

Predicting Disease Outbreaks: A Comprehensive Survey and A Proposed Framework for Early Detection

Reham Abdallah¹, Sayed A. AbdelGaber¹, and Hanan Ali Sayed²

¹Information System Dept., Faculty of Computers and Artificial Intelligence, Helwan University, Egypt

²Public Health Dept., Faculty of Medicine, Helwan University, Egypt

Reham-abdallah@eru.edu.eg, sgaber14@gmail.com, Hananali1993@yahoo.com

Abstract— In recent years, the COVID-19 pandemic has emerged as a global crisis, underscoring the importance of early detection and analysis in controlling disease outbreaks. However, due to high uncertainty and a lack of essential outbreak data, traditional models have struggled with accuracy in long-term predictions. While the literature review highlights various attempts to address this challenge, existing models still require improvement in terms of generalization and robustness. Recent studies suggest that Machine Learning (ML) techniques offer a promising approach to analyzing health-related data, enabling the identification of potential disease outbreaks, facilitating timely interventions, and ultimately reducing healthcare costs. This research seeks to evaluate the performance and predictive capabilities of various machine learning algorithms to determine the most accurate and reliable models for disease prediction. The findings aim to propose a novel framework for early outbreak detection using ML techniques and to conduct a comparative analysis of studies that have employed ML for detecting disease outbreaks.

Keywords—Public Health, Epidemic, Outbreak, Risk Factors, Machine Learning.

I. INTRODUCTION

Imagine a world free from the global crises caused by infectious diseases. The year 2020 brought unprecedented disruptions as the rapid spread of COVID-19 exposed vulnerabilities in economies, healthcare systems, and daily life. The swift progression of the virus strained existing strategies, underscoring the urgent need for innovative and adaptable solutions. In the event of an infectious outbreak, it is imperative for governments and authorities to enforce specific regulations across society. Infectious diseases remain a constant threat to global public health, often leading to widespread illness, fatalities, and significant economic strain. [1].

Controlling and accurately predicting disease outbreaks has always been a top priority in public health. Accurate predictions are essential for public health professionals to implement effective prevention and control measures and to make well-informed decisions.

Manually detecting pandemics and infectious diseases is a time-consuming process that is prone to human error, presenting a significant challenge [2]. As a result, the use of Artificial Intelligence (AI) has become urgently necessary to address this issue. Predicting the course of an outbreak, identifying at-risk populations, efficiently allocating resources, and implementing effective containment strategies are complex tasks that require real-time analysis of large and diverse datasets [3]. AI is widely used across various industries and has a significant impact on the field of public health. Predictive models, developed using machine learning algorithms, have been applied in both infectious and non-infectious contexts, such as using image analysis to forecast disease progression. Accurate outbreak prediction models are essential for gaining a deeper understanding of the potential spread and impact of infectious diseases. Leveraging AI and ML to enhance public health offers an opportunity to bridge the gap between scientific predictions and their real-world application. [4].

In public health, understanding the terms outbreak, epidemic, pandemic, and endemic is essential for grasping the dynamics of diseases and their impact on populations. An epidemic refers to a rapid increase in the number of cases of a particular disease within a specific geographic area or community, typically exceeding the normal expected levels. When an epidemic spreads across multiple countries and regions, it is classified as a pandemic. In contrast, an endemic disease is characterized by a relatively stable, consistent presence within a particular population or region, without the sharp increases observed in epidemics [5].

Finally, an outbreak is a localized rise in disease cases that may or may not develop into an epidemic, as shown in Fig 1. The purpose of the research is to further our understanding of the potential applications of machine learning as a preventative measure against disease outbreaks.

The present study attempts to investigate the subsequent research questions:

- Which machine learning techniques are applied to diagnose infectious diseases?
- Which dataset sources are utilized in the models that are predicted?
- To what extent do the methods employed for illness diagnostic models perform?

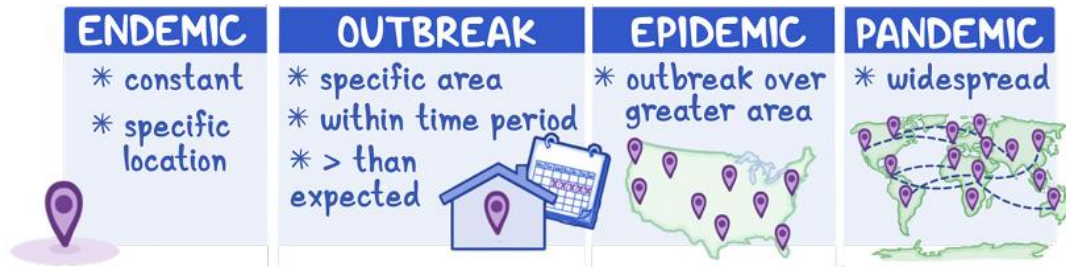


Fig 1: Summary of Pandemic, Epidemic, Outbreak and Endemic [6]

II. BACKGROUND OF STUDY

To minimize the risk of communicable disease outbreaks, it is essential to understand the key risk factors associated with infectious diseases. Consequently, it is crucial to evaluate how socioeconomic, demographic, geographic, climatic, behavioral, and health-related factors impact the patterns of these diseases.

A. Public Health risk factors

Risk factors for public health are important in determining how healthy a population is. A health risk factor is any characteristic, action, or state that raises an individual's chance of developing a certain illness or health issue. Numerous influences, including socioeconomic, environmental, genetic, and demographic factors, might be included in this category of risk factors [7]. Thus, it is essential to identify these risk variables in order to develop effective public health interventions and to avoid diseases. Based on shared traits and influences, risk factors can be generally divided into multiple groupings, which include the following:

1) Behavioural risk factors

These are connected to how people behave and make lifestyle choices. The prognoses and susceptibility of common infectious diseases are significantly shaped by behavioral variables. Among these are things like obesity, smoking, alcohol consumption, and physical inactivity [8].

2) Demographic risk factors

These elements are strongly related to the socioeconomic status of people, places, or eras. These elements have a big impact on how infectious diseases propagate and affect people. Age, sex, population density, occupation, education level, and ethnicity are important demographic risk variables [9] [61].

3) Environmental risk factors

Environmental risk factors are those outside circumstances or environmental components that have the potential to raise the probability of unfavorable health

outcomes, environmental issues, or detrimental effects on ecosystems. These elements have the potential to contribute to issues including pollution, climate change, habitat destruction, and disease transmission [10].

B. Outbreak / Epidemic Stages

When the quantity of instances in a certain area or population during a given period exceeds what would be expected, it is called an epidemic or outbreak of disease, as illustrated in Fig 2 [11].

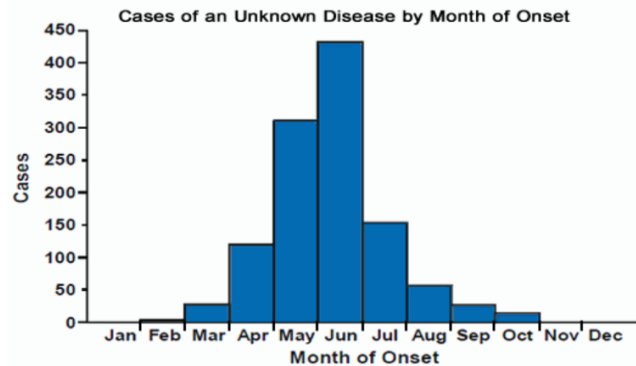


Fig 2: Trend of the outbreak throughout time [11]

Epidemics usually develop in phases, each distinguished by unique dynamics and characteristics. These stages are essential for understanding and controlling the outbreak and directing researchers' and public health officials' actions. The following are the stages of the epidemic investigation [12] [13]:

1. Preparation for the investigation
2. Confirm outbreak and diagnosis.
3. Develop case definition.
4. Formulate hypothesis.
5. Testing hypothesis
6. Conclusion and recommendations for control and prevention.
7. Communicate the findings.

C. An Overview of Machine Learning(ML) in the Health Sector

As illustrated in Fig. 3, AI is a broad area of computer science that includes various specialized areas such as image processing, machine learning, deep learning, artificial neural networks, and cloud-based AI solutions. ML is a subset of AI, that focuses on building models and algorithms that let computers learn from data and without explicit programming, make predictions or conclusions [1].

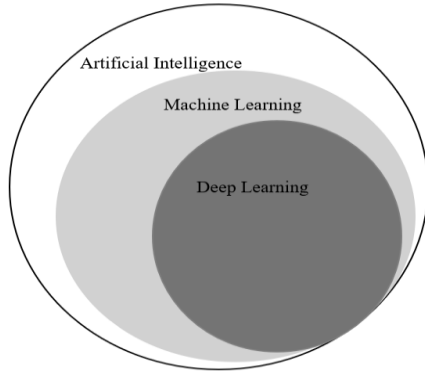


Fig3: The link between AL, ML, DL

Machine learning has demonstrated substantial benefits in the medical sector, especially in areas like disease prediction and medical diagnosis. ML algorithms can evaluate huge and complicated datasets, identifying patterns and trends that could be hard for people to observe. The three primary categories into which these algorithms are typically divided are reinforcement learning, unsupervised learning, and supervised learning, as depicted in Fig 4.

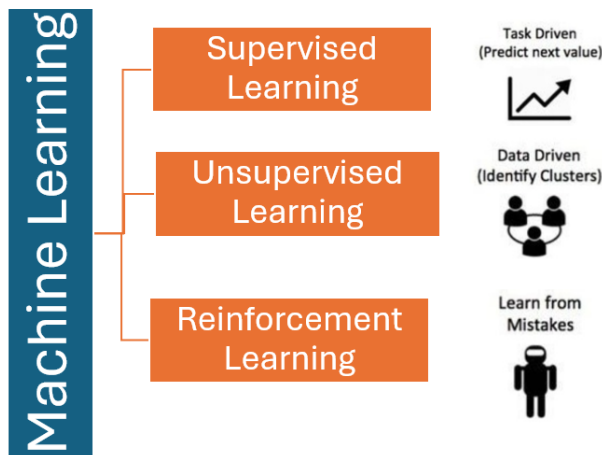


Fig 4: Machine Learning categories

Supervised learning involves training a model to link input data with corresponding target labels using a labeled dataset. In this method, if the target variables are continuous real numbers, the task is referred to as regression. Conversely, when the target variables are categorical, it is known as

classification. Unsupervised learning, however, focuses on identifying patterns in datasets that consist only of input data, without specific objective labels. Reinforcement learning is teaching sentient agents to learn from their mistakes, keep trying to achieve the best result and maximize rewards over time.

This research will concentrate specifically on supervised learning, which includes both classification and regression techniques. The following section will outline the most commonly used methods in these categories.

1. *Logistic Regression (LR)* is a statistical technique that's applied to forecast binary outcomes, where the dependent variable can take one of two possible values, usually represented as 0 and 1. Many tools, such as R, Python, Java, and MATLAB, can be used to implement logistic regression [15][16][17].
2. *Naïve Bayes (NB)*: is a classification technique that utilizes Bayes' theorem, enabling the estimation of an event's probability using past knowledge of relevant circumstances. NB is a popular ML algorithm for the two binary and multiclass classification tasks, known for its simplicity, speed, and efficiency. This makes it particularly well-suited for a variety of applications, especially when working with small datasets [15][16].
3. *K-nearest Neighbors (KNN)*: is a simple, conventional classification algorithm. It utilizes a "voting" mechanism among the 'K' nearest neighbors as the basis for its classification conclusion. The approach can be used for tasks involving continuous qualities and computes distances using the Euclidean distance metric [16].
4. *Decision Tree (DT)*: is a popular algorithm in machine learning and data mining, valued for its simplicity and adaptability in managing activities related to regression and classification. This algorithm uses a structure like to a tree to represent decision logic, where data items are classified through nodes, with the top node known as the root node. Decision Trees can be deployed using different tools, including Python with Scikit-Learn, R Studio, Orange, KNIME, and Weka [16][17].
5. *Support Vector Machines (SVM)*: is a supervised learning technique that is applied to regression and classification tasks, especially in the medical diagnosis industry. By using different kernel functions, SVM can handle data that is both linear and non-linear. It functions by displaying data points in a high-dimensional space and locating the hyperplane that efficiently divides classes. Tools like SVMlight with C, LibSVM with Python, MATLAB, SAS, Kernlab, Scikit-Learn, and Weka can all be used to build SVM. [15][16][17].
6. *Random Forest (RF)*: is an ensemble classifier that mimics a forest composed of several trees by using

several decision trees. Both classification and regression tasks can be completed with it. To create a model that can predict outcomes and categorize unlabeled input, reinforcement learning uses a labeled dataset. It works very well with big datasets. R, Scikit-Learn, and Python are some of the tools used in RF implementation [16][17].

III. RESEARCH METHODOLOGY

A. Research Procedure

The search procedure is a crucial component of research, ensuring that the study is accurate and grounded through comprehension of the state of the art in the field of public health. This search focuses on selecting relevant published articles from 2018 to 2023 that pertain to infectious disease prediction, with the aim of forecasting future outbreaks. To ensure thorough and adequate coverage of studies aligned with our research objectives, the search was conducted using databases such as PubMed, Web of Science, Scopus, and Google Scholar. The upcoming search keywords were used to find papers that could be related: (Forecast* OR Predict) AND (Machine learning) AND (infectious diseases OR communicable diseases) AND (Risk factors).

B. Data Analysis

The selected studies were categorized based on factors such as the country of origin, type of disease, disease classification, data source, sample size, and the techniques employed.

IV. LITERATURE REVIEW

This section summarizes previous research on machine learning-based public health forecasts. It provides a thorough examination of the approaches, formulas, evaluation criteria, and modeling instruments used in forecasting of infectious diseases.

Yadav et al. performed a study to predict malaria in Senegal using four distinct machine learning algorithms. They used datasets 1 and 2, which had 21,083 and 5,809 records, respectively. Artificial neural networks (ANN), Support Vector Machines (SVM) with Gaussian kernels, and Random Forests (RF) were used in the study. The outcomes showed that ANN produced the best accuracy, with at least 92%, 85%, and 89% of precision, recall, and F1-scores, respectively, in both datasets [20].

In a study, Mushtaq et al. concentrated on identifying key input variables for machine learning-based waterborne disease prediction. To forecast cases of waterborne diseases that are positive, they used a variety of machine learning methods, such as KNN, DT, Random Forest (RF), (SVM), and Logistic Regression LR. The authors received patient data on typhoid and malaria from Ayub Medical Hospital in Pakistan between 2017 and 2020. A total of 22,916 instances

of malaria and 68,624 cases of typhoid were included in the data. According to the experiment's results, RF outperformed other machine-learning models in its ability to forecast cases of typhoid disease (77%), and malaria (60%). Furthermore, RF outperformed other ML models by a large margin in the malaria dataset [21].

Ousseynou et al. evaluated the effectiveness of the most used machine learning models for predicting Malaria occurrence. They compared several ML algorithms, including NB, LR, DT, SVM, RF, and ANN. Using a real-world dataset containing 21,083 instances of patients residing in Senegal, the study found that Naive Bayes (NB) outperformed the other techniques, delivering the best precision, recall, and AUC scores [22].

P. Mohapatra et al. proposed a model to forecast malaria risk in India using different ML algorithms. The study utilized data from the Directorate of Public Health Services, Government of Odisha, which included various climate datasets and temporal scales. The research was conducted using the WEKA platform, employing Multilayer Perceptron (MLP) and J48 classifier techniques. The findings revealed that the J48 classifier outperformed the MLP, achieving a performance accuracy of 71%. [23].

Salim et al. assessed a highly efficient machine-learning model for forecasting Dengue outbreaks across five different districts in Malaysia. The study utilized CART, ANN, SVM, and NB models to predict dengue outbreaks based on climate variables. Data was gathered from private clinics, public clinics, and hospitals, comprising 1,300 instances over five years. The findings indicated that the SVM model delivered the highest predictive performance among the various ML techniques, achieving an accuracy of 70%. This study highlights how SVM with a linear kernel can accurately forecast dengue outbreaks without overfitting. [24].

Samrat et al. employed an ML model to identify the factors contributing to dengue epidemics in a specific area of Bangladesh. The study used two machine learning techniques: Multiple Linear Regression (MLR) and Support Vector Regression (SVR). A dataset named DengueBD was created using daily press releases from the Directorate General of Health Services (DGHS) in Bangladesh for this study. The findings indicate that the MLR model achieved 67% accuracy, while the SVR model reached 75% accuracy [25].

McGough et al. suggested a novel method based on machine learning for forecasting dengue outbreaks in Brazil. The study used data on cases of dengue fever per year, collected from the Brazilian Ministry of Health between 2001 and 2017, as well as daily temperature and precipitation data from GMAO-NASA for the years 2000-2016. The researchers employed SVM model and an ensemble learning model to predict whether a year would be epidemic or non-epidemic for dengue fever. This method focused on time-series feature extraction, and the model achieved an accuracy of 75% [26].

Dourjoy et al. identified key signs of dengue fever and developed an early projection system for dengue in Bangladesh. They used SVM and RF classifier algorithms to forecast the occurrence of dengue fever, analyzing around 1,047 data points gathered through surveys and online sources. Unlike other studies, their approach focused on symptoms as the main parameter for predicting future dengue cases. The results indicated that SVM achieved a slightly higher accuracy of 69%, compared to 68% for Random Forest [27].

Gupta et al. created an early prediction model for dengue disease diagnosis by applying a number of machine learning techniques, such as support vector classifier (SVC), DT, RF, Gaussian Naive Bayes, and KNN classifier. An analysis of 1,872 cases from two cities was conducted using data from the DengAI competition. With the best accuracy and mean score of 8.72, Random Forest emerged as the most successful strategy among the evaluated ones, according to the data [28].

Zia Farooq et al. created a framework to provide rapid and accurate notifications for potential West Nile virus outbreaks in Europe. The study leveraged the XGBoost machine learning algorithm and SHAP for ranking feature importance. They applied the framework to four distinct datasets from 2010 to 2019. The findings revealed that the model accomplished AUC scores of 0.97 and 0.93 on two of the datasets, highlighting its strong performance [29].

Ajith et al. carried out research to forecast the trend of West Nile Virus (WNV) outbreaks in India using machine learning models. They used a dataset from Kaggle as the primary data source for their analysis. The study employed multiple techniques, including RF Classifier, NB Classifier, and Adaptive Boost. The results revealed that the RF algorithm delivered the highest accuracy in predicting WNV presence, surpassing the other methods utilized [30].

Kim et al. applied SVM to estimate the prevalence of influenza by analyzing Internet articles collected from the Centre for Health Protection (CHP) in Hong Kong. The dataset, which included 7,791 news articles from sources like Twitter, spanned from 2004 to 2018. The research found that SVM obtained an accuracy of 86.7% in predicting influenza cases. They also proposed using a weighting index to determine if a peak is local or global and how big it is [31].

Chiu et al. utilized ML methods to examine disease prevention in the setting of Emerging Infectious Diseases (EID). An analysis of 83,227 Taiwanese hospital admissions for influenza-like illness (ILI) was conducted. Deep Neural Network (DNN), logistic regression-based, and DT models are the three categories of prediction models that were created. The outcomes showed that DT models and DNN models had comparable performance in predicting the severity of ILI [32].

Gupta proposed a machine learning model to forecast Zika disease and predict future outbreak locations in rural and suburban areas. Feature selection was performed using RF and XGBoost algorithms. The model was built using

ADABOost, XGBoost, SVM, Multi-Layer Perceptron, and LR. The findings demonstrated that XGBoost had the highest accuracy, achieving 95% [33].

Zheng et al. predicted the probability of snail infestations with *Schistosoma* using a random forest machine learning technique. Data was collected through a snail survey conducted at 2,369 sites. The results showed that the model achieved an AUC of 0.889 and a Kappa value of 0.618 [34].

Ali et al. developed a model for forecasting Schistosomiasis by identifying key features using machine learning techniques in China. The dataset comprised 4,316 instances collected from the Hubei Institute of Schistosomiasis Prevention and Control (HISPC). Four machine learning techniques were employed to build the model: Gradient Boosting, Light Gradient Boosting, Extreme Gradient Boosting, and CatBoost. The results demonstrated that the CatBoost model achieved the highest accuracy at 87.1%, outperforming the other models [35].

Jiang et al. utilized machine learning techniques to create a model that predicted variables influencing patients with schistosomiasis one year after discharge. The Hunan Institute for Schistosomiasis Control (HISC) in China provided the data for the study, which comprised 9,541 patients. LR, DT, RF, ANN, and XGBoost are the five machine learning algorithms the prediction model was constructed using. XGBoost yielded the best results, as evidenced by an AUC value of 0.846 (95% confidence interval: 0.821 - 0.871) when the model's performance was evaluated using AUC [36].

Soni et al. developed a system for predicting Ebola disease using various machine learning techniques, including DT, Bagging classifiers, KNN, SVM, Stochastic Gradient Descent classifiers, LR, RF, Gradient Boosting classifiers, Ridge Classifier, and Hybrid Neural Networks. They utilized a combination of classification methods and hybrid models, particularly focusing on Stochastic Gradient Descent, Random Forest, and KNN techniques. The dataset, obtained from the WHO, revealed that while Random Forest achieved the highest accuracy among the individual techniques, the hybrid model combining Random Forest and KNN outperformed it, achieving an accuracy of 96% [37].

In conclusion, Table 1 provides a comparative analysis of recent studies that have applied machine learning techniques for infectious disease prediction.

Table 1: Comparative study of ML approaches for the prediction of infectious diseases adapted from [64].

Ref#	Year	Disease Type	Country	Aim	Data Source	Sample size	Techniques used	Outcomes
21	2023	Waterborne typhoid and malaria	Pakistan	Forecasting waterborne disease	Ayub Medical Hospital	Malaria= 22,916 Typhoid= 68,624	DT, RF, LR, KNN, SVM	RF achieved high accuracy with 60% for malaria and 77% for typhoid
62	2023	COVID-19	Korea	Detecting COVID-19 outbreak	KDCA	Thousands of cases	SVM, RF, XGB	RF achieved high accuracy with 95.80%
63	2023	covid19 , zika, malaria	India	Forecasting epidemic diseases	Tweepy library, Kaggle.	1000 tweets	DT, NB, SVM,RF,LR	LR achieved the highest accuracy of 94%
38	2022	COVID-19	US	Predicting COVID-19 pandemic	Corona Data Scraper	Thousands of cases	LR, SVR, KNN, MLP, RF, XGB	The prediction accuracy improved by 33.78%
40	2022	Dengue, Zika, influenza, COVID-19	Brazil	Predicting prevalence of Dengue, zika, COVID-19	NDIS	Thousands of cases	RF, TrAdaBoost, NN	RF performed better for zika, while NN were more effective for COVID-19
15	2022	COVID-19	India	Predicting COVID-19 pandemic	Kaggle	Millions of cases	NB, SVM, LR	Naïve Bayes is more accurate for predicting COVID-19
39	2022	Hepatitis C	Egypt	Forecasting Hepatitis C disease	Egyptian National Committee	4,962	MLP, NB, DT, SVM, RF, BN	BN algorithm achieved an accuracy of 68.9%
25	2022	Dengue	Bangladesh	Predicting Dengue outbreaks	Directorate General of Health Services	162272 +149485	SVR, MLR	SVM achieved high accuracy with 75%
29	2022	West Nile virus	Sweden	Developing West Nile virus outbreak model	European Center for Disease Prevention and Control	data set from 2010-2019	XG-Boost, SHAP, AXI	AUC with scores of 0.97 and 0.93
32	2022	Infectious diseases	Taiwan	Address the challenges of Infectious diseases	National Health Insurance	83,227	DT, LR, DNN	DT model with 85% sensitivity

Ref#	Year	Disease Type	Country	Aim	Data Source	Sample size	Techniques used	Outcomes
					Research Database			
35	2022	schistosome	China	predict Schistosomiasis features	Resdc, WorldCom	4316	GB, LG-Boost, XG—Boost, Cat-Boost	Cat-Boost model achieved the highest accuracy of 87.1%
41	2022	COVID-19	Brazil	Predicting COVID-19	Israelita Albert Einstein Hospital	5,644	RF, XGB, SVM, LR	SVM achieved the highest performance on the external validation data set with an accuracy of 91.18% while RF on the present study's dataset achieved the best performance with 85.3%
42	2022	COVID-19	-	Predicting trend of COVID-19	Open Data Resources from Mexico and Brazil	Not reported	LR- DT - Boosted RF	LR achieved the best results
48	2021	COVID-19	Brazil, US, India, France, UK, Russia	Predicting the trend of COVID-19 pandemic	Johns Hopkins University, WHO, CDC, COVID19 India	Thousands of cases	ARIMA models	The model achieved 85% accuracy for all the countries
44	2021	COVID-19	World, China, India, Italy	Predicting the trend of COVID-19 pandemic	WHO	Millions of cases	GPR , MTGP, LR, SVR, RF	MTGP model is better than the traditional forecasting model
31	2021	Multiple Infectious diseases	237 different Countries	Predicting incidence of Infectious diseases	Media articles, Medisys	Millions of cases	SVM, SSL, DNN	SSL performs very well, with an accuracy of over 0.7
45	2021	Hand, foot, and mouth disease	China	Predicting the trend of hand, foot, and mouth disease	DCCPHC, CMDSC, Tencent Location Big Data	Thousands of cases	RF, XGB	XGB model achieved the highest accuracy
46	2021	COVID-19	India	Predicting the COVID-19 pandemic	Kaggle	Thousands of cases	KNN, DT, SVM, RT	KNN is better with an accuracy of 98.34%
20	2021	Malaria	India	detecting malaria disease	two real-world datasets	D1=21083 D2=5809	RF, LR, SVM, ANN	For Dataset 1 the best model was ANN which achieved 99%,99%, and 99% for Precision, Recall, and F1-score respectively.

Ref#	Year	Disease Type	Country	Aim	Data Source	Sample size	Techniques used	Outcomes
22	2021	Malaria	Senegal	Assessing models for malaria prediction	real-world dataset about patients	21083	NB, LR, DT, SVM, RF, ANN	NB provides the best precision, recall, and AUC by 9% in the prediction.
24	2021	Dengue	Malaysia	predicting outbreaks of Dengue	private clinics, public clinics, and hospitals	1300	CART, ANN, SVM, BN	SVM achieved the highest performance with accuracy 70%
26	2021	Dengue	Brazil	Predicting Dengue outbreak	Brazilian Ministry of Health, GMAO-NASA	432	SVM, ensemble learning model	accuracy of 75%
27	2021	Dengue	Bangladesh	Identify symptoms of Dengue fever	surveying and using the Internet	1047	SVM, RF	SVM Accuracy: 69% and RF 68%
34	2021	schistosome	China	Forecast the likelihood of snail infection	survey of snail records	2369 sites	RF	AUC = 0.889
36	2021	schistosome	China	Predict Schistosomiasis factors	HISC	9541	LR, DT, RF, ANN, XGB	XGB achieved the highest results with AUC of 0.846, CI of 95%
37	2021	Ebola	India	Creating hybrid models of Ebola virus prediction	Harvard Dataverse	213	DT, KNN, SVM, GD, LR, RF, GB, KNN, hybrid NN	the hybrid model of RF classifier and KNN models achieved the highest accuracy with 96%
23	2021	Malaria	India	Predicting Future Malaria Cases	Directorate of Public Health Services	NA	MLP, J48	J48 achieved high accuracy with 71%
47	2020	Dengue	Perù, Puerto Rico,	Predicting weekly Dengue cases	Passive surveillance systems	Thousands of cases	RF, RF-UFA, ARIMA, LASSO	RF was more accurate for near-term predictions while the ARIMA model performed best for long-term predictions

Ref#	Year	Disease Type	Country	Aim	Data Source	Sample size	Techniques used	Outcomes
			Singapore					
43	2020	Hepatitis E	China	Predicting incidence of hepatitis E	Shandong Center for Disease Control and Prevention	Thousands of cases	ARIMA, SVM, LSTM	LSTM obtained the best performance in all three metrics of RSME, MAPE, MAE
49	2020	COVID-19	Iran	Predicting COVID-19 pandemic	Ministry of Health and Medical Education	Thousands of cases	SVM, ARIMA	SVM achieved an AUC value of 0.786
50	2020	COVID-19	World	Predicting trend of COVID-19	Kaggle	Thousands of cases	LR, BR, RF, KN	BR model gives the highest accuracy at 69.61%
51	2020	Swine fever		Predicting the African swine fever outbreak	African swine fever outbreak data and the WorldClim database meteorological data	Not reported	RF, ANN, SVM, AdaBoost, C4.5, and NB	RF achieved the highest accuracy with 98.29%
30	2020	West Nile Virus	India	predicting the trend of West Nile Virus	Kaggle	10,506	RF, NB, adaptive boost	RF achieved the highest accuracy
52	2020	schistosome	China	evaluating KAPs of schistosomiasis	standardized questionnaire	601	LR	the results approved that CI was 95%, P-value < 0.05
53	2020	COVID-19	Italy	Predicting trend of COVID-19	San Raffaele Hospital	279	DT, ET, KNN, RF, LR, NB, SVM	RF achieved a higher accuracy of 82 %
54	2019	Influenza-like illness (ILI)	Hong Kong	Predicting weekly incidence of ILI	(CHP)	Thousands of cases	SVM	SVM achieved an accuracy of 86.7%
55	2019	Dengue	Russian	Predicting incidence of Dengue	NOAA , Philippine Department of Health, CDC, ISID ,	Thousands of cases	RF, LSTM, XGB, SARIMA, LightGBM, LR,	The most effective predictions were given by a mathematical model based on a combination of spatial analysis techniques (MGWR) and neural networks based on the LSTM architecture

Ref#	Year	Disease Type	Country	Aim	Data Source	Sample size	Techniques used	Outcomes
					Landsat, Sentinel		KNN, CatBoost, Keras	
56	2019	Dengue	India	predict Dengue outbreak	Dengue disease dataset	75	LogitBoost, kNN, ANN, SVM, NB, DT	LogitBoost ensemble model achieved the best performance with an accuracy of 92%
57	2019	Influenza	Iran	predict influenza outbreak	Who	73483	SVM, ANN, RF	The accuracy of the SV was 89.2%
58	2019	zika	Not reported	Predicting zika epidemic	CDC, NASA, IMF	Not reported	ADABOost, XGB SVM, MLP, LR.	XGB has the highest accuracy with 95%
59	2018	Dengue, malaria, chickenpox, hand foot & mouth disease	Japan, Taiwan, Thailand, Singapore	Predicting the trend of Dengue, malaria, chickenpox, hand foot and mouth disease	National Institute of Infectious Diseases	Thousands of cases	LASSO	short-term predictions perform better than longer term predictions
60	2018	Hand, foot, and mouth disease	China	Predicting the trend of hand, foot, and mouth disease	Shenzhen Health Information Center, Weather Underground app	Thousands of cases	XGB	XGB works better by 16.7%

NN : Neuron Network
 NDIS : Notifiable Diseases Information System
 NB : Naïve Bayes
 XGB : XGboost
 GBT: Gradient-Boosted Trees
 SVM: Support vector machine
 LR : Logistic regression
 GPR : Gaussian process regression
 MTGP : Multi-Task Gaussian Process
 SSL: Semi-supervised Learning

DCCPHC: Data Center of China Public Health Science
 CMDSC : China Meteorological Data Service Centre
 BR : bagging regressor
 RF : Random forest
 NHANES : National Health and Nutrition Examination Survey
 CHP : Media Articles Centre for Health Protection
 NOAA : National Oceanic and Atmospheric Administration
 ISID : International Society for Infectious Diseases
 ET : Extremely Randomized Trees

V. OUTBREAK PREDICTION

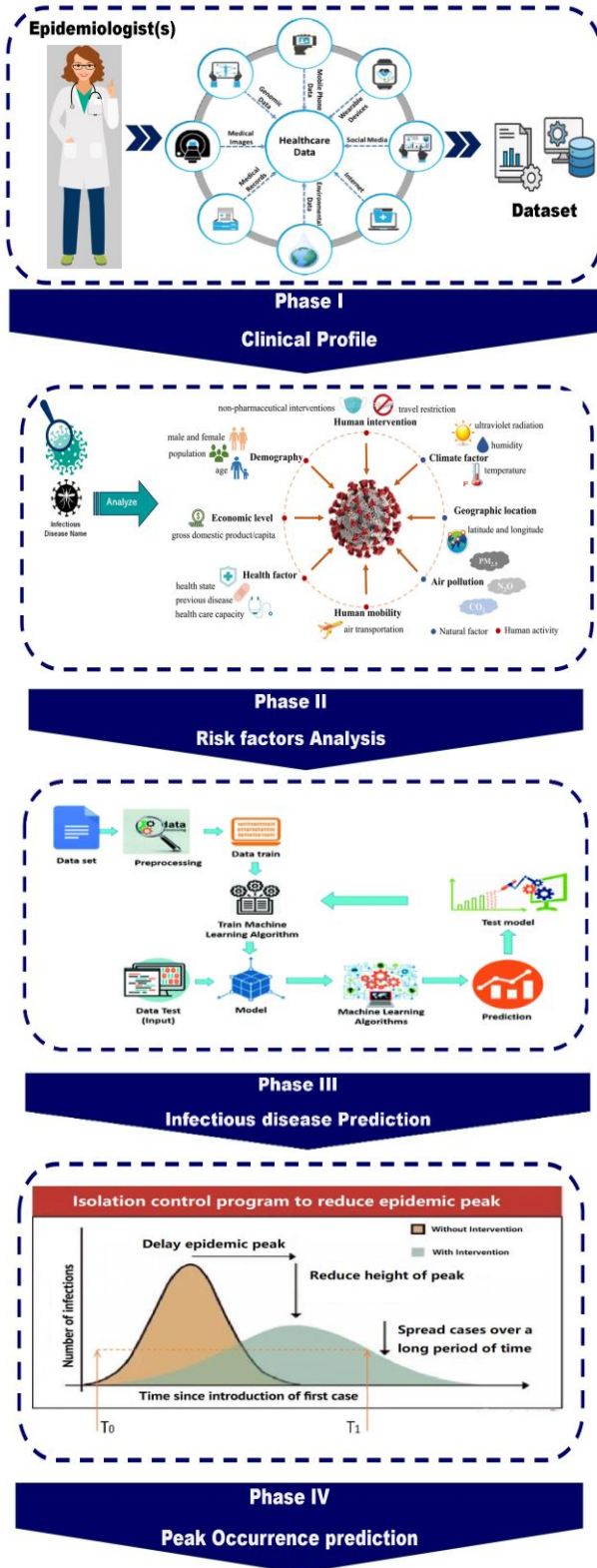


Fig 5: A proposed Intelligent novel framework for Early Epidemic Outbreak Detection

Machine learning models have recently gained significant attention for their effectiveness in predicting outbreaks. These models are adept at detecting hidden trends in data that conventional methods might overlook. Techniques such as NB, SVM, Neural Networks, and RF are commonly utilized to predict epidemic diseases. Despite the challenges in infectious disease prediction, machine learning models can overcome these by handling large volumes of data and identifying intricate relationships between input features. SVM and tree-based algorithms like RF are frequently used in infectious disease prediction due to their ease of use and effectiveness in time series analysis [1].

In a connected world, the efficiency of public health relies heavily on promptly identifying outbreaks and epidemics. Although traditional surveillance systems are important, they often face challenges related to data collection, processing speed, and adapting to new threats. [61].

To tackle these challenges, this research proposes a unique framework for the early detection of outbreaks and epidemics using machine learning techniques. As illustrated in Fig 5, this framework leverages the ability of machine learning to improve the early identification of epidemics. It is comprised of four distinct phases:

1. Clinical profile illustration:

In the first phase, the framework prioritizes the collection and presentation of clinical profiles from individuals in the affected population, assisted by experts like epidemiologists. This involves compiling detailed data on symptoms, medical history, and diagnostic test results. These clinical profiles are crucial for detecting patterns and anomalies that could suggest the onset of an outbreak.

2. Risk Factors Analysis

In this phase, a detailed analysis is performed to identify the risk factors that might contribute to the outbreak. This includes evaluating various elements such as demographic data, environmental conditions, travel histories, and exposure to specific sources or environments. To achieve a thorough understanding, more detailed insights into these factors are gathered. The analysis also incorporates epidemiological features and the disease's natural history to offer a complete assessment of the factors involved in the outbreak [19].

3. Infectious Disease Prediction using ML

ML algorithms are employed to forecast the risk of infectious disease spreading among the population during this crucial stage. The dynamics of disease propagation can be simulated by using a various of machine-learning methods to identify affected individuals. Pre-processing to prepare the raw data, data training—using training data to train the created model—and prediction-gathering from a second set of test data are some of the processes that make up this phase. The last stage is the model evaluation testing phase.

4. Peak Occurrence Prediction Using Time Series

One of the most important aspects of preparing a successful public health response is estimating the peak incidence of an outbreak or epidemic. At this point, the outbreak's anticipated peak is predicted with accuracy and timeliness using time series analytic techniques. Knowing when the peak will occur helps public health professionals and policymakers more effectively deploy resources, implement interventions, and coordinate responses.

VI. COMPARATIVE ANALYSIS STUDY

The algorithms employed annually are outlined in research published from 2018 to 2023, as depicted in Fig 6. Moreover, all of these studies address challenges related to disease prediction.

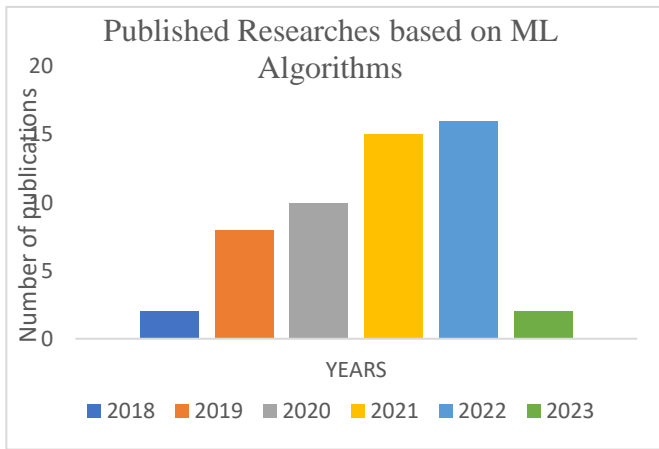


Fig 6: Published studies applying ML methods.

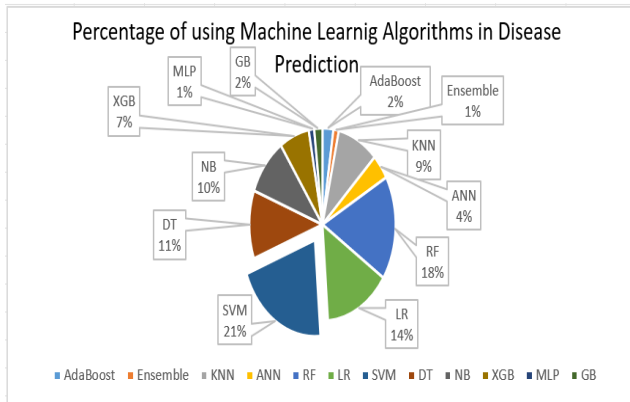


Fig 7: Percentage of diseases predicted by ML techniques.

The use of ML algorithms for disease prediction, as illustrated in Fig. 7, indicates that Support Vector Machine has become the top performer, achieving the greatest percentage at 21%. On the other hand, studies or publications that used ensemble and Multilayer Perceptron (MLP) approaches had the lowest percentage, which was 1%.

VII. DISCUSSION

The summary of the selected research highlights two main findings. Firstly, it identifies the challenge of using small datasets for predicting epidemics or disease outbreaks, emphasizing the need for larger, more comprehensive datasets to enhance the predictability and accuracy of predictive models. Secondly, it demonstrates that the performance of machine learning models can be enhanced by using a combination of techniques to extract a wide range of features, which helps capture diverse patterns and characteristics.

To address these challenges, the proposed model aims to utilize multiple datasets for diagnosing infectious diseases. By incorporating various data sources, the model seeks to provide valuable insights to decision-makers, enabling earlier detection and more timely responses to potential outbreaks. The importance of data quality and diversity is also crucial, as large-scale, high-quality datasets and the integration of different sources—such as clinical records, environmental data, and social media—are essential for developing accurate and robust predictive models. This comprehensive approach will enhance understanding of disease patterns and support effective public health interventions and timely actions.

VIII. CONCLUSION

Determining the end of infectious disease outbreaks is crucial for effective response, but current models in the literature often struggle with generalization and robustness. There is a need to improve these models to provide more reliable and practical solutions. This research offers a comparative analysis of various machine learning models for predicting disease outbreaks, serving as an initial benchmark that highlights machine learning's suitability for further study. It also recommends that significant progress in outbreak forecasting can be made by involving machine learning techniques in existing approaches.

IX. LIMITATIONS AND FUTURE WORK

Predicting or measuring epidemics presents significant challenges due to the absence of fixed standards, driven by the unpredictable nature of epidemics and the diverse factors influencing them, particularly in infectious diseases. A review of current studies highlights several critical issues: the scarcity of large datasets, which are crucial for the effectiveness of machine learning algorithms; the presence of noisy datasets filled with irrelevant information, which decreases model accuracy and heightens the risk of overfitting; and the challenge of feature selection, which is crucial for model performance as it involves identifying the most relevant features from a broader set. While open-source libraries address some of these challenges, integrating the Analytic Hierarchy Process (AHP) technique could prove valuable for future research.

Moreover, many studies are focused on specific diseases, which limits the applicability of their findings to broader populations or other health conditions. Creating a more generalized model capable of predicting a wider range of diseases could be a promising area for future investigation. Additionally, many studies rely solely on clinical profiles, overlooking environmental factors. Including these factors in future research is crucial for improving prediction accuracy. Due to the significant ambiguity and absence of necessary historical data, common models often struggle with inadequate long-term prediction accuracy.

REFERENCES

- [1] Keshavamurthy, R., Dixon, S., Pazdernik, K. T., & Charles, L. E. (2022). Predicting infectious disease for biopreparedness and response: A systematic review of machine learning and deep learning approaches. *One Health*, 100439.
- [2] Ajagbe, S. A., & Adigun, M. O. (2023). Deep learning techniques for detection and prediction of pandemic diseases: a systematic literature review. *Multimedia Tools and Applications*, 1-35.
- [3] Santangelo, O. E., Gentile, V., Pizzo, S., Giordano, D., & Cedrone, F. (2023). Machine Learning and Prediction of Infectious Diseases: A Systematic Review. *Machine Learning and Knowledge Extraction*, 5(1), 175-198.
- [4] Battineni, G., Sagaro, G. G., Chinatalapudi, N., & Amenta, F. (2020). Applications of machine learning predictive models in the chronic disease diagnosis. *Journal of personalized medicine*, 10(2), 21.
- [5] Dicker, R. C., Coronado, F., Koo, D., & Parrish, R. G. (2006). *Principles of epidemiology in public health practice; an introduction to applied epidemiology and biostatistics*.
- [6] www.cdc.gov/scienceambassador/nerdacademy.
- [7] World Health Organization. (2009). Global health risks: mortality and burden of disease attributable to selected major risks. World Health Organization.
- [8] Wood, S., Harrison, S. E., Judd, N., Bellis, M. A., Hughes, K., & Jones, A. (2021). The impact of behavioral risk factors on communicable diseases: a systematic review of reviews. *BMC Public Health*, 21(1), 1-16.
- [9] Pijls, B. G., Jolani, S., Atherley, A., Derckx, R. T., Dijkstra, J. I., Franssen, G. H., ... & Zeegers, M. P. (2021). Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. *BMJ open*, 11(1), e044640.
- [10] Rojas-Rueda, D., Morales-Zamora, E., Alsufyani, W. A., Herbst, C. H., AlBalawi, S. M., Alsukait, R., & Alomran, M. (2021). Environmental risk factors and health: an umbrella review of meta-analyses. *International journal of environmental research and public health*, 18(2), 704.
- [11] <https://www.cdc.gov/training/quicklearns/createepi/>
- [12] Philip S. Brachman, *Control of Communicable Diseases Manual*, 17th Edition, American Journal of Epidemiology, Volume 154, Issue 8, 15 October 2001, Pages 783–784.
- [13] Alexander, L., Koshiol, J., & MacDonald, P. D. M. (2003). An overview of outbreak investigations. *FOCUS on Field Epidemiology*, 1(1).
- [14] Linthicum, K. P., Schafer, K. M., & Ribeiro, J. D. (2019). Machine learning in suicide science: Applications and ethics. *Behavioral sciences & the law*, 37(3), 214-222.
- [15] Tiwari, D., Bhati, B. S., Al-Turjman, F., & Nagpal, B. (2022). Pandemic coronavirus disease (Covid-19): World effects analysis and prediction using machine-learning techniques. *Expert Systems*, 39(3), e12714.
- [16] Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making*, 19(1), 1-16.
- [17] Badawy, M., Ramadan, N., & Hefny, H. A. (2023). Healthcare predictive analytics using machine learning and deep learning techniques: a survey. *Journal of Electrical Systems and Information Technology*, 10(1), 40.
- [18] Palaniappan, S., V. R., & David, B. (2022). Prediction of epidemic disease dynamics on the infection risk using machine learning algorithms. *SN computer science*, 3(1), 47
- [19] Charnley, G. E., Kelman, I., Gaythorpe, K. A., & Murray, K. A. (2021). Traits and risk factors of post-disaster infectious disease outbreaks: a systematic review. *Scientific reports*, 11(1), 5616.
- [20] Yadav, S. S., Kadam, V. J., Jadhav, S. M., Jagtap, S., & Pathak, P. R. (2021, March). Machine learning based malaria prediction using clinical findings. In 2021 International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 216-222).
- [21] Hussain, M., Cifci, M. A., Sehar, T., Nabi, S., Cheikhrouhou, O., Maqsood, H., ... & Mohammad, F. (2023). Machine learning based efficient prediction of positive cases of waterborne diseases. *BMC Medical Informatics and Decision Making*, 23(1), 1-16.
- [22] Mbaye, O., Ba, M. L., & Sy, A. (2021). On the Efficiency of Machine Learning Models in Malaria Prediction. In *Public Health and Informatics* (pp. 437-441). IOS Press.
- [23] Mohapatra, P., Tripathi, N. K., Pal, I., & Shrestha, S. (2022). Determining suitable machine learning classifier technique for prediction of malaria incidents attributed to climate of Odisha. *International Journal of Environmental Health Research*, 32(8), 1716-1732.
- [24] Salim, N. A. M., Wah, Y. B., Reeves, C., Smith, M., Yaacob, W. F. W., Mudin, R. N., ... & Haque, U. (2021). Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Scientific reports*, 11(1), 939.
- [25] Dey, S. K., Rahman, M. M., Howlader, A., Siddiqi, U. R., Uddin, K. M. M., Borhan, R., & Rahman, E. U. (2022). Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh: A machine learning approach. *PLoS One*, 17(7), e0270933.
- [26] McGough, S. F., Clemente, L., Kutz, J. N., & Santillana, M. (2021). A dynamic, ensemble learning approach to forecast dengue fever epidemic years in Brazil using weather and population susceptibility cycles. *Journal of the Royal Society Interface*, 18(179), 20201006.
- [27] Dourjoy, S. M. K., Rafi, A. M. G. R., Tumpa, Z. N., & Saifuzzaman, M. (2021). A comparative study on prediction of dengue fever using machine learning algorithm. In *Advances in Distributed Computing and Machine Learning: Proceedings of ICADCML 2020* (pp. 501-510). Springer Singapore.
- [28] Gupta, G., Khan, S., Guleria, V., Almjally, A., Alabdullah, B. I., Siddiqui, T., ... & Al-Subaie, M. (2023). DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment Analysis and Machine Learning Algorithms. *Diagnostics*, 13(6), 1093.
- [29] Farooq, Z., Rocklöv, J., Wallin, J., Abiri, N., Sewe, M. O., Sjödin, H., & Semenza, J. C. (2022). Artificial intelligence to predict West Nile virus outbreaks with eco-climatic drivers. *The Lancet Regional Health—Europe*, 17.
- [30] Ajith, A., Manoj, K., Kiran, H., Pillai, P. J., & Nair, J. J. (2020, July). A study on prediction and spreading of epidemic diseases. In 2020 International Conference on Communication and Signal Processing (ICCSP) (pp. 1265-1268). IEEE.
- [31] Kim, J., & Ahn, I. (2021). Infectious disease outbreak prediction using media articles with machine learning models. *Scientific reports*, 11(1), 4413.
- [32] Chiu, H. Y. R., Hwang, C. K., Chen, S. Y., Shih, F. Y., Han, H. C., King, C. C., ... & Oyang, Y. J. (2022). Machine learning for emerging infectious disease field responses. *Scientific Reports*, 12(1), 328.
- [33] Abhijith Gupta, "Prediction of Epidemic Disease Dynamics using Machine Learning", 2019.

- [34] Zheng, J. X., Xia, S., Lv, S., Zhang, Y., Bergquist, R., & Zhou, X. N. (2021). Infestation risk of the intermediate snail host of *Schistosoma japonicum* in the Yangtze River Basin: improved results by spatial reassessment and a random forest approach. *Infectious diseases of poverty*, 10(03), 34-46.
- [35] Ali, Z., Hayat, M. F., Shaukat, K., Alam, T. M., Hameed, I. A., Luo, S., ... & Ksibi, A. (2022). A proposed framework for early prediction of Schistosomiasis. *Diagnostics*, 12(12), 3138.
- [36] Jiang, H., Deng, W., Zhou, J., Ren, G., Cai, X., Li, S., ... & Zhou, Y. (2021). Machine learning algorithms to predict the 1 year unfavourable prognosis for advanced schistosomiasis. *International Journal for Parasitology*, 51(11), 959-965.
- [37] Soni, U., & Gupta, N. (2021, August). An Artificial Intelligence Approach for Forecasting Ebola Disease. In *Journal of Physics: Conference Series* (Vol. 1950, No. 1, p. 012038). IOP Publishing.
- [38] Katragadda, S., Bhupatiraju, R. T., Raghavan, V., Ashkar, Z., & Gottumukkala, R. (2022). Examining the COVID-19 case growth rate due to visitor vs. local mobility in the United States using machine learning. *Scientific Reports*, 12(1), 12337.
- [39] Abd El-Salam, S. M., Ezz, M. M., Hashem, S., Elakel, W., Salama, R., ElMakhzangy, H., & ElHefnawi, M. (2019). Performance of machine learning approaches on prediction of esophageal varices for Egyptian chronic hepatitis C patients. *Informatics in Medicine Unlocked*, 17, 100267.
- [40] Roster, K., Connaughton, C., & Rodrigues, F. A. (2022). Forecasting new diseases in low-data settings using transfer learning. *Chaos, Solitons & Fractals*, 161, 112306.
- [41] Çubukçu, H. C., Topcu, D. İ., Bayraktar, N., Gülşen, M., Sari, N., & Arslan, A. H. (2022). Detection of COVID-19 by machine learning using routine laboratory tests. *American journal of clinical pathology*, 157(5), 758-766.
- [42] Iwendi, C., Huescas, C. G. Y., Chakraborty, C., & Mohan, S. (2022). COVID-19 health analysis and prediction using machine learning algorithms for Mexico and Brazil patients. *Journal of Experimental & Theoretical Artificial Intelligence*, 1-21.
- [43] Guo, Y., Feng, Y., Qu, F., Zhang, L., Yan, B., & Lv, J. (2020). Prediction of hepatitis E using machine learning models. *Plos one*, 15(9), e0237750.
- [44] Ketu, S., & Mishra, P. K. (2021). Enhanced Gaussian process regression-based forecasting model for COVID-19 outbreak and significance of IoT for its detection. *Applied Intelligence*, 51, 1492-1512.
- [45] Abdel-Fattah, M. A., Othman, N. A., & Goher, N. (2022). Predicting chronic kidney disease using hybrid machine learning based on apache spark. *Computational Intelligence and Neuroscience*, 2022.
- [46] Rohini, M., Naveena, K. R., Jothipriya, G., Kameshwaran, S., & Jagadeeswari, M. (2021, March). A comparative approach to predict corona virus using machine learning. In *2021 international conference on artificial intelligence and smart systems (ICAIS)* (pp. 331-337).
- [47] Benedum, C. M., Shea, K. M., Jenkins, H. E., Kim, L. Y., & Markuzon, N. (2020). Weekly dengue forecasts in Iquitos, Peru; San Juan, Puerto Rico; and Singapore. *PLoS neglected tropical diseases*, 14(10), e0008710.
- [48] Dash, S., Chakraborty, C., Giri, S. K., Pani, S. K., & Frnda, J. (2021). BIFM: Big-data driven intelligent forecasting model for COVID-19. *IEEE Access*, 9, 97505-97517.
- [49] Pourghasemi, H. R., Pouyan, S., Farajzadeh, Z., Sadhasivam, N., Heidari, B., Babaei, S., & Tiefenbacher, J. P. (2020). Assessment of the outbreak risk, mapping and infection behavior of COVID-19: Application of the autoregressive integrated-moving average (ARIMA) and polynomial models. *Plos one*, 15(7), e0236238.
- [50] Mahima, Y., & Ginige, T. N. D. S. (2020, November). COVID-19 Spread prediction Based on Food Categories using Data Science. In *2020 IEEE International Conference for Innovation in Technology (INOCON)* (pp. 1-7). IEEE.
- [51] Liang, R., Lu, Y., Qu, X., Su, Q., Li, C., Xia, S., ... & Niu, B. (2020). Prediction for global African swine fever outbreaks based on a combination of random forest algorithms and meteorological data. *Transboundary and emerging diseases*, 67(2), 935-946.
- [52] Guan, Z., Dai, S. M., Zhou, J., Ren, X. B., Qin, Z. Q., Li, Y. L., ... & Xu, J. (2020). Assessment of knowledge, attitude and practices and the analysis of risk factors regarding schistosomiasis among fishermen and boatmen in the Dongting Lake Basin, the People's Republic of China. *Parasites & Vectors*, 13(1), 1-9.
- [53] Brinati, D., Campagner, A., Ferrari, D., Locatelli, M., Banfi, G., & Cabitza, F. (2020). Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. *Journal of medical systems*, 44, 1-12.
- [54] Kim, J., & Ahn, I. (2019). Weekly ILI patient ratio change prediction using news articles with support vector machine. *BMC bioinformatics*, 20, 1-16.
- [55] Kolesnikov, A. A., Kikin, P. M., & Portnov, A. M. (2019). Diseases spread prediction in tropical areas by machine learning methods ensembling and spatial analysis techniques. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 221-226.
- [56] Iqbal, N., & Islam, M. (2019). Machine learning for dengue outbreak prediction: A performance evaluation of different prominent classifiers. *Informatica*, 43(3).
- [57] Tapak, L., Hamidi, O., Fathian, M., & Karami, M. (2019). Comparative evaluation of time series models for predicting influenza outbreaks: Application of influenza-like illness data from sentinel sites of healthcare centers in Iran. *BMC research notes*, 12, 1-6.
- [58] <https://medium.com/intel-student-ambassadors/prediction-of-epidemic-disease-dynamics-using-machine-learning-22cf3b7129f3>.
- [59] Chen, Y., Chu, C. W., Chen, M. I., & Cook, A. R. (2018). The utility of LASSO-based models for real time forecasts of endemic infectious diseases: A cross country comparison. *Journal of biomedical informatics*, 81, 16-30.
- [60] Zhong, R., Wu, Y., Cai, Y., Wang, R., Zheng, J., Lin, D., ... & Li, Y. (2018). Forecasting hand, foot, and mouth disease in Shenzhen based on daily level clinical data and multiple environmental factors. *Bioscience trends*, 12(5), 450-455.
- [61] Cooper, D. J., Lear, S., Sithole, N., Shaw, A., Stark, H., Ferris, M., ... & Baker, S. (2022). Demographic, behavioural and occupational risk factors associated with SARS-CoV-2 infection in UK healthcare workers: a retrospective observational study. *BMJ open*, 12(11), e063159.
- [62] Cho, G., Park, J. R., Choi, Y., Ahn, H., & Lee, H. (2023). Detection of COVID-19 epidemic outbreak using machine learning. *Frontiers in Public Health*, 11, 1252357.
- [63] Potnis, N., & Tiple, B. (2023). Machine Learning Techniques for Disease Prediction. In *ITM Web of Conferences* (Vol. 57, p. 01004). EDP Sciences.
- [64] R. Abdallah, S. A. Abdel Gaber and H. A. Sayed, "Disease Outbreak /Epidemic in Public Health Sector," 2024 6th International Conference on Computing and Informatics (ICCI), New Cairo - Cairo, Egypt, 2024, pp. 203-216, doi: 10.1109/ICCI61671.2024.10485007.