peer-reviewed version

A Systematic Review of Machine Learning Applications in Infectious Disease Prediction, Diagnosis, and Outbreak Forecasting

Yiting Wang [*] , Jiachen Zhong , Rohan Kumar

Posted Date: 15 April 2025

doi: 10.20944/preprints202504.1250.v1

Keywords: infectious disease; machine learning; outbreak prediction; disease diagnosis; systematic review

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: Posted: 15 April 2025 doi:10.20944/preprints202504.1250.v1

Review

A Systematic Review of Machine Learning Applications in Infectious Disease Prediction, Diagnosis, and Outbreak Forecasting

Yiting Wang [1], Jiachen Zhong [2] and Rohan Kumar [3,*]

[1] Department of Data Science, University of Southern California, Los Angeles, 90089, CA, USA
[2] Department of Applied Mathematics, University of Washington, Seattle, 98105, WA, USA
[3] Independent Researcher, Miami, FL, 33174, USA
* Correspondence: rs419222@gmail.com

Abstract: Infectious diseases pose a significant global health burden, contributing to millions of deaths annually despite advancements in sanitation and healthcare access. This review systematically examines the role of machine learning in infectious disease prediction, diagnosis, and outbreak forecasting in the United States. We first categorize existing studies according to the type of disease and the ML methodology, highlighting key findings and emerging trends. We then examine the integration of hybrid and deep learning models, the application of natural language processing (NLP) in public health monitoring, and the use of generative models for medical image enhancement. In addition, we discuss the applications of machine learning in five diseases, including coronavirus disease 2019 (COVID-19), influenza (flu), human immunodeficiency virus (HIV), tuberculosis, and hepatitis, focusing on its role in diagnosis, outbreak prediction, and early detection. Our findings suggest that while machine learning has significantly improved disease detection and prediction, challenges remain in model generalizability, data quality, and interpretability.

Keywords: infectious disease; machine learning; outbreak prediction; disease diagnosis; systematic review

---

1. Introduction

Infectious diseases are considered a concerning global health burden for the millions of deaths out of it [1]. Despite improved sanitation and access to health care, recent global changes have greatly increased the risk and consequence of the outbreak of infectious diseases [2]. Infectious disease also remains a continued challenge on public health resources and individual health in the United States, where millions of people are affected by endemic diseases such as chronic hepatitis, HIV, and other sexually transmitted infections [3]. The increasing burden of infectious diseases highlights the importance of having predictive solutions, whether for estimating outbreak trends or diagnosing diseases [4]. Such solutions enable public health institutions to respond quickly and accurately at an early stage, reducing the risk and consequences of disease progression. This makes machine learning techniques a highly favorable tool in this field, with many studies and experiments demonstrating strong predictive performance [5].

Statistical methods [6], machine learning [7–11], and deep learning are emerging technologies that have been widely introduced in the medical field. Traditional methods in the prediction and detection of infectious diseases include random forest boost [12], XGBoost [13], enhanced k-nearest neighbors (KNN) classifier enhanced [14], naive bayes [15], neural networks [16], gradient boost-based method [17], support vector machine (SVM) [18] and linear discriminant analysis (LDA) [19]. Beyond these traditional approaches, researchers have explored advanced machine learning techniques [20–22] and hybrid methods to achieve better performance. Wan et al. developed a model-based reinforcement learning framework for managing infectious diseases [23]. Mishra et al. concluded that ensemble methods hold significant potential for improving diagnostic accuracy in the early stages of disease detection [24].

Deep learning [25–31] has demonstrated exceptional capability in analyzing medical images with high accuracy [32]. Traditional deep learning methods for image processing in infectious disease detection include convolutional neural networks (CNNs) [33–35], long short-term memory (LSTM) [36] and ResNet [37,38]. To further enhance performance, Farooq et al. fine-tuned a pre-trained ResNet-50 model on chest X-ray images for early disease screening [39]. Islam et al. integrated a CNN-LSTM model, combining CNN with LSTM to enable automated disease diagnosis from X-ray images [40]. Similiarly, Mehta et al. utilized a conditional generative adversarial network (cGAN) combined with a fine-tuned deep transfer learning model to classify chest X-ray images, improving diagnostic accuracy [41]. Kalane et al. implemented a universal network (U-Net) architecture to develop an automated detection system for disease identification using insights from Computer Tomography (CT) images [42].

Generative models [43,44,44,45] have also been applied to infectious disease diagnosis and prediction with notable examples including generative adversarial networks (GANs) [46,47], variational autoencoders (VAEs) [48] and diffusion model [49]. ranolo proposed a deep convolutional neural network integrated with an autoencoder (AE), referred to as CAE-COVIDX, for disease detection [50]. Beyond generative models, language models [51–55] have also been applied to disease diagnosis and outbreak prediction, including recurrent neural networks (RNNs) [56], long short-term memory (LSTM) [57,58], and natural language processing (NLP) [59,60] to enhance the accuracy and efficiency. Besides these models, more advanced approaches also exist [61,62]. Agarwal et al. utilized a hierarchical multi-modal approach based on bidirectional encoder representations from transformers (BERT) to precisely predict patient outcomes [63]. Cheng et al. found that the chat generative pre-trained transformer (ChatGPT) is valuable in infectious diseases, with significant potential for further applications [64].

Recognizing the significant potential of machine learning in infectious disease prediction, this review systematically explores the development and recent advancements in machine learning applications, with a particular focus on major infectious diseases prevalent in the United States. The Review Design section outlines the search strategy and selection criteria employed to identify relevant studies. The Results section presents a detailed overview of machine learning applications across specific infectious diseases. Finally, the strengths and limitations of this review are discussed, followed by a concluding section that summarizes key findings and implications for future research.

Table 1. Summary of Infectious Disease Research Papers on ML/DL.

Paper	Year	ML	DL	Generative	LLM	Algorithm
[19]	2009	✓	✗	✗	✗	LDA
[18]	2012	✓	✗	✗	✗	SVM
[59]	2019	✓	✗	✗	✗	NLP
[33]	2016	✗	✓	✗	✗	CNN
[35]	2020	✗	✓	✗	✗	CNN
[39]	2020	✗	✓	✗	✗	ResNet-50
[40]	2020	✗	✓	✗	✗	CNN-LSTM
[46]	2020	✗	✓	✓	✗	GAN
[60]	2020	✓	✗	✗	✗	NLP
[34]	2021	✗	✓	✗	✗	CNN
[37]	2021	✗	✓	✗	✗	ResNet
[38]	2021	✗	✓	✗	✗	ResNet

Table 1. *Cont.*

Paper	Year	ML	DL	Generative	LLM	Algorithm
[42]	2021	✗	✓	✗	✗	U-Net
[47]	2021	✗	✓	✓	✗	GAN
[48]	2021	✗	✓	✓	✗	VAE
[50]	2021	✗	✓	✓	✗	CNN-Autoencoder
[56]	2021	✓	✗	✗	✗	RNN
[23]	2021	✓	✗	✗	✗	Reinforcement
[41]	2021	✗	✓	✓	✗	cGAN + Deep Transfer
[57]	2022	✓	✗	✗	✗	LSTM
[58]	2022	✓	✗	✗	✗	LSTM
[20]	2022	✓	✗	✗	✗	Advanced ML
[63]	2022	✓	✗	✗	✓	BERT-based
[62]	2023	✗	✓	✗	✓	BERT-based
[64]	2023	✗	✓	✗	✓	ChatGPT
[61]	2024	✗	✓	✗	✓	BERT
[49]	2024	✗	✓	✓	✗	Diffusion
[21]	2025	✓	✗	✗	✗	Hybrid ML

2. Review Design

The primary objective of this review is to summarize and present the current development of machine learning applications in infectious diseases within the United States. All literature included was rigorously selected based on a predefined and systematic search strategy.

2.1. Search Strategy

We searched for the keywords "infectious disease" and "machine learning" together. The infectious diseases considered were COVID-19, Influenza (Flu), HIV/AIDS, syphilis, gonorrhea, chlamydia, sexually transmitted infections (STIs), hepatitis (A, B, C), and tuberculosis. These diseases were selected based on three factors: their widespread prevalence and impact (e.g., COVID-19 and influenza), their public health significance, and their continuing burden and relevance in the United States [65].

For the machine learning search, we used terms such as statistical learning, machine learning, deep learning, decision algorithms, clinical decision-making, artificial intelligence, neural networks, support vector machines, supervised learning, unsupervised learning, ensemble learning, Bayesian networks, random forests, natural language processing, clinical prediction, prognostic models, diagnostic analysis, and time-series analysis.

2.2. Search Selection

The topics of the selected articles include diagnosis, outbreak prediction, mortality and transmission forecasting and treatment. Several selection criteria were applied: (1) machine learning must be the primary technique used in the study; (2) only articles written in English were included; (3) studies involving humans were selected, while animal studies were excluded; (4) studies focusing on infectious diseases' genes, genomes, sequencing, or other non-machine learning analysis methods were excluded; (5) only studies with full-text availability were considered; and (6) studies related to public opinion or sentiment analysis were not included.

doi:10.20944/preprints202504.1250.v1

3. Results

In recent years, machine learning and deep learning techniques have gained significant traction in the medical domain, offering powerful tools for diagnosis, prognosis, and treatment optimization across various health conditions [66–76]. Among the numerous applications, their role in understanding and managing infectious diseases has become particularly prominent. In this section, we highlight five carefully selected infectious diseases that illustrate the breadth and impact of these technologies in modern healthcare.

3.0.1. Coronavirus Disease 2019 (COVID-19)

Coronavirus Disease 2019 (COVID-19) is a respiratory illness caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). [77]. It spreads primarily through direct human contact or respiratory droplets released when an infected person coughs or sneezes [78]. Approximately 14.9 million people lost their lives, directly or indirectly, due to the COVID-19 pandemic in 2020 and 2021 [79]. There have been approximately 103.4 million confirmed cases of COVID-19 and 1.2 million related deaths in the United States [80].

Forecasting the COVID-19 outbreak has become an increasingly critical challenge. Susceptible-Exposed-Infectious-Recovered (SEIR), Susceptible-Infectious-Recovered (SIR) models, agent-based models, curve-fitting models and traditional machine learning methods were widely applied [81]. Moulaei et al. predicted mortality in hospitalized patients with COVID-19 using machine learning methods, including Java 8 (J48) Decision Tree, EXtreme Gradient Boosting (XGBoost), Reinforcement Learning (RL), k-Nearest Neighbors (kNN), Random forest and Naive Bayes [82]. Arpaci et al. investigated various models, including BayesNet, Logistic, Instance-Based k-Nearest Neighbors (IBk), Classification Rule (CR), Partial Decision Tree (PART) and J48, for diagnosing COVID-19. Among these, CR demonstrated the best performance, achieving an accuracy of 84.21% [83].

Multilayer perceptron (MLP) and adaptive neuro-fuzzy inference system (ANFIS) playing a vital role. Ardabili et al. found MLP and ANFIS demonstrated strong generalization capabilities for long-term forecasting in their analysis of COVID-19 outbreak prediction [84]. The multilayered perceptron-imperialist competitive algorithm (MLP-ICA) and ANFIS are also applied to predict the number of infected individuals and the mortality rate in Hungary [85].

Various deep learning and hybrid models have also been utilized in this field. Dairi et al. used various machine learning and hybrid models, including LSTM-CNN (Long Short-Term Memory - Convolutional Neural Network) and GAN-GRU (Generative Adversarial Network - Gated Recurrent Unit), to forecast COVID-19 transmission, finding that LSTM-CNN achieved the best performance with a 3.718% error rate [86].

The application of machine learning (ML) and deep learning (DL) in COVID-19 forecasting has significantly evolved, demonstrating the potential of AI-driven models in epidemic prediction. Table 2 summarizes selected research papers in this domain. Early studies focused on statistical models such as SEIR and SIR, often integrated with ML techniques for enhanced accuracy. As the pandemic progressed, more advanced approaches, including hybrid deep learning models (LSTM-CNN, GAN-GRU), were introduced to improve predictive performance and adaptability to dynamic outbreak conditions.

Table 2. Summary of Selected Papers on COVID-19 Forecasting.

Paper	Year	ML	DL	Algorithm	Complexity	Scalability
[81]	2020	✓	✗	SEIR, SIR, ABM, CF	Low–Med	High
[84]	2020	✓	✗	MLP, ANFIS	Medium	Medium
[85]	2020	✓	✗	MLP-ICA, ANFIS	Med–High	Medium
[83]	2021	✓	✗	BayesNet, IBk, J48	Low–Med	Medium
[86]	2021	✓	✓	LSTM-CNN, GAN-GRU	High	High
[82]	2022	✓	✗	J48, XGBoost, kNN, RF	Medium	Med–High

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: Posted: 15 April 2025 doi:10.20944/preprints202504.1250.v1

Influenza (flu)

Influenza (flu) is a contagious respiratory illness resulting from influenza virus infection [87]. There are four types of influenza viruses: A, B, C, and D. Influenza A is the only type that can lead to severe outbreaks [87]. From 2010 to 2024, the annual impact of the flu in the U.S. ranged from 9.3 million to 41 million illnesses, 120,000 to 710,000 hospital admissions, and 6,300 to 52,000 deaths [88].

Thus, forecasting flu outbreaks is crucial, and machine learning has been widely applied in recent studies. Alessa et al. compared FastText (FT) with six traditional machine learning algorithms, achieving an F-measure of 89.9%. They also combined FT with linear regression, resulting in a 96. 29% correlation to predict outbreaks [89]. Khan et al. used a feedforward propagation neural network (MSDII-FFNN) to predict outbreaks with 90% precision [90]. Zhang et al. first applied LSTM to predict influenza outbreaks, finding that a 5-layer LSTM model with regularization achieved the lowest root mean squared error (RMSE) of 0.002 [91].

Twitter data are widely utilized for predicting outbreaks using machine learning methods. Allen et al. applied geographic information science (GIS) and support vector machine (SVM) with Twitter data to track influenza outbreaks, finding a statistically significant correlation with national, regional, and local flu reports [92]. Amin et al. also use Twitter data to identify seasonal outbreaks early and found that Random Forest outperformed other methods [93].

Hemagglutinin Type 1 and Neuraminidase Type 1 (H1N1) is a subtype of the influenza virus that caused a global pandemic in 2009 [94]. Machine learning also plays a role in supporting H1N1 vaccine prevention. Inampudi et al. determined the probability that individuals would receive the H1N1 and seasonal flu vaccines, finding that the SVM model achieved 83. 97% precision for the prediction of H1N1, while the Artificial Neural Network (ANN) model reached 86. 10% precision for the prediction of seasonal flu [95]. Ayachit et al. used ensemble learning to predict the likelihood of vaccination for H1N1 and seasonal flu, with CatBoost achieving the best performance and an accuracy of 0.8617 [96].

Influenza forecasting is crucial for public health, with research evolving from traditional statistical models and basic machine learning (ML) techniques like SVM, LR, and Decision Trees to more advanced methods. Table 3 highlights studies applying ML and deep learning (DL) techniques, including LSTM networks, to improve flu prediction accuracy. Hybrid models combining ML and DL have also been explored. These methods, leveraging real-time data from sources like social media and health records, enhance forecasting precision. Models like Random Forest and CatBoost aid in feature selection and early outbreak detection, while LSTM and ANNs capture long-term flu transmission patterns.

Table 3. Summary of Selected Papers on Influenza Forecasting

Paper	Year	ML	DL	Algorithm	Complexity	Scalability
[92]	2016	✓	✗	SVM, GIS	Medium	Medium
[91]	2017	✗	✓	LSTM	High	Medium
[89]	2019	✓	✗	FastText, LR, SVM	Medium	Medium
[96]	2020	✓	✗	CatBoost, Ensemble	Medium	High
[90]	2020	✗	✓	FFNN	Medium	Medium
[93]	2021	✓	✗	RF, SVM, NB	Medium	High
[95]	2021	✓	✓	SVM, ANN	Medium	Medium

Human Immunodeficiency Virus (HIV) / Immunodeficiency Syndrome (AIDS)

Human immunodeficiency virus (HIV) weakens the body's immune system by targeting essential defense cells. When the infection progresses to its most severe stage, it leads to acquired immunodeficiency syndrome (AIDS) [97]. HIV continues to be a critical global health concern, resulting in an estimated 42.3 million deaths worldwide [98]. More than 38,000 individuals were diagnosed with HIV in the United States [99].

Machine learning is crucial in HIV prevention and behavior prediction. Wang et al. used SVM and Random Forest to predict high-risk HIV behaviors, with cost-sensitive SVM achieving an AUC of 0.86 for multiple sexual partners and Random Forest excelling in predicting sexual activity [100]. Pan et al. used Random Forest to identify key predictors of HIV testing uptake, highlighting condomless sex, self-efficacy, condom attitudes, and depression as significant factors [101]. Nisa et al. utilized the SMOTE technique to address classification bias and applied Random Forest, achieving an 82% accuracy in predicting the probability of future HIV acquisition in high-risk groups [102].

Men who have sex with men (MSM) are disproportionately affected by HIV, accounting for 68% of new HIV diagnoses in the United States [103]. Researchers have also explored the application of machine learning methods in this field. Bao et al. used Gradient Boosting Machine (GBM) to achieve the highest Area Under the Curve (AUC) of 76.3% for predicting HIV in Australian men who have sex with men (MSM), finding that past syphilis infection was the top predictor, contributing 16.7% [104]. Chingombe et al. employed traditional machine learning algorithms, Bagging Classifier and RNNs, to predict HIV status among MSM and found that RNNs performed best, achieving an accuracy of 0.98 [105].

Deep learning has shown superior performance compared to traditional methods and machine learning techniques in both diagnosis and prediction. Turbe et al. used SVM and CNN to analyze field-based rapid diagnostic test (RDT) images for HIV, achieving 98.9% accuracy, significantly outperforming traditional visual interpretation methods [106]. Wang et al. found that LSTM outperformed Autoregressive Integrated Moving Average (ARIMA), Generalized Regression Neural Network (GRNN), and Exponential Smoothing (ES) models in forecasting HIV incidence, achieving the lowest Mean Squared Error (MSE) [107].

HIV prevention and prediction has been a critical area of research, with studies evolving from traditional statistical methods to more advanced machine learning (ML) and deep learning (DL) techniques. Table 4 highlights research applying both ML and DL methods, such as Random Forest (RF), Support Vector Machines (SVM), and Long Short-Term Memory (LSTM) networks, to enhance prediction accuracy. Hybrid models combining ML and DL approaches have been explored to improve predictive performance. These models, leveraging various data sources, such as electronic health records and social media, offer better precision in predicting HIV. Techniques like Gradient Boosting Machine (GBM) and Random Forest excel in feature selection, while deep learning models such as LSTM and Convolutional Neural Networks (CNN) capture complex patterns in HIV transmission.

Table 4. Summary of Selected Papers on HIV Prevention and Prediction

Paper	Year	ML	DL	Algorithm	Complexity	Scalability
[101]	2017	✓	✗	RF	Low–Med	Medium
[107]	2019	✗	✓	LSTM, ARIMA, GRNN	High	Medium
[106]	2021	✓	✓	SVM, CNN	High	High
[100]	2021	✓	✗	SVM, RF	Medium	Medium
[104]	2021	✓	✗	GBM	Medium	Medium
[105]	2022	✓	✓	Bagging, RNN	Medium	Medium
[102]	2023	✓	✗	RF, SMOTE	Medium	Medium

Tuberculosis

It has been known that early diagnosis of tuberculosis is favorable for disease management, and it is beneficial for both the patient and the public health to reduce the further transmission of the disease in society [108]. However, most of the conventional early detection techniques such as

X-ray, conventional light microscopy, and other similar traditional techniques turn out to be very time-consuming and complex to operate [109]. There have been many immunoassay techniques developed to rapid diagnosis of TB, but they are financially costly and require more skilled staff to be involved [109]. Therefore, it will be very helpful if there can be reliable cost-effective techniques to diagnose TB at an early stage, and the ML-based techniques turn out to be promising toward this goal [110].

The early diagonis was firstly benefieted from the introduction of Artificial neural network(ANN) in 1990, which showed good performance in recognizing the structural patterns in the X-ray images [111]. In 1998, the first automated neural methods were developed to identify TB bacilli in sputum smears stained with auramine, realizing 93.5% sensitivity in diagnosing TB without too much involvement of the stuff. In 1999, the first ANN based early diagonisis method was developed, using general regression neural network (GRNN) to diagonize patients with active pulmonary TB [112]. This model used 21 different parameters to form input patterns and achieved a sensitivity of 100% (95% CI, 91 to 100%), significantly outperforming the clinical evaluation of physicians with respect to the precision of diagnosis [112]. Some other machine learning methods such as decision trees and random forest were also found very useful in predicting the diagnosis results [113].

However, there was a challenge of obtaining reliable and accurate prediction from these models, which led the researchers more to the deep learning and AI advancement that with better ability to understand images data [109]. Benefit from the introduction of genetic algorithm and fuzzy logic and single hidden layer feed-forward neural networks, the accuracy of the diagnosis of tuberculosis was further improved since 2010 [114]. Hooda et al. proposed a CNN architecture with 7 convolutional layers and 3 fully connected layers, reaching a validation accuracy of 82.09% with Adam optimizer [115]. In 2018, Kant and Srivastava presented another new neural network based TB diagnosis method which achieved a recall of 83.78%, providing great potential of evolution into an efficient and reliable TB early detection tool with its high sensitivity [116]. Hrizi et al. a comparisons study evaluating the diagonisis ability of different machine learning models such as KNN, CART, RF, NB, LDA and SVM, concluding SVM the best performing model with mean accuracy of 84% [117]. Hansun et al. further confirmed the high potential of ML and DL methods, with ML models showing a higher average precision (93.1 71%) and sensitivity (93.2 55%) and DL models showed greater average AUC(92.12%) and specificity(91.54%) [118].

Overall, tuberculosis diagnosis has greatly benefited from the evolution of machine learning and deep learning techniques. Early approaches relied on artificial neural networks and decision trees, while more recent models utilize convolutional neural networks (CNNs), genetic algorithms, and support vector machines (SVMs) to improve diagnostic accuracy and scalability. Comparative studies confirm that ML and DL models can achieve high sensitivity and specificity, making them promising tools for early and cost-effective TB detection in both clinical and remote settings.

Table 5. Summary of Selected Papers on Tuberculosis Diagnosis

Paper	Year	ML	DL	Algorithm	Complexity	Scalability
[111]	1990	✓	✗	ANN	Low	Low
[113]	1997	✓	✗	DT, RF	Low–Med	Medium
[112]	1999	✓	✗	GRNN	Low	Low
[114]	2011	✓	✗	Feed-Forward ANN	Medium	Medium
[115]	2017	✗	✓	CNN	High	High
[116]	2018	✗	✓	Deep NN	High	High
[117]	2022	✓	✗	KNN, RF, NB, LDA, SVM	Medium	Medium

Hepatitis

Hepatitis is a transmissible viral disease that has affected 350 million people in the world with only around 10% of them diagnosed [119]. Hepatitis virus A, B, C, D, and E are the five primary varities of the disease.The disease can be caused various factors such as excessive alcohol use, reactions to drugs, and viral or bacterial infections, and the chance of early detection greatly enhance the likelihood of a good recovery [120].

Different kinds of Artificial Intelligence techniques such as predictive analytics, Natural Language Processing and Machine Learning are found effective in the Hepatitis early detection, and ML is the most emphasized one for detection for its ease of use and excellent performance [121]. Bharathi et al. compared several classification techniques such as SVM, DT, LR and RF on detecting Hepatitis and without feature selection, concluding RF is the best one with an accuracy of 89% [122]. Yaganoglu employed machine learning methods for Hepatitis virus detection, achieving an accuracy of 99.31% by adding some new features and eliminating class imbalance with SMOTE [123]. Harabor et al. evaluated four machine learning models support vector machine (SVM), random forest (RF), naive Bayes, and K nearest neighbors for hepatitis screening based on a structured survery data from Romania, concluding the KNN the best model with excellent accuracy of 98.1% [124].

Deep learning models are also found very useful in Hepatitis early detection based on the works of past researchers. Wang et al. proposed a rapid screening method for hepatitis based on Long short-term memory(LSTM) neural network, realizing excellent results with an accuracy of 97.32% and AUC of 0.995 [125]. In 2022, Chen et al. proposed a Multilayer Perceptron (MLP)-based model that outperformed traditional machine learning techniques, such as LightGBM and XGBoost, in diagnosing Hepatitis [126]. Besides, deep learning models are also found to be effective in predicting the infection scale.Guo et al. adopted and compared ARIMA, SVM, and LSTM to predict case number and incidence of Hepatitis with the real data rom 2005 to 2017 in China, finding out the LSTM was the best-performing model [127].

In the case of hepatitis, both traditional ML methods and deep learning models have demonstrated strong predictive capabilities. Techniques such as Random Forest, SVM, and LSTM have been used successfully for early detection and outbreak forecasting. Recent advancements in deep learning, particularly with LSTM and multilayer perceptrons (MLP), have shown superior performance in diagnosis. These findings support the continued integration of AI-based tools into hepatitis screening programs, especially in resource-constrained healthcare environments.

Table 6. Summary of Selected Papers on Hepatitis Early Detection

Paper	Year	ML	DL	Algorithm	Complexity	Scalability
[125]	2020	✗	✓	LSTM	High	High
[127]	2020	✓	✓	ARIMA, SVM, LSTM	Medium	High
[123]	2022	✓	✗	SVM, SMOTE	Medium	Medium
[126]	2022	✗	✓	MLP	Medium	High
[124]	2023	✓	✗	SVM, RF, Naive Bayes, KNN	Medium	High
[121]	2024	✓	✗	Predictive Analytics, NLP, ML	High	High
[122]	2024	✓	✗	SVM, DT, LR, RF	Medium	Medium

Strength and Limitations

This review focuses on the application of machine learning in predicting infectious diseases specifically within the United States, providing targeted insights relevant to domestic healthcare settings. By organizing findings across five high-impact diseases—COVID-19, influenza, HIV, tuberculosis, and hepatitis—this review presents a clear picture of the developments and advancements in machine learning techniques for infectious disease estimation from a disease-specific perspective. Supported by a well-designed search strategy and a solid systematic review process, this work serves as a practical guideline for researchers, highlighting recent innovations such as hybrid models, deep learning, generative frameworks, and language models, which help to capture emerging directions in this rapidly evolving field.

Due to resource limitations, this review only includes publications written in English, which may have led to the exclusion of valuable studies published in other languages. In addition, while this review emphasizes a disease-specific approach to presenting and discussing findings, it does not provide a statistical comparison of model performance across studies from a model-specific perspective. Addressing this aspect in future work may offer additional insights into the comparative effectiveness of different machine learning techniques in infectious disease research.

Discussion

Infectious diseases continue to pose a significant global health threat, with rising risks due to recent environmental and social changes. Traditional prediction and diagnostic methods have been increasingly supplemented by machine learning (ML) and deep learning (DL) techniques, which have shown strong potential in improving accuracy and early disease detection. This paper reviews the evolution of ML and DL approaches, including Random Forest, XGBoost, SVM, CNN, LSTM, and generative models like GANs and VAEs, emphasizing their contributions to predicting and diagnosing major infectious diseases such as HIV and chronic hepatitis. The trend in research is shifting towards hybrid models and advanced methods, which leverage complex data such as medical images and health records to provide more precise predictions.

The contribution of this review lies in its comprehensive examination of these technologies, identifying the strengths and limitations of different models, while exploring the potential of generative models and natural language processing for disease diagnosis and outbreak prediction. Future studies can build on this work by exploring the development of hybrid models that integrate multiple approaches, as well as leveraging emerging AI tools, such as ChatGPT, to further improve diagnostic speed and accuracy. The scalability of these models also suggests they can be adapted to both resource-rich and resource-limited settings, offering significant promise for improving public health responses worldwide.

Conclusions

Machine learning has become a valuable tool in the fight against infectious diseases, helping improve how we track, diagnose, and predict outbreaks. From COVID-19 and influenza to HIV, tuberculosis, and hepatitis, a wide range of ML techniques—such as ensemble models, deep learning, and natural language processing—have shown strong potential in making public health responses faster and more accurate. This review also highlights the growing interest in hybrid approaches and the increasing exploration of advanced models like transformers and generative networks.

That said, several important challenges remain. Many models still struggle with generalizability, interpretability, and practical deployment in real-world healthcare settings. Issues such as data imbalance, limited transparency in model decisions, and data privacy concerns must be addressed before broader adoption can take place. In addition, future research should expand beyond English-language literature and pay closer attention to infectious diseases that are more prevalent or impactful outside the United States. With continued progress, machine learning holds strong promise for making healthcare systems smarter, more proactive, and better equipped to handle future outbreaks.

Author Contributions: Yiting Wang and Jiachen Zhong contributed equally to the review and analysis of the literature. Rohan Kumar provided additional support but contributed less extensively. All authors reviewed and approved the final manuscript.

Funding: Not applicable.

Informed Consent Statement: All authors provide their consent for the publication of this manuscript. There are no individual participant data included, and all co-authors agree to the content and submission of this work. This article does not include any original research involving human or animal subjects. Data used were anonymized and obtained from publicly available sources or with proper permissions, ensuring participant privacy. Direct human involvement was not part of the study; thus, written consent was not applicable.

Data Availability Statement: The code used in this study is available from the corresponding author upon reasonable request.

Conflicts of Interest: Not applicable.

References

1. Organization, W.H.; et al. *Global report on infection prevention and control*; World Health Organization: Geneva, 2022.
2. Baker, R.E.; Mahmud, A.S.; Miller, I.F.; Rajeev, M.; Rasambainarivo, F.; Rice, B.L.; Takahashi, S.; Tatem, A.J.; Wagner, C.E.; Wang, L.F.; et al. Infectious disease in an era of global change. *Nature reviews microbiology* **2022**, *20*, 193–205.
3. Khabbaz, R.F.; Moseley, R.R.; Steiner, R.J.; Levitt, A.M.; Bell, B.P. Challenges of infectious diseases in the USA. *The Lancet* **2014**, *384*, 53–63.
4. Santangelo, O.E.; Gentile, V.; Pizzo, S.; Giordano, D.; Cedrone, F. Machine learning and prediction of infectious diseases: a systematic review. *Machine Learning and Knowledge Extraction* **2023**, *5*, 175–198.
5. Liu, M.; Liu, Y.; Liu, J. Machine learning for infectious disease risk prediction: a survey. *ACM Computing Surveys* **2023**.
6. Weng, Y.; Wu, J.; et al. Fortifying the global data fortress: a multidimensional examination of cyber security indexes and data protection measures across 193 nations. *International Journal of Frontiers in Engineering Technology* **2024**, *6*, 13–28.
7. Lu, B.; Dan, H.C.; Zhang, Y.; Huang, Z. Journey into Automation: Image-Derived Pavement Texture Extraction and Evaluation. *arXiv preprint arXiv:2501.02414* **2025**.
8. Zhang, T.; Liu, X.; Guo, Z.; Tian, Y. Adaptive Music Recommendation: Applying Machine Learning Algorithms Using Low Computing Device. *Journal of Software Engineering and Applications* **2024**, *17*, 817–831.
9. Weng, Y.; Wu, J. Leveraging artificial intelligence to enhance data security and combat cyber attacks. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023* **2024**, *5*, 392–399.
10. Li, K.; Liu, L.; Chen, J.; Yu, D.; Zhou, X.; Li, M.; Wang, C.; Li, Z. Research on reinforcement learning based warehouse robot navigation algorithm in complex warehouse layout. In Proceedings of the 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA). IEEE, 2024, pp. 296–301.
11. Zhao, Y.; Chen, P.; Cao, Z.; Wang, X. NoncovANM: Gridless DOA estimation for LPDF system. *IEEE Transactions on Vehicular Technology* **2023**, *72*, 15563–15573.
12. Indhumathi, K.; Kumar, K.S. Seasonal Infectious Disease Prediction based on Electronic Patient Health Records using Boosted Random Forest Algorithms. In Proceedings of the 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). IEEE, 2022, pp. 2025–2032.
13. Nikhil, V.; Mary, S.S.A.; Natarajan, B.; Bajulunisha, A.; Anuradha, R.; Venkatraman, K. Revolutionizing Influenza A Detection in Humans: Unleashing the Power of XGBoost Classification. In Proceedings of the 2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE). IEEE, 2024, pp. 229–234.
14. Shaban, W.M.; Rabie, A.H.; Saleh, A.I.; Abo-Elsoud, M. A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier. *Knowledge-Based Systems* **2020**, *205*, 106270.
15. Yoshikawa, H. Can naive Bayes classifier predict infection in a close contact of COVID-19? A comparative test for predictability of the predictive model and healthcare workers in Japan. *Journal of Infection and Chemotherapy* **2022**, *28*, 774–779.
16. Shaier, S.; Raissi, M.; Seshaiyer, P. Data-driven approaches for predicting spread of infectious diseases through DINNs: Disease Informed Neural Networks. *arXiv preprint arXiv:2110.05445* **2021**.
17. Ramón, A.; Torres, A.M.; Milara, J.; Cascón, J.; Blasco, P.; Mateo, J. eXtreme Gradient Boosting-based method to classify patients with COVID-19. *Journal of Investigative Medicine* **2022**, *70*, 1472–1480.

18. Sartakhti, J.S.; Zangooei, M.H.; Mozafari, K. Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA). *Computer methods and programs in biomedicine* **2012**, *108*, 570–579.
19. Dogantekin, E.; Dogantekin, A.; Avci, D. Automatic hepatitis diagnosis system based on linear discriminant analysis and adaptive network based on fuzzy inference system. *Expert Systems with Applications* **2009**, *36*, 11282–11286.
20. Tang, J.; Qian, W.; Song, L.; Dong, X.; Li, L.; Bai, X. Optimal boxes: boosting end-to-end scene text recognition by adjusting annotated bounding boxes via reinforcement learning. In Proceedings of the European Conference on Computer Vision. Springer, 2022, pp. 233–248.
21. Mao, Y.; Tao, D.; Zhang, S.; Qi, T.; Li, K. Research and Design on Intelligent Recognition of Unordered Targets for Robots Based on Reinforcement Learning. *arXiv preprint arXiv:2503.07340* **2025**.
22. Zhong, J.; Wang, Y. Enhancing Thyroid Disease Prediction Using Machine Learning: A Comparative Study of Ensemble Models and Class Balancing Techniques **2025**.
23. Wan, R.; Zhang, X.; Song, R. Multi-objective model-based reinforcement learning for infectious disease control. In Proceedings of the Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining, 2021, pp. 1634–1644.
24. Mishra, S.; Kumar, R.; Tiwari, S.K.; Ranjan, P. Machine learning approaches in the diagnosis of infectious diseases: a review. *Bulletin of Electrical Engineering and Informatics* **2022**, *11*, 3509–3520. https://doi.org/10.11591/eei.v11i6.4406.
25. Li, X.; Mohammed, R.; Mangin, T.; Saha, S.; Kelly, K.; Whitaker, R.; Tasdizen, T. Joint audio-visual idling vehicle detection with streamlined input dependencies. In Proceedings of the Proceedings of the Winter Conference on Applications of Computer Vision, 2025, pp. 885–894.
26. Lu, Z.; Lu, B.; Wang, F. CausalSR: Structural Causal Model-Driven Super-Resolution with Counterfactual Inference. *arXiv preprint arXiv:2501.15852* **2025**.
27. Dan, H.C.; Lu, B.; Li, M. Evaluation of asphalt pavement texture using multiview stereo reconstruction based on deep learning. *Construction and Building Materials* **2024**, *412*, 134837.
28. Li, X.; Mangin, T.; Saha, S.; Mohammed, R.; Blanchard, E.; Tang, D.; Poppe, H.; Choi, O.; Kelly, K.; Whitaker, R. Real-time idling vehicles detection using combined audio-visual deep learning. In *Emerging Cutting-Edge Developments in Intelligent Traffic and Transportation Systems*; IOS Press, 2024; pp. 142–158.
29. Dan, H.C.; Huang, Z.; Lu, B.; Li, M. Image-driven prediction system: Automatic extraction of aggregate gradation of pavement core samples integrating deep learning and interactive image processing framework. *Construction and Building Materials* **2024**, *453*, 139056.
30. Zhao, Y.; Chen, P.; Cao, Z. NFLnet: A Deep Neural Network-Based Localization Method for Near-Field Vehicles. *IEEE Transactions on Instrumentation and Measurement* **2024**.
31. Li, Z.; Qiu, S.; Ke, Z. Revolutionizing Drug Discovery: Integrating Spatial Transcriptomics with Advanced Computer Vision Techniques. In Proceedings of the 1st CVPR Workshop on Computer Vision For Drug Discovery (CVDD): Where are we and What is Beyond?
32. Weng, Y.; Wu, J.; Kelly, T.; Johnson, W. Comprehensive overview of artificial intelligence applications in modern industries. *arXiv preprint arXiv:2409.13059* **2024**.
33. Liang, Z.; Powell, A.; Ersoy, I.; Poostchi, M.; Silamut, K.; Palaniappan, K.; Guo, P.; Hossain, M.A.; Sameer, A.; Maude, R.J.; et al. CNN-based image analysis for malaria diagnosis. In Proceedings of the 2016 IEEE international conference on bioinformatics and biomedicine (BIBM). IEEE, 2016, pp. 493–496.
34. Hira, S.; Bai, A.; Hira, S. An automatic approach based on CNN architecture to detect Covid-19 disease from chest X-ray images. *Applied Intelligence* **2021**, *51*, 2864–2889.
35. Bukhari, S.U.K.; Bukhari, S.S.K.; Syed, A.; Shah, S.S.H. The diagnostic evaluation of Convolutional Neural Network (CNN) for the assessment of chest X-ray of patients infected with COVID-19. *MedRxiv* **2020**, pp. 2020–03.
36. Demir, F. DeepCoroNet: A deep LSTM approach for automated detection of COVID-19 cases from chest X-ray images. *Applied soft computing* **2021**, *103*, 107160.
37. Showkat, S.; Qureshi, S. Efficacy of Transfer Learning-based ResNet models in Chest X-ray image classification for detecting COVID-19 Pneumonia. *Chemometrics and Intelligent Laboratory Systems* **2022**, *224*, 104534.
38. Bharati, S.; Podder, P.; Mondal, M.R.H.; Prasath, V.S. CO-ResNet: Optimized ResNet model for COVID-19 diagnosis from X-ray images. *International Journal of Hybrid Intelligent Systems* **2021**, *17*, 71–85.
39. Farooq, M.; Hafeez, A. Covid-resnet: A deep learning framework for screening of covid19 from radiographs. *arXiv preprint arXiv:2003.14395* **2020**.

40. Islam, M.Z.; Islam, M.M.; Asraf, A. A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Informatics in medicine unlocked* **2020**, *20*, 100412.
41. Mehta, T.; Mehendale, N. Classification of X-ray images into COVID-19, pneumonia, and TB using cGAN and fine-tuned deep transfer learning models. *Research on Biomedical Engineering* **2021**, *37*, 803–813.
42. Kalane, P.; Patil, S.; Patil, B.; Sharma, D.P. Automatic detection of COVID-19 disease using U-Net architecture based fully convolutional network. *Biomedical Signal Processing and Control* **2021**, *67*, 102518.
43. Yu, Q.; Ke, Z.; Xiong, G.; Cheng, Y.; Guo, X. Identifying money laundering risks in digital asset transactions based on ai algorithms. In Proceedings of the 2024 4th International Conference on Electronic Information Engineering and Computer Communication (EIECC). IEEE, 2024, pp. 1081–1085.
44. Ke, Z.; Zhou, S.; Zhou, Y.; Chang, C.H.; Zhang, R. Detection of ai deepfake and fraud in online payments using gan-based models. *arXiv preprint arXiv:2501.07033* **2025**.
45. Zhao, Y.; Chen, P.; Sun, M.; Luo, T.; Cao, Z. RIS-aided sensing system with localization function: Fundamental and practical design. *IEEE Sensors Journal* **2023**, *24*, 506–514.
46. Albahli, S. Efficient GAN-based Chest Radiographs (CXR) augmentation to diagnose coronavirus disease pneumonia. *International journal of medical sciences* **2020**, *17*, 1439.
47. Motamed, S.; Rogalla, P.; Khalvati, F. RANDGAN: randomized generative adversarial network for detection of COVID-19 in chest X-ray. *Scientific Reports* **2021**, *11*, 8602.
48. Mansour, R.F.; Escorcia-Gutierrez, J.; Gamarra, M.; Gupta, D.; Castillo, O.; Kumar, S. Unsupervised deep learning based variational autoencoder model for COVID-19 diagnosis and classification. *Pattern Recognition Letters* **2021**, *151*, 267–274.
49. Centres, P.; Perez-Morelo, D.; Guzman, R.; Reinaudi, L.; Gimenez, M. Diffusion model for the spread of infectious diseases: SIR model with mobile agents. *Physica A: Statistical Mechanics and its Applications* **2024**, *633*, 129399.
50. Pranolo, A.; Mao, Y.; et al. CAE-COVIDX: automatic covid-19 disease detection based on x-ray images using enhanced deep convolutional and autoencoder. *International Journal of Advances in Intelligent Informatics* **2021**, *7*.
51. Tang, J.; Zhang, W.; Liu, H.; Yang, M.; Jiang, B.; Hu, G.; Bai, X. Few could be better than all: Feature sampling and grouping for scene text detection. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 4563–4572.
52. Zhao, Z.; Tang, J.; Lin, C.; Wu, B.; Huang, C.; Liu, H.; Tan, X.; Zhang, Z.; Xie, Y. Multi-modal in-context learning makes an ego-evolving scene text recognizer. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 15567–15576.
53. You, W.; Hooi, B.; Wang, Y.; Wang, Y.; Ke, Z.; Yang, M.H.; Huang, Z.; Cai, Y. MIRAGE: Multimodal Immersive Reasoning and Guided Exploration for Red-Team Jailbreak Attacks. *arXiv preprint arXiv:2503.19134* **2025**.
54. Tang, J.; Liu, Q.; Ye, Y.; Lu, J.; Wei, S.; Lin, C.; Li, W.; Mahmood, M.F.F.B.; Feng, H.; Zhao, Z.; et al. Mtvqa: Benchmarking multilingual text-centric visual question answering. *arXiv preprint arXiv:2405.11985* **2024**.
55. Zhao, Z.; Tang, J.; Wu, B.; Lin, C.; Wei, S.; Liu, H.; Tan, X.; Zhang, Z.; Huang, C.; Xie, Y. Harmonizing visual text comprehension and generation. *arXiv preprint arXiv:2407.16364* **2024**.
56. Puleio, A. Recurrent neural network ensemble, a new instrument for the prediction of infectious diseases. *The European Physical Journal Plus* **2021**, *136*, 319.
57. Absar, N.; Uddin, N.; Khandaker, M.U.; Ullah, H. The efficacy of deep learning based LSTM model in forecasting the outbreak of contagious diseases. *Infectious Disease Modelling* **2022**, *7*, 170–183.
58. Chandra, R.; Jain, A.; Singh Chauhan, D. Deep learning via LSTM models for COVID-19 infection forecasting in India. *PloS one* **2022**, *17*, e0262708.
59. Erraguntla, M.; Zapletal, J.; Lawley, M. Framework for Infectious Disease Analysis: A comprehensive and integrative multi-modeling approach to disease prediction and management. *Health informatics journal* **2019**, *25*, 1170–1187.
60. Kim, M.; Chae, K.; Lee, S.; Jang, H.J.; Kim, S. Automated classification of online sources for infectious disease occurrences using machine-learning-based natural language processing approaches. *International Journal of Environmental Research and Public Health* **2020**, *17*, 9467.
61. Barman, R.; Rajamohan, K. Detecting Infectious Disease Based on Social Media Data Using BERT Model. In Proceedings of the 2024 IEEE International Conference on Contemporary Computing and Communications (InC4). IEEE, 2024, Vol. 1, pp. 1–6.
62. Su, M.; Cheng, D.; Xu, Y.; Weng, F. An improved BERT method for the evolution of network public opinion of major infectious diseases: Case Study of COVID-19. *Expert Systems with Applications* **2023**, *233*, 120938.

63. Agarwal, K.; Choudhury, S.; Tipirneni, S.; Mukherjee, P.; Ham, C.; Tamang, S.; Baker, M.; Tang, S.; Kocaman, V.; Gevaert, O.; et al. Preparing for the next pandemic via transfer learning from existing diseases with hierarchical multi-modal BERT: a study on COVID-19 outcome prediction. *Scientific reports* **2022**, *12*, 10748.
64. Cheng, K.; Li, Z.; He, Y.; Guo, Q.; Lu, Y.; Gu, S.; Wu, H. Potential use of artificial intelligence in infectious disease: take ChatGPT as an example. *Annals of biomedical engineering* **2023**, *51*, 1130–1135.
65. Adams, D.A. Summary of Notifiable Infectious Diseases and Conditions — United States, 2014. *MMWR. Morbidity and Mortality Weekly Report* **2016**, *63*.
66. Zhao, C.; Wu, Z.; Pan, B.; Zhang, R.; Golestani, A.; Feng, Z.; Ge, Y.; Yang, H. Functional biomacromolecules-based microneedle patch for the treatment of diabetic wound. *International Journal of Biological Macromolecules* **2024**, p. 131650.
67. Lu, Y.; Wu, C.T.; Parker, S.J.; Cheng, Z.; Saylor, G.; Van Eyk, J.E.; Yu, G.; Clarke, R.; Herrington, D.M.; Wang, Y. COT: an efficient and accurate method for detecting marker genes among many subtypes. *Bioinformatics Advances* **2022**, *2*, vbac037.
68. Zhao, C.; Pan, B.; Wang, T.; Yang, H.; Vance, D.; Li, X.; Zhao, H.; Hu, X.; Yang, T.; Chen, Z.; et al. Advances in NIR-responsive natural macromolecular hydrogel assembly drugs for cancer treatment. *Pharmaceutics* **2023**, *15*, 2729.
69. Chen, L.; Lu, Y.; Wu, C.T.; Clarke, R.; Yu, G.; Van Eyk, J.E.; Herrington, D.M.; Wang, Y. Data-driven detection of subtype-specific differentially expressed genes. *Scientific reports* **2021**, *11*, 332.
70. Zhao, C.; Feng, M.; Gluchman, M.; Ma, X.; Li, J.; Wang, H. Acellular fish skin grafts in the treatment of diabetic wounds: Advantages and clinical translation. *Journal of Diabetes* **2024**, *16*, e13554.
71. Du, D.; Bhardwaj, S.; Lu, Y.; Wang, Y.; Parker, S.J.; Zhang, Z.; Van Eyk, J.E.; Yu, G.; Clarke, R.; Herrington, D.M.; et al. Embracing the informative missingness and silent gene in analyzing biologically diverse samples. *Scientific reports* **2024**, *14*, 28265.
72. Nguyen, L.T.; Hains, A.E.; Aziz-Zanjani, M.O.; Dalsass, M.; Farooqee, S.B.; Lu, Y.; Jackson, P.K.; Van Rechem, C. Absence of SMARCB1 in rhabdoid tumor cells increases sensitivity to translation inhibition and alters translation efficiency of specific mRNAs. *Journal of Biological Chemistry* **2024**, *300*.
73. Zhao, C.; Zhang, R.; Yang, H.; Gao, Y.; Zou, Y.; Zhang, X. Antibody-drug conjugates for non-small cell lung cancer: Advantages and challenges in clinical translation. *Biochemical Pharmacology* **2024**, p. 116378.
74. Parker, S.; Mao, C.; Wang, Y.; Seals, A.; Kim, D.K.; Lu, Y.; bhardwaj, s.; Du, D.; Karere, G.; Caudell, D.; et al. Mapping the Molecular Progression of Human Coronary Atherosclerosis Confirms Key Role of Smooth Muscle Cell Phenotype and Highlights Novel Regulators of Phenotypic Fate. *Circulation* **2024**, *150*, A4139230–A4139230.
75. Zhao, C.; Fan, Z.; Zhang, R.; Sun, Y.; Li, W.Y. The Role of the Major Histocompatibility Complex Region on Chromosome 6 in Skin Atrophy: A Mendelian Randomization Study. *Journal of Cosmetic Dermatology* **2025**, *24*, e70040.
76. Yang, J.M.; Zhang, N.; Luo, T.; Yang, M.; Shen, W.K.; Tan, Z.L.; Xia, Y.; Zhang, L.; Zhou, X.; Lei, Q.; et al. TCellSI: A novel method for T cell state assessment and its applications in immune environment prediction. *Imeta* **2024**, *3*, e231.
77. World Health Organization. Coronavirus disease (COVID-19), 2020. Accessed: February 7, 2025.
78. Rothan, H.A.; Byrareddy, S.N. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of autoimmunity* **2020**, *109*, 102433.
79. World Health Organization. WHO Coronavirus (COVID-19) Dashboard: Death Toll, 2025. Accessed: February 7, 2025.
80. Mathieu, E.; Ritchie, H.; Rodés-Guirao, L.; Appel, C.; Gavrilov, D.; Giattino, C.; Hasell, J.; Macdonald, B.; Dattani, S.; Beltekian, D.; et al. COVID-19 Pandemic. *Our World in Data* **2020**. https://ourworldindata.org/coronavirus.
81. Santosh, K. COVID-19 prediction models and unexploited data. *Journal of medical systems* **2020**, *44*, 170.
82. Moulaei, K.; Shanbehzadeh, M.; Mohammadi-Taghiabad, Z.; Kazemi-Arpanahi, H. Comparing machine learning algorithms for predicting COVID-19 mortality. *BMC medical informatics and decision making* **2022**, *22*, 2.
83. Arpaci, I.; Huang, S.; Al-Emran, M.; Al-Kabi, M.N.; Peng, M. Predicting the COVID-19 infection with fourteen clinical features using machine learning classification algorithms. *Multimedia Tools and Applications* **2021**, *80*, 11943–11957.
84. Ardabili, S.F.; Mosavi, A.; Ghamisi, P.; Ferdinand, F.; Varkonyi-Koczy, A.R.; Reuter, U.; Rabczuk, T.; Atkinson, P.M. Covid-19 outbreak prediction with machine learning. *Algorithms* **2020**, *13*, 249.

85. Pinter, G.; Felde, I.; Mosavi, A.; Ghamisi, P.; Gloaguen, R. COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. *Mathematics* **2020**, *8*, 890.
86. Dairi, A.; Harrou, F.; Zeroual, A.; Hittawe, M.M.; Sun, Y. Comparative study of machine learning methods for COVID-19 transmission forecasting. *Journal of biomedical informatics* **2021**, *118*, 103791.
87. World Health Organization. Influenza (Seasonal), 2023. Accessed: February 7, 2025.
88. Centers for Disease Control and Prevention. Influenza (Flu) Burden in the U.S., 2023. Accessed: February 7, 2025.
89. Alessa, A.; Faezipour, M.; et al. Preliminary flu outbreak prediction using twitter posts classification and linear regression with historical centers for disease control and prevention reports: Prediction framework study. *JMIR public health and surveillance* **2019**, *5*, e12383.
90. Khan, M.A.; Abidi, W.U.H.; Al Ghamdi, M.A.; Almotiri, S.H.; Saqib, S.; Alyas, T.; Khan, K.M.; Mahmood, N. Forecast the influenza pandemic using machine learning. *Computers, Materials and Continua* **2020**, *66*, 331–340.
91. Zhang, J.; Nawata, K. A comparative study on predicting influenza outbreaks. *Bioscience trends* **2017**, *11*, 533–541.
92. Allen, C.; Tsou, M.H.; Aslam, A.; Nagel, A.; Gawron, J.M. Applying GIS and machine learning methods to Twitter data for multiscale surveillance of influenza. *PloS one* **2016**, *11*, e0157734.
93. Amin, S.; Uddin, M.I.; AlSaeed, D.H.; Khan, A.; Adnan, M. Early detection of seasonal outbreaks from twitter data using machine learning approaches. *Complexity* **2021**, *2021*, 5520366.
94. World Health Organization. Influenza A (H1N1) outbreak, 2009. Accessed: February 7, 2025.
95. Inampudi, S.; Johnson, G.; Jhaveri, J.; Niranjan, S.; Chaurasia, K.; Dixit, M. Machine learning based prediction of h1n1 and seasonal flu vaccination. In Proceedings of the Advanced Computing: 10th International Conference, IACC 2020, Panaji, Goa, India, December 5–6, 2020, Revised Selected Papers, Part I 10. Springer, 2021, pp. 139–150.
96. Ayachit, S.S.; Kumar, T.; Deshpande, S.; Sharma, N.; Chaurasia, K.; Dixit, M. Predicting h1n1 and seasonal flu: Vaccine cases using ensemble learning approach. In Proceedings of the 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN). IEEE, 2020, pp. 172–176.
97. World Health Organization. HIV/AIDS **2024**. Accessed: February 5, 2025.
98. World Health Organization. HIV/AIDS Key Facts, 2024. Accessed: February 11, 2025.
99. Centers for Disease Control and Prevention. HIV Diagnoses, Deaths, and Prevalence, 2023. Accessed: February 11, 2025.
100. Wang, B.; Liu, F.; Deveaux, L.; Ash, A.; Gosh, S.; Li, X.; Rundensteiner, E.; Cottrell, L.; Adderley, R.; Stanton, B. Adolescent HIV-related behavioural prediction using machine learning: a foundation for precision HIV prevention. *Aids* **2021**, *35*, S75–S84.
101. Pan, Y.; Liu, H.; Metsch, L.R.; Feaster, D.J. Factors associated with HIV testing among participants from substance use disorder treatment programs in the US: A machine learning approach. *AIDS and Behavior* **2017**, *21*, 534–546.
102. Nisa, S.U.; Mahmood, A.; Ujager, F.S.; Malik, M. HIV/AIDS predictive model using random forest based on socio-demographical, biological and behavioral data. *Egyptian Informatics Journal* **2023**, *24*, 107–115.
103. Centers for Disease Control and Prevention. HIV Surveillance Report, 2020 **2022**. Accessed: 2025-02-11.
104. Bao, Y.; Medland, N.A.; Fairley, C.K.; Wu, J.; Shang, X.; Chow, E.P.; Xu, X.; Ge, Z.; Zhuang, X.; Zhang, L. Predicting the diagnosis of HIV and sexually transmitted infections among men who have sex with men using machine learning approaches. *Journal of Infection* **2021**, *82*, 48–59.
105. Chingombe, I.; Dzinamarira, T.; Cuadros, D.; Mapingure, M.P.; Mbunge, E.; Chaputsira, S.; Madziva, R.; Chiurunge, P.; Samba, C.; Herrera, H.; et al. Predicting HIV status among men who have sex with men in Bulawayo & Harare, Zimbabwe using bio-behavioural data, recurrent neural networks, and machine learning techniques. *Tropical Medicine and Infectious Disease* **2022**, *7*, 231.
106. Turbé, V.; Herbst, C.; Mngomezulu, T.; Meshkinfamfard, S.; Dlamini, N.; Mhlongo, T.; Smit, T.; Cherepanova, V.; Shimada, K.; Budd, J.; et al. Deep learning of HIV field-based rapid tests. *Nature medicine* **2021**, *27*, 1165–1170.
107. Wang, G.; Wei, W.; Jiang, J.; Ning, C.; Chen, H.; Huang, J.; Liang, B.; Zang, N.; Liao, Y.; Chen, R.; et al. Application of a long short-term memory neural network: a burgeoning method of deep learning in forecasting HIV incidence in Guangxi, China. *Epidemiology & Infection* **2019**, *147*, e194.
108. Bhirud, P.; Joshi, A.; Hirani, N.; Chowdhary, A. Rapid laboratory diagnosis of pulmonary tuberculosis. *The International Journal of Mycobacteriology* **2017**, *6*, 296–301.

109. Singh, M.; Pujar, G.V.; Kumar, S.A.; Bhagyalalitha, M.; Akshatha, H.S.; Abuhaija, B.; Alsoud, A.R.; Abualigah, L.; Beeraka, N.M.; Gandomi, A.H. Evolution of machine learning in tuberculosis diagnosis: a review of deep learning-based medical applications. *Electronics* **2022**, *11*, 2634.
110. Rabehi, A.; Kumar, P. Improving tuberculosis diagnosis and forecasting through machine learning techniques: A systematic review. *Metaheuristic Optim. Rev.* **2024**, *1*, 35–44.
111. Asada, N.; Doi, K.; MacMahon, H.; Montner, S.; Giger, M.; Abe, C.; Wu, Y. Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung diseases: pilot study. *Radiology* **1990**, *177*, 857–860.
112. El-Solh, A.A.; Hsiao, C.B.; Goodnough, S.; Serghani, J.; Grant, B.J. Predicting active pulmonary tuberculosis using an artificial neural network. *Chest* **1999**, *116*, 968–973.
113. El-Solh, A.; Mylotte, J.; Sherif, S.; Serghani, J.; Grant, B. Validity of a decision tree for predicting active pulmonary tuberculosis. *American journal of respiratory and critical care medicine* **1997**, *155*, 1711–1716.
114. Elveren, E.; Yumuşak, N. Tuberculosis disease diagnosis using artificial neural network trained with genetic algorithm. *Journal of medical systems* **2011**, *35*, 329–332.
115. Hooda, R.; Sofat, S.; Kaur, S.; Mittal, A.; Meriaudeau, F. Deep-learning: A potential method for tuberculosis detection using chest radiography. In Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2017, pp. 497–502. https://doi.org/10.1109/ICSIPA.2017.8120663.
116. Kant, S.; Srivastava, M.M. Towards Automated Tuberculosis detection using Deep Learning. In Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence (SSCI), 2018, pp. 1250–1253. https://doi.org/10.1109/SSCI.2018.8628800.
117. Hrizi, O.; Gasmi, K.; Ben Ltaifa, I.; Alshammari, H.; Karamti, H.; Krichen, M.; Ben Ammar, L.; Mahmood, M.A. Tuberculosis disease diagnosis based on an optimized machine learning model. *Journal of Healthcare Engineering* **2022**, *2022*, 8950243.
118. Hansun, S.; Argha, A.; Liaw, S.T.; Celler, B.G.; Marks, G.B. Machine and deep learning for tuberculosis detection on chest x-rays: systematic literature review. *Journal of medical Internet research* **2023**, *25*, e43154.
119. Ramrakhiani, N.S.; Chen, V.L.; Le, M.; Yeo, Y.H.; Barnett, S.D.; Waljee, A.K.; Zhu, J.; Nguyen, M.H. Optimizing hepatitis B virus screening in the United States using a simple demographics-based model. *Hepatology* **2022**, *75*, 430–437.
120. Saleem, H. Hepatitis Diagnosis: A Comprehensive Review of Machine Learning Classification Algorithms. *The Indonesian Journal of Computer Science* **2024**, *13*.
121. Ali, G.; Mijwil, M.M.; Adamopoulos, I.; Buruga, B.A.; Gök, M.; Sallam, M. Harnessing the potential of artificial intelligence in managing viral hepatitis. *Mesopotamian Journal of Big Data* **2024**, *2024*, 128–163.
122. Bharathi, P.T.; Bindu, S.N.; Deepthi, S.G.; Gunakeerthi, H.U.; Harshitha, K.U. AI based solution for Predicting Hepatitis C Virus from Blood Samples. In Proceedings of the 2024 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES), 2024, pp. 1–6. https://doi.org/10.1109/ICSSES62373.2024.10561391.
123. Yağanoğlu, M. Hepatitis C virus data analysis and prediction using machine learning. *Data & Knowledge Engineering* **2022**, *142*, 102087. https://doi.org/https://doi.org/10.1016/j.datak.2022.102087.
124. Harabor, V.; Mogos, R.; Nechita, A.; Adam, A.M.; Adam, G.; Melinte-Popescu, A.S.; Melinte-Popescu, M.; Stuparu-Cretu, M.; Vasilache, I.A.; Mihalceanu, E.; et al. Machine learning approaches for the prediction of hepatitis B and C seropositivity. *International journal of environmental research and public health* **2023**, *20*, 2380.
125. Wang, X.; Tian, S.; Yu, L.; Lv, X.; Zhang, Z. Rapid screening of hepatitis B using Raman spectroscopy and long short-term memory neural network. *Lasers in medical science* **2020**, *35*, 1791–1799.
126. Chen, L.; Ji, P.; Ma, Y. Machine Learning Model for Hepatitis C Diagnosis Customized to Each Patient. *IEEE Access* **2022**, *10*, 106655–106672. https://doi.org/10.1109/ACCESS.2022.3210347.
127. Guo, Y.; Feng, Y.; Qu, F.; Zhang, L.; Yan, B.; Lv, J. Prediction of hepatitis E using machine learning models. *Plos one* **2020**, *15*, e0237750.