

RESEARCH ARTICLE



Enhancing Diarrheal Disease Diagnosis and Detection with a Hybrid Deep Learning Model

 OPEN ACCESS

Received: 18-01-2024

Accepted: 12-02-2024

Published: 28-02-2024

Nilamba Vala¹, Shruti Yagnik², Yogesh Kumar^{3*}¹ Research Scholar, INDUS University, Ahmedabad, Gujarat, India² Associate Professor, INDUS University, Ahmedabad, Gujarat, India³ Department of CSE, School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India

Citation: Vala N, Yagnik S, Kumar Y (2024) Enhancing Diarrheal Disease Diagnosis and Detection with a Hybrid Deep Learning Model. Indian Journal of Science and Technology 17(11): 967-978. <https://doi.org/10.17485/IJST/v17i11.132>

* Corresponding author.

yogesh.arora10744@gmail.com**Funding:** None**Competing Interests:** None

Copyright: © 2024 Vala et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([ISee](https://www.isee.org/))

ISSN

Print: 0974-6846

Electronic: 0974-5645

Abstract

Background: Diarrheal disease is a major source of mortality as well as morbidity in children under five in underdeveloped nations. Traditional diagnostic methods for diarrheal disease are time-consuming and often lack accuracy which leads to delays in treatment and the potential for misdiagnosis.

Objectives: The objective of this review paper is to provide a comprehensive analysis of the existing body of research on the detection and diagnosis of diarrheal disease using artificial intelligence based techniques such as machine learning and deep learning. **Method:** A review has been conducted where the papers from the year 2015-2023 have been screened using PRISMA criteria. **Findings:** This review underlines how machine and deep learning can diagnose diarrheal disorders and work on their limitations that still persist. It synthesizes information from peer-reviewed papers and research studies to show how these techniques can improve diagnosis, which is essential for treating and preventing diarrheal diseases. While reviewing the existing techniques, it has been found that Random Forest achieved 97.48% accuracy, Logistic Regression and Support Vector Machine obtained 100% recall, while Naive Bayes produced the maximum precision of 96.55% with a reduced error rate of 2.52%. The review paper also stresses the necessity of standardizing feature representation, model architecture, and assessment criteria to optimize diarrheal illness detection models. **Novelty:** While most of the researchers had used only machine learning techniques, this review emphasizes the importance of using deep learning techniques as well as hyper parameter optimizers for generating the optimal results. Apart from this, the novelty of study lies in its hybridization of advanced Convolutional Neural Networks (CNNs) and their diverse combinations to represent a significant advancement in the detection and diagnosis of diarrheal diseases.

Keywords: Healthcare; Diarrheal disease; Pathogens; Artificial intelligence; Machine Learning; Transfer Learning; Deep Learning

1 Introduction

Diarrheal disease ranks as the second most common cause of mortality among children aged five and below that results in the death of approximately 525,000 children annually. Globally, there are 780 million people who do not have access to better quality drinking water, and 2.5 billion people who do not have access to improved sanitation facilities⁽¹⁾. Diarrhoea can persist for multiple days, depleting the body of vital water and minerals essential for survival. Previously, the primary factors contributing to fatalities from diarrhoea were severe dehydration and excessive fluid loss. Currently, it is probable that septic bacterial infections are becoming a more significant factor in the overall number of deaths related to diarrhoea⁽²⁾. The individuals who are most susceptible to life-threatening diarrhoea are children who suffer from malnutrition or have compromised immune systems, as well as individuals who are HIV-positive. Diarrhoea commonly arises as a manifestation of an illness in the gastrointestinal tract, resulting from a range of bacterial, viral, and parasite microorganisms. The transmission of infection occurs by contaminated food or drinking-water, or through person-to-person contact due to inadequate hygiene practices⁽³⁾. Diarrhoea is characterized by the occurrence of three or more loose or watery stools per day, or a higher frequency of bowel movements than what is considered typical for the person. The regular passage of well-formed stools does not constitute diarrhoea, nor does the passage of loose, "pasty" stools in breastfed infants. Figure 1 presents the types of diarrhoea⁽⁴⁾.

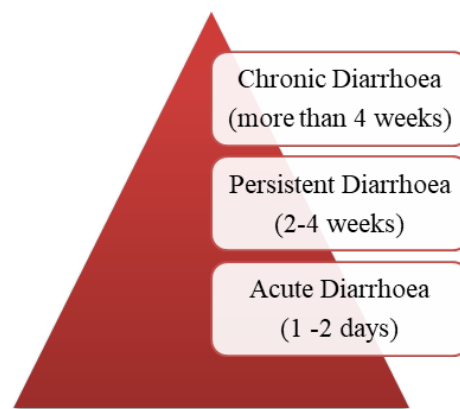


Fig 1. Types of diarrhoea

The conventional methods for identifying and managing diarrhoea encompass a blend of clinical evaluation, stool analysis, polymerase chain reaction testing, rehydration, dietary adjustments, antibiotics and anti-parasitic drugs, alleviation of symptoms, and ongoing monitoring and follow-up⁽⁵⁾. Conventional methods for identifying and managing diarrhoea include constraints, such as the laborious process of microbiological analysis, possible delays in obtaining outcomes, and the dependence on subjective clinical evaluations. Furthermore, these methods may not consistently identify the exact causal factor, resulting in the use of broad-spectrum medications or delays in giving specific therapies⁽⁶⁾.

Artificial Intelligence (AI) techniques play a crucial role in overcoming these limitations by expediting the diagnostic process and enhancing accuracy. Machine learning algorithms can analyze vast datasets of clinical information, laboratory results, and medical literature for the identification of patterns and correlations that may not be immediately apparent to human observers. AI can facilitate rapid analysis of diagnostic tests, providing quicker and more accurate identification of pathogens responsible for diarrhoea⁽⁷⁾. Moreover, AI-driven decision support systems can assist healthcare professionals in devising personalized treatment plans based on the specific characteristics of the infection, optimizing therapeutic interventions and potentially reducing the duration and severity of diarrhoea episodes⁽⁸⁾. An artificial intelligence system can accurately identify cases of diarrhoea with a precision of up to 98% by analyzing the auditory signals produced by toilets. Maia Gatlin and her Georgia Tech colleagues collected 350 toilet sounds from YouTube as well as Soundsnap. These recordings included ordinary faeces, diarrhoea, urine, and farts. Then, they used 70% of the recordings to train an AI system to distinguish the four faeces groups by sound. They tested the AI's performance with the remaining 20% of the recordings after establishing its consistency with 10% of the data⁽⁹⁾. A group of scientists from Duke University used machine learning algorithms to discover complex patterns in the bacterial communities that live in the human gut. These patterns are so intricate that they cannot be detected by humans alone. These patterns may serve as indicators of which individuals, out of the roughly one billion people worldwide who are at risk of contracting cholera, will actually become ill with this diarrheal disease⁽¹⁰⁾.

Hence based on such inventions and ideas, the aim of the article is to perform and sum up the work of the researchers who have given their contribution for the detection and diagnosis of diarrheal diseases using machine and deep learning techniques.

1.1 Research contribution

The main contribution of the study is mentioned as following:

1. Extensive background information on diarrheal diseases which include their epidemiology, traditional treatment techniques, limitations, and how AI can address these limitations has been provided.
2. A systematic review of relevant research papers has been conducted using the PRISMA criteria, and several research questions have been formulated. Apart from this, the proposed methodology has been also presented that involves hybridized deep learning techniques for detecting and classifying diarrheal disease.
3. A comprehensive survey as well as analysis of researchers' contributions to the use of machine and deep learning techniques for detecting and classifying diarrheal pathogens has been done which also include the challenges they face.

1.2 Research questions

In addition, the study also covered few research questions that are discussed in Section 4:

RQ 1: How can deep learning algorithms be applied as well as optimized for the accurate and rapid detection of diarrheal pathogens, considering the diverse nature of microbial species responsible for diarrhoea?

RQ 2: What is the cost-effectiveness of implementing AI-based diarrheal disease detection systems in healthcare settings?

RQ 3: What are the ways in which collaborating with public health institutions might improve the incorporation of AI-driven technologies into surveillance and monitoring systems for diarrheal diseases?

1.3 Organization of the manuscript

After a thorough introduction of the diarrheal disease, its traditional as well as AI based techniques to diagnose in Section 1, Section 2 presents the methodology used to conduct this study which mainly includes the PRISMA part of the article. Section 3 informs us regarding the work done by the researchers in the same realm, followed by Section 4 and Section 5 where the article is discussed and concluded respectively.

2 Methodology

This review has adhered to the PRISMA (Preferred reporting items for systematic Reviews and Meta-Analyses) standards, as shown in Figure 2, to minimize the bias process and provide a transparent as well as systematic approach to conducting reviews.

A comprehensive search was conducted manually between 2015 and 2023 in five distinct publication databases, namely Google Scholar (<https://scholar.google.co.in>), ScienceDirect (<https://www.sciencedirect.com>), PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>), Springer (<https://www.springer.com/in>), and Scopus (<https://www.scopus.com>) with the aim of identifying the relevant papers in order to ensure the thoroughness of the review. The publication has been queried using the keywords "diarrheal diseases", "*E.coli*", "cholera", "dehydration", "irritable bowel syndrome", "machine learning", "salmonellosis", "Cryptosporidium", "artificial Intelligence", and "deep learning", as well as numerous keyword combinations.

The method of selecting research articles is determined by a set of inclusion and exclusion criteria, which involves four distinct phases: identification, when documents are identified by visiting multiple repositories, Screening is the process of transparently selecting papers by examining the decisions made at various phases of the systematic review. Eligibility refers to the process of evaluating all full-length articles, while included refers to the total number of papers accepted for the article. PRISMA improves the quality of reporting for systematic reviews, meta-analyses, and peer-reviewed publications are referred to understand current research on machine and deep learning techniques for detecting diarrheal diseases.

2.1 Overview

Mbunge et al. (2022)⁽¹¹⁾ did an experiment to predict diarrhoea among the children of Zimbabwe by using various machine learning classifiers such as decision tree, logistic regression, and support vector machine. Apart from this, they also highlighted the importance of understanding various risk factors such as socioeconomic, climatic, as well as demographic factors to address the diarrhoea among the survey which they did. The authors also stated that machine learning could help policymakers to create effective and adaptable preventive treatments, control programs, and approaches to combat diarrhoea in children under five.

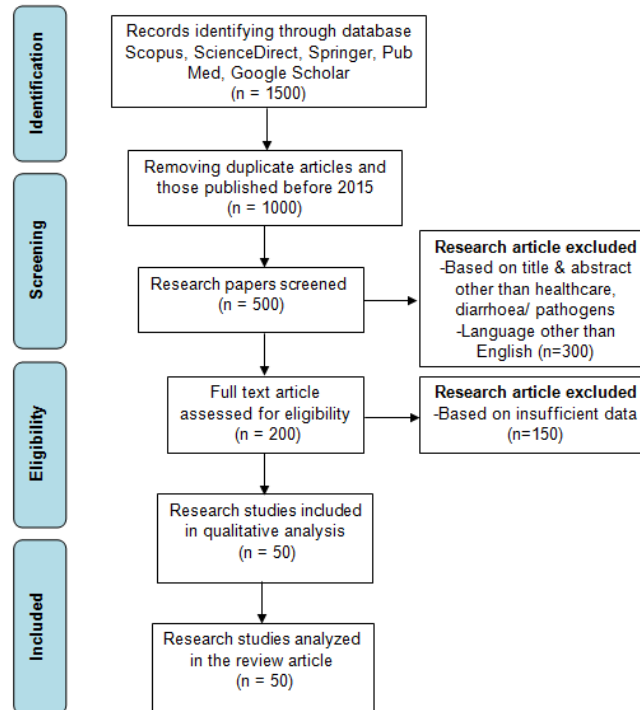


Fig 2. PRISMA criteria

Abdullahi et al. (2022)⁽¹²⁾ used machine learning techniques to investigate the impact of climate change on diarrhoea which outbreaks in nine South African provinces as well as integrated climate data for the prediction of daily diarrhoea cases. To generate synthetic data for augmentation and optimize the parameters of machine learning classifiers, Generative Adversarial Networks and Relevance Estimation and Value Calibration (REVAC) techniques had been used respectively. Results indicated the deep learning methods surpassing SVM along with the excellent performance of CNN with real-world data, while LSTM outperforms with synthetic data augmentation. Besides this, ML accuracy improved by at least 30% with data augmentation and on the contrary, REVAC enhanced CNN accuracy by approximately 2.5% in each province. Zahirzda et al. (2021)⁽¹³⁾ focused on the prediction of diarrhoea in children under five in Afghanistan. To conduct this experiment, they used three machine learning techniques such as Random Forest, Naïve Bayes, as well as Support Vector Machine (SVM). Initially, they cleaned the data using data preprocessing and later used Information Gain and Correlation-based Feature Selection algorithms for selecting features. They mentioned that their work extended to health departments would improve prevention decisions as well as aid policymaking in Afghanistan to provide a valuable tool for detecting childhood diarrhoea effectively. Yonanda et al. (2023)⁽¹⁴⁾ addressed the detection of pathogenic *Escherichia coli* causing diarrhoea in Indonesia, emphasizing the limitations of conventional diagnostic methods and proposing an automated approach using the Faster R-CNN deep learning method. The study employed Faster R-CNN with Inception v2 and ResNet-50 architectures, integrating augmentation and Image Enhancement into the Tryptic Soy Agar image dataset. Results highlighted the significant impact of Image Enhancement on model performance and offered a more efficient and accurate alternative to traditional diagnostic methods by holding potential for improving *Escherichia coli* detection and treatment in diarrheal diseases in Indonesia. Momand et al. (2023)⁽¹⁵⁾ addressed child dehydration in Afghanistan and introduced machine learning techniques such as Multilayer Perceptron, Random Forest, J48, Support Vector Machine, along with the Logistic Regression to diagnose it. The study was conducted on Afghanistan Demographic and Health Survey (ADHS) dataset, where Random Forest emerged as the most effective classifier by achieving precision of 91%, accuracy of 91.46%, and AUC of 94%. Their proposed model held promise for healthcare professionals by offering a tool for prompt and accurate identification of dehydration in young children as well as it also highlighted the potential of machine learning in enhancing early diagnosis of dehydration in Afghan children under five. Do et al. (2023)⁽¹⁶⁾ used standard and sophisticated machine learning approaches like SARIMA, CNN, LSTM, SVM, Xgboost, Catboost, BlockRNN, LightGBM, N-HiST, TFT, TCN, as well as Transformer to forecast diarrhea outbreaks. The study evaluated model performance in 55 provinces with various climates, topography, and socio-economic aspects to help researchers design early-warning systems.

The article also presented Dynamic Weighted Ensemble (DWE), a two-layer ensemble model that dynamically combined meta models for diarrhea prediction.

Wang et al. (2015)⁽¹⁷⁾ presented a three-layered feed-forward backpropagation Artificial Neural Network (ANN) model using the Levenberg–Marquardt algorithm for the prediction of weekly infectious diarrhoea cases based on meteorological factors. The chosen meteorological variables exhibited a strong correlation with infectious diarrhoea. Their study compared the ANN model with Random Forests Regression (RFR), Support Vector Regression (SVR), and Multivariate Linear Regression (MLR) using the same dataset and employed 5-fold cross-validation to prevent overfitting. The research indicated that the BPNN model was an effective quantitative instrument to evaluate and predict infectious diarrhea based on meteorological conditions, highlighting its use and accuracy. Tabata et al. (2023)⁽¹⁸⁾ explored the potential of an AI technique to detect subtle endoscopic changes associated with Irritable Bowel Syndrome (IBS). The researchers used Google Cloud Platform AutoML Vision where the AI image models exhibited a high AUC of 0.95 in discriminating between healthy and IBS individuals, with 30.8% sensitivity and 97.6% specificity for detecting IBS. The overall AUC for discriminating between different IBS subtypes and healthy subjects was 0.83. Chusyairi et al. (2021)⁽¹⁹⁾ mentioned about the infant deaths caused by diarrhoea and to identify health care centers with varying levels of diarrhoea cases, their study employed the Fuzzy C-Means algorithm which categorized into three groups based on the severity of diarrhoea sufferers. Their findings revealed that two health care sectors had the lowest, 14 had moderate, and the remaining exhibited a high number of diarrhoea cases. Their research suggested that these results could guide health departments in managing diarrheal diseases, potentially reducing infant mortality rates in regions with high occurrences of diarrhoea. The Fuzzy C-Means algorithm proved effective in stratifying health care centres based on the prevalence of diarrhoea cases, providing valuable insights for targeted healthcare interventions. Wahyudi and Andriani (2021)⁽²⁰⁾ developed a classification system for diarrhoea outbreaks data to uncover patterns and create classification rules for detecting the Case Fatality Rate (CFR) of diarrhoea. The study employed the C4.5 algorithm, known for its decision tree approach, and the Naïve Bayes algorithm, used a probabilistic approach in classification. They followed the Knowledge Discovery in Databases stages. After obtaining classification rules, evaluation was conducted using Confusion Matrix and Receiver Operating Characteristic Curve on both training and testing data. Results indicated that the C4.5 algorithm achieved higher accuracy compared to Naïve Bayes. Fang et al. (2020)⁽²¹⁾ used meteorological factors to create an ideal Random Forest (RF) model for predicting infectious diarrhoea in Jiangsu Province, China. Classic Autoregressive Integrated Moving Average (ARIMA)/X models were compared to the RF model, which included precipitation, atmospheric pressure, relative humidity, lagged components, 1–4 week lag morbidity, and a time Model construction using 2012–2016 morbidity and meteorological data, with testing using 2017 data. The RF model outperformed ARIMA/X models with a 20% MAPE. Afaq and Manocha (2022)⁽²²⁾ used dew computing-assisted smart monitoring to forecast diarrhoea severity by assessing the association between health, indoor climatic conditions, and food components. Smart sensors collect data, which can be classified using PW-NB at the cyber layer to identify abnormal health events. Through health, food, and environmental association analysis, a Multi-scale Gated Recurrent Unit (M-GRU) was proposed to measure severity. M-GRU outperformed LSTM, KNN, SVM, RNN, and DT in precision.

Ogwel et al. (2023)⁽²³⁾ covered the period of 1990–2021 and used PubMed by focusing on predictive modeling of diarrhoea in children. Among 2671 initially retrieved articles, 38 were included, with disease forecasts (36.8%), vaccine-related predictions (23.7%), and disease/pathogen detection (13.2%) being common research topics. The majority of studies (73.7%) were published between 2011 and 2020, employing machine learning techniques (31.6%) with various algorithms. Zaeni et al. (2021)⁽²⁴⁾ characterized toddler diarrhoea diseases and gave parents basic treatment options. Bivariate Pearson correlation testing identified eight significant attributes from 166 cases and 21 attributes from the Mojopanggung public health centre. Faeces frequency, consistency, temperature, eye condition, nausea/vomiting, thirst, as well as skin turgor were assessed. An oversampling strategy rectified class imbalance, yielding 318 dehydration cases. According to Confusion Matrix results, the Naive Bayes Classifier (NBC) algorithm had an precision of 96.55%, average accuracy of 97.48%, error rate of 2.52%, and recall of 96.27%. Their study helped to classify toddler diarrhoea by dehydration. Leo et al. (2019)⁽²⁵⁾ explored the application of machine learning techniques to model cholera epidemics by addressing data imbalance through the Principal Component Analysis (PCA) and Adaptive Synthetic Sampling Approach (ADASYN) as well as on applying the Wilcoxon sign-rank test identified the XGBoost classifier as the most effective model. In their study, Campbell et al. (2020)⁽²⁶⁾ used a machine learning technique to predict the risk of cholera in coastal India. The study focused on an area with more than 200 million people. The model utilized essential climate variables derived from terrestrial, atmospheric, along with the oceanic satellite data. They applied a random forest classifier to a dataset on cholera outbreaks in coastal Indian districts from 2010 to 2018. The model demonstrated high accuracy (0.99), F1 score (0.942), and sensitivity (0.895), indicating a correct identification rate of 89.5% for outbreaks. The study highlighted the potential of using random forest classifiers and satellite-derived climate variables to assess the risk of cholera in the environment. Yanik et al. (2020)⁽²⁷⁾ introduced a novel *E. coli* bacteria detection method,

utilizing deep learning instead of manual observation of color changes in endo agar medium. Experimental data demonstrated that automated detection of *E. coli* could be achieved within 6-10 hours, a significant improvement compared to the 22 ± 2 hours required by the traditional method. This innovative approach holds promise for quicker and more efficient detection of bacteria in water samples, with potential implications for public health and environmental monitoring.

Lopez et al. (2023)⁽²⁸⁾ used YOLOv5, a machine learning technique, to detect *Cryptosporidium* in optical along with phase contrast microscopy pictures. Two tests employing randomly initialized weights assessed model accuracy using confusion matrix, mean average accuracy, and F1 scores on 520 and 1200 images. The best model used phase contrast imaging and the optimal optical imaging training and retraining epoch. Considering computational expenses, YOLOv5s was the best *Cryptosporidium* detection model. The study indicated that machine learning can improve foodborne pathogen detection. Kumar et al. (2023)⁽²⁹⁾ explored the impact of foodborne diseases on public health and the economy, and the use of artificial intelligence in predicting and detecting pathogens in contaminated food. The study compared the accuracy of various prediction methods which provided valuable insights for researchers and practitioners in the field. The paper also discussed the most prevalent harmful bacteria and viruses that could cause foodborne diseases. Overall, the study highlighted the potential of artificial intelligence in improving food safety and reducing the incidence of foodborne illnesses such as diarrhoea.

Furthermore, the researchers' work has been thoroughly analyzed and contrasted based on specific factors, as indicated in Table 1.

Table 1. Analysis of the previous work

Author's Name	Dataset	Techniques	Outcomes	Challenges
Mbunge et al. (2022) ⁽¹¹⁾	Zimbabwe Demographic Health Survey dataset	Logistic Regression	Accuracy = 85% Precision = 86% Recall = 100% F1 score = 94%	Limited dataset, class imbalance
		Support Vector Machine	Accuracy = 84% Precision = 85% Recall = 100% F1 score = 92%	
Abdullahi et al. (2022) ⁽¹²⁾	Data of 9 features was collected from South Africa	CNN LSTM SVM	RMSE = 31.55% RMSE = 32.91% RMSE = 33.89%	Lack of generalization as the model has been trained only with the diarrhoea cases of South Africa people only
Zahirzda et al. (2021) ⁽¹³⁾	Afghanistan Demographic and Health Survey 2015	Random Forest	Accuracy = 77.10%	
Yonanda et al. (2023) ⁽¹⁴⁾	Tryptic Soy Agar image dataset	Naïve Bayes	AUC = 86.4% Precision = 77.3% Recall = 77.1% Accuracy = 69.85%	The work was conducted only on the subset of the collected data and the exclusion of environmental and social factors that may contribute to diarrhoea.
		Support Vector Machine	AUC = 75.9% Precision = 70.3% Recall = 69.9% Accuracy = 72.61%	
Momand et al. (2023) ⁽¹⁵⁾	Afghanistan Demographic and Health Survey	FasterRCNN + ResNet50	Accuracy = 92% Precision = 96% Recall = 95% F1 Score = 96%	The work was conducted on small dataset of 510 images
		Random Forest	Accuracy = 91.46%	
Tabata et al. (2023) ⁽¹⁸⁾	2479 endoscopy images 538 endoscopy images	MLP	Accuracy = 85.48%	Small dataset, class imbalance
		J48	Accuracy = 84.56%	
Tabata et al. (2023) ⁽¹⁸⁾	2479 endoscopy images 538 endoscopy images	SVM	Accuracy = 82.05%	Lack of generalization, Not reliable
		LR	Accuracy = 70.95%	
		Google cloud platform AutoML	AUC = 0.90 AUC = 0.45	

Continued on next page

Table 1 continued

		484 endoscopy images		AUC = 0.60	
Wahyudi and Andriani (2021) ⁽²⁰⁾		Data of 231 individuals	Naïve Bayes C4.5	Accuracy = 79.85%	The parameters of the model could be fine tuned to optimize its performance
Zaeni et al. (2021) ⁽²⁴⁾		Data collected from 166 cases	Naive Bayes Classifier	Accuracy = 82.82%	
				Accuracy = 97.48%	Underfitting
				Precision = 96.55%	
				Recall = 96.27%	
Leo et al. (2019) ⁽²⁵⁾		Data collected from Dar es Salaam region	XGB classifier	Error rate = 2.52%	Misclassification issue
				Sensitivity = 80.5%	
				Specificity = 73%	
				Accuracy = 76.7%	
Campbell et al. (2020) ⁽²⁶⁾		six ECV datasets	Random Forest	Accuracy = 99%	The research was conducted on under-reporting of cholera cases in coastal India
				F1 score = 94.2%	
				Sensitivity = 89.5%	
Karant et al. (2023) ⁽³⁰⁾	National Reporting System (NORS)	Outbreak System	Elastic Net	R square =0.669	Limited dataset
Ali et al. (2020) ⁽³¹⁾		Real time data	U-Net, Support vector machine	AUC = 83%	The model should be examined using various performance metrics

2.2 Research gaps

Table 1 provides a comparative analysis of various research studies on diarrheal disease detection and diagnosis, highlighting the datasets, techniques, outcomes, and limitations of each study. The findings reveal that most studies rely on machine learning techniques, such as logistic regression, random forest, and support vector machine, to predict and diagnose diarrheal disease. However, the limitations of small datasets^(11,30) and classification issue^(15,25) are common issues faced by these studies which hampers the performance of the models. Hence, to improve the performance, some studies have employed advanced techniques such as FasterRCNN with ResNet50 architectures, which have shown promising results in improving the accuracy of diagnosis and detection. However, these studies also face limitations such as the exclusion of environmental and social factors that may contribute to diarrhoea⁽¹³⁾ and the lack of generalization^(12,18,26). Furthermore, some studies have employed AI algorithms such as Naive Bayes^(13,20,24) and Elastic Net⁽³⁰⁾, which have shown high accuracy in detecting diarrheal disease. However, these studies also face limitations such as underfitting, limited dataset⁽¹⁴⁾, and the need for fine-tuning the model parameters to optimize performance.

Hence, the proposed modification to the methodology towards integrating advanced Convolutional Neural Networks (CNNs) and combining them in diverse ways marks a notable progress in the field of detecting and diagnosing diarrheal diseases. The research seeks to overcome the limits of short datasets and traditional machine learning methods by using advanced or hybridized CNN architectures and adopting various hyperparameter optimization techniques. This methodology has the ability to provide a new era of powerful and advanced models for detecting and diagnosing diarrhoea. This will help to improvise the diagnosis accuracy as well as bring innovation to the field of analyzing diarrheal diseases. Additionally, feature representation, model architecture, and assessment criteria can also be standardized to optimize diarrheal disease detection models.

RQ1: How can deep learning algorithms be applied as well as optimized for the accurate and rapid detection of diarrheal pathogens, considering the diverse nature of microbial species responsible for diarrhoea?

To optimize deep learning algorithms for the accurate and rapid detection of diarrheal pathogens considering the diverse nature of microbial species, a comprehensive framework involving several key components should be proposed, as shown in Figure 3

1. **Data collection:** Firstly, it is important to identify particular pathogens that cause diarrhoea, such as viruses (e.g., Rotavirus, Norovirus), bacteria (e.g., *Salmonella*, *Escherichia coli*), and parasites (e.g., *Giardia*, *Cryptosporidium*). These pathogens should be categorized based on their genetic and functional characteristics. In addition to this, it is important to examine the use of varied and inclusive datasets that are obtained from clinical samples, public health databases, and research studies. It is also crucial to promote collaboration between healthcare facilities and research organizations.

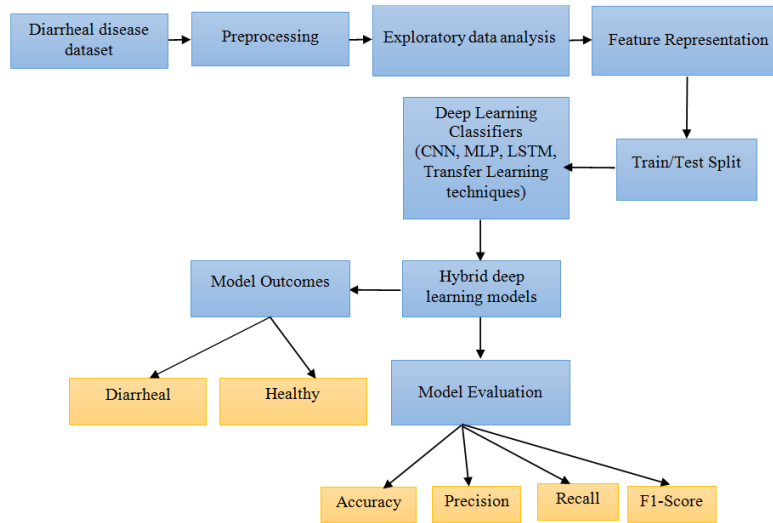


Fig 3. Proposed system to detect and classify diarrhoea

2. **Data Preprocessing:** Data preparation is a crucial step in detecting or diagnosing diarrheal diseases, especially when working with image datasets. In the earliest phases of preprocessing, the tasks of scaling and normalization are carried out. The purpose is to ensure that all images have the same input dimensions and consistent pixel intensity ranges⁽³²⁾. When dealing with numerical datasets concerning diarrheal diseases, particular emphasis is placed on managing missing values or NaN values. Imputation approaches customized to the unique attributes of the dataset are used to guarantee a thorough and reliable dataset for subsequent analysis. The selection of imputation techniques should be based on statistical rigor and domain-specific expertise, recognizing the potential influence of missing data on the precision and validity of the outcomes derived from the dataset⁽³³⁾.
3. **Exploratory data analysis:** Within the realm of diarrheal disease, exploratory data analysis (EDA) entails a methodical investigation of datasets pertaining to the identification of outliers or abnormalities that may necessitate particular attention. Visualization methods, such as bar charts and time-series plots, can be used to emphasize patterns in the occurrence of diseases. An examination of class distribution can uncover discrepancies in the intensity or types of microorganisms responsible for resulting in diarrhoea. Feature analysis involves examining the relationships between potential risk factors such as water quality, sanitation, and geographical location. In addition, exploratory data analysis (EDA) methods can be used to identify patterns in the prevalence of diarrheal diseases in different locations or countries. By analyzing the long term data on diarrheal disease, one can identify recurring seasonal patterns or instances of widespread prevalence of the disease^(34,35).
4. **Feature Representation:** When it comes to improving the detection of diarrheal diseases, it is crucial to have a careful approach to extracting or selecting features. The aim is to use advanced techniques like contour feature extraction, scaling, normalization, and principal component analysis to identify and extract specific features that are closely related to the unique characteristics of diarrheal diseases. At the same time, when we use rigorous feature selection methods, it helps to improve the dataset by keeping only the most important attributes. There are two benefits to this approach. Firstly, it helps to prevent overfitting and later improves the execution of the process by reducing the computational complexity^(36,37).
5. **Model Architecture:** In the analysis of diarrheal data, it is essential to carefully select or develop deep learning architectures, such as Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), as well as to consider the benefits of incorporating transfer learning⁽³⁸⁾. Additionally, systematic exploration of the hyperparameter space using techniques like grid search and random search is crucial for identifying the optimal combination that yields the highest performance⁽³⁹⁾. Furthermore, the inclusion of regularization techniques like dropout or L2 regularization is important to mitigate the risk of overfitting. Practical investigations into optimization algorithms such as Adam and RMSProp, which are variations of gradient descent, are also valuable. Moreover, the use of learning rate scheduling to dynamically adjust learning rates during training is essential for efficient and adaptive model convergence⁽⁴⁰⁾.
6. **Evaluation Metrics:** Ultimately, it is crucial to monitor the performance of the model on the validation set. Adjustments to the hyper-parameters should be made, and the best model should be selected based on key metrics such as precision,

accuracy, recall, and loss. These metrics are instrumental in evaluating the model’s effectiveness in detecting bacterial infections that cause diarrhoea⁽⁴¹⁾. The formula for computing these metrics is presented in Table 2.

Table 2. Formula of performance metrics

Metrics	Formula
Accuracy	$\frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$
Loss	$\frac{(Actual\ Value - Predicted\ Value)^2}{Number\ of\ observations}$
Precision	$\frac{True\ Positive}{True\ Positive + False\ Positive}$
Recall	$\frac{True\ Positive}{True\ Positive + False\ Negative}$
F1 score	$2 \frac{Precision * Recall}{Recall + Precision}$

RQ2: What is the cost-effectiveness of implementing AI-based diarrheal disease detection systems in healthcare settings?

Evaluating the cost-effectiveness of integrating AI-powered systems for detecting diarrheal diseases in healthcare settings requires a comprehensive and intricate methodology. The incorporation of such systems offers significant benefits, such as decreased diagnostic time, enhanced precision, and optimized utilization of resources. Rapidly analyzing large datasets allows for quick detection of trends, resulting in more focused and efficient therapies, thus reducing wasteful interventions and related expenses⁽⁴²⁾. Moreover, the improved effectiveness in diagnostic procedures leads to the efficient use of healthcare resources, potentially leading to financial savings and enhanced facility utilization. AI systems enable early detection, which can prompt timely preventative interventions, reducing the severity of illnesses and consequent healthcare expenses. Nevertheless, it is essential to recognize the primary expenses associated with algorithmic development, personnel instruction, and system incorporation⁽⁴³⁾. Apart from this, evaluating the overall effectiveness and results for patients, it is important to think about the relationship between initial investments and long-term advantages. Continual expenses for maintenance and updates are crucial to guarantee long-term efficiency. It is important to focus on dealing with both false positive and bad results, improving algorithms, and taking into account the wider social consequences, such as accessibility⁽⁴⁴⁾.

RQ3: What are the ways in which collaborating with public health institutions might improve the incorporation of AI-driven technologies into surveillance and monitoring systems for diarrheal diseases?

Statista reports that the global value of AI in healthcare was approximately \$11 billion in 2021 and is projected to reach \$188 billion by 2030. This growth is expected to occur at a compound annual growth rate (CAGR) of 37% from 2022 to 2030. Figure 4 shows the growth of AI market in the realm of healthcare up to year 2030⁽⁴⁵⁾.

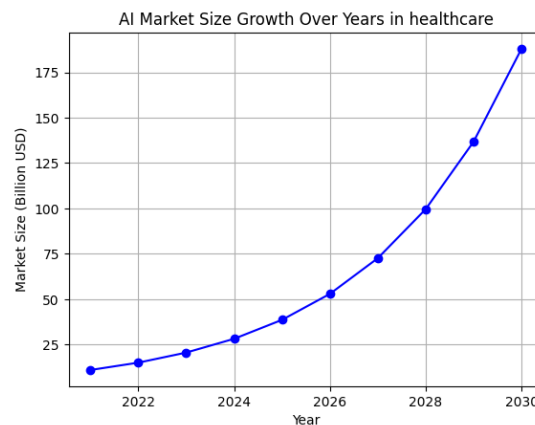


Fig 4. Growth of AI in healthcare

Therefore, considering these facts, it can be concluded that doing collaborative research with public health institutes is a crucial pathway for promoting the incorporation of AI-based tools into surveillance and monitoring systems. One important

aspect is to establish strong data sharing and integration methods by working together with public health organizations to get and combine various datasets such as clinical records, environmental factors, and demographic information. This requires a dedication to interoperability and defined data formats in order to smoothly include AI algorithms⁽⁴⁶⁾. In addition, the collaborative effort encompasses the development and validation of algorithms to necessitate strong collaboration with public health professionals and ensures that AI algorithms are in line with the unique requirements of diarrheal disease surveillance. It is also crucial to conduct real-world validation in order to confirm the accuracy, dependability, and generalizability of results in various populations and circumstances⁽⁴⁷⁾. Besides this, adhering to norms and regulations is necessary to ensure ethical and privacy considerations. It is important to actively collaborate on developing secure data-sharing frameworks that protect sensitive information. Collaboration also involves providing user training and capacity building programs to help healthcare professionals better understand AI technology and how to use it for surveillance and monitoring purposes⁽⁴⁸⁾. In addition to this, it is also essential to develop AI solutions that can be easily expanded and adapted to public health settings with limited resources. This includes taking into account aspects such as infrastructure, knowledge, and financial constraints⁽⁴⁹⁾. The primary focus is on community engagement, which highlights the participation of local communities in the construction of AI-based surveillance systems to guarantee cultural sensitivity and community acceptance. Open and clear communication about the advantages of technology fosters confidence and effectively tackles any concerns within the community. The successful incorporation of AI-based technologies into surveillance and monitoring systems for diarrheal illnesses in public health contexts can be achieved by thoroughly addressing these aspects through collaborative research⁽⁵⁰⁾.

3 Conclusion

The paper examines the need of precise diagnosis and identification of diarrheal disease, specifically in relation to public health in poor nations. It underscores the constraints of existing diagnostic approaches and highlights the potential of AI techniques, particularly machine learning algorithms, in accelerating the diagnosis process and improving efficiency. The study suggests employing a hybrid deep learning model to enhance the precision of diagnosing and detecting diarrheal diseases. Furthermore, it examines the economic viability of deploying AI-powered systems for detecting diarrheal diseases in healthcare sector and the potential advantages of partnering with public health organizations to improve the integration of AI-driven technologies into surveillance and monitoring systems for such diseases. The existing research is limited by the use of small and imbalanced datasets, which results in misclassification, lack of generalization, modelling errors, and reduced effectiveness of various learning models in detecting and classifying diarrheal diseases or their pathogens. Hence, it is imperative for future study to prioritize the enlargement of the dataset to a more extensive magnitude, integrating a hybrid deep learning model and fine-tuning its parameters. Furthermore, it is crucial to acknowledge and tackle ethical and privacy issues while developing cooperation between public health institutions and AI specialists to successfully incorporate AI-driven solutions in the fight against diarrheal diseases.

References

- 1) Diarrhoeal disease . 2017. Available from: <https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease>.
- 2) Nemeth V, Pflieghaar N, Diarrhea. Diarrhea. 1920.
- 3) Shah S, Kongre V, Kumar V, Bharadwaj R. A Study of Parasitic and Bacterial Pathogens Associated with Diarrhea in HIV-Positive Patients. *Cureus*. 2016;8(9). Available from: <https://doi.org/10.7759/cureus.807>.
- 4) Clinic C. Professional CCM. Diarrhea. 2023. Available from: <https://my.clevelandclinic.org/health/diseases/4108-diarrhea>.
- 5) Shane AL, Mody RK, Crump JA, Tarr PI, Steiner TS, Kotloff K, et al. 2017 Infectious Diseases Society of America Clinical Practice Guidelines for the Diagnosis and Management of Infectious Diarrhea. *Clinical Infectious Diseases*. 2017;65(12):e45–e80. Available from: <https://doi.org/10.1093/cid/cix669>.
- 6) Robins-Browne RM, Levine MM. Laboratory Diagnostic Challenges in Case/Control Studies of Diarrhea in Developing Countries. *Clinical Infectious Diseases*. 2012;55(suppl_4):S312–S316. Available from: <https://doi.org/10.1093/cid/cis756>.
- 7) Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *Journal of Ambient Intelligence and Humanized Computing*. 2023;14(7):8459–8486. Available from: <https://doi.org/10.1007/s12652-021-03612-z>.
- 8) Giordano C, Brennan M, Mohamed B, Rashidi P, Modave F, Tighe P. Accessing Artificial Intelligence for Clinical Decision-Making. *Frontiers in Digital Health*. 2021;3. Available from: <https://doi.org/10.3389/fdgh.2021.645232>.
- 9) Wong C. AI Listens to Toilet Sounds to Guess Whether People Have Diarrhoea. . Available from: <https://www.newscientist.com/article/2350082-ai-listens-to-toilet-sounds-to-guess-whether-people-have-diarrhoea/>.
- 10) Broadfoot M. AI Detects Patterns of Gut Microbes for Cholera Risk. *Research Medicine*. 2018. Available from: <https://today.duke.edu/2018/05/ai-detects-patterns-gut-microbes-cholera-risk>.
- 11) Mbunge E, Chemhaka G, Batani J, Gurajena C, Dzinamarira T, Musuka G, et al. Predicting Diarrhoea Among Children Under Five Years Using Machine Learning Techniques. In: *Artificial Intelligence Trends in Systems*. Springer International Publishing. 2022;p. 94–109. Available from: https://doi.org/10.1007/978-3-031-09076-9_9.
- 12) Abdullahi T, Nitschke G, Sweijd N. Predicting diarrhoea outbreaks with climate change. *PLOS ONE*. 2022;17(4):e0262008–e0262008. Available from: <https://doi.org/10.1371/journal.pone.0262008>.

- 13) Zahirzda A, Chanmas G, Chan JH. A data mining model for predicting diarrhea in Afghan children. Institute of Electrical and Electronics Engineers (IEEE). 2021. Available from: <https://doi.org/10.36227/techrxiv.13711747.v1>.
- 14) Yonanda YP, Putra AP, Purwanti E. Automatic Detection of Escherichia coli Bacteria from Tryptic Soy Agar Image Using Deep Learning Method. *Indonesian Applied Physics Letters*;4(2):45–56. Available from: <https://doi.org/10.20473/iapl.v4i2.46793>.
- 15) Momand Z, Pal D, Mongkolnam P, Chan JH. A Machine Learning Approach to Detect Dehydration in Afghan Children. 2023. Available from: <https://doi.org/10.48550/arXiv.2305.13275>.
- 16) Do TD, Nguyen TD, Ta VC, Anh DT, Thi THT, Phan DH, et al. Dynamic weighted ensemble for diarrhoea incidence predictions. *Machine Learning*. 2023;p. 1–24. Available from: <https://doi.org/10.1007/s10994-023-06465-z>.
- 17) Wang Y, Li J, Gu J, Zhou Z, Wang Z. Artificial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China). *Applied Soft Computing*. 2015;35:280–290. Available from: <https://doi.org/10.1016/j.asoc.2015.05.047>.
- 18) Tabata K, Mihara H, Nanjo S, Motoo I, Ando T, Teramoto A, et al. Artificial intelligence model for analyzing colonic endoscopy images to detect changes associated with irritable bowel syndrome. *PLOS Digital Health*. 2023;2(2):e0000058–e0000058. Available from: <https://doi.org/10.1371/journal.pdig.0000058>.
- 19) Chusyairi A, Saputra PRN. Fuzzy C-Means Clustering Algorithm For Grouping Health Care Centers On Diarrhea Disease. *International Journal of Artificial Intelligence Research*. 2021;5(1):35–43. Available from: <https://doi.org/10.29099/ijair.v5i1.191>.
- 20) Wahyudi M, Andriani A. Application of C4.5 and Naïve Bayes Algorithm for Detection of Potential Increased Case Fatality Rate Diarrhea. *Journal of Physics: Conference Series*. 2021;1830(1):012016–012016. Available from: <https://doi.org/10.1088/1742-6596/1830/1/012016>.
- 21) Fang X, Liu W, Ai J, He MJ, Wu Y, Shi Y, et al. Forecasting incidence of infectious diarrhea using random forest in Jiangsu Province, China. *BMC Infectious Diseases*. 2020;20(1). Available from: <https://doi.org/10.1186/s12879-020-4930-2>.
- 22) Afaq Y, Manocha A. Dew computing-assisted cognitive Intelligence-inspired smart environment for diarrhea prediction. *Computing*. 2022;104(11):2511–2540. Available from: <https://doi.org/10.1007/s00607-022-01097-y>.
- 23) Ogwel B, Mzazi V, Nyawanda BO, Otieno G, Omoro R. Predictive modeling for infectious diarrheal disease in pediatric populations: A systematic review. *Learning Health Systems*. 2024;8(1). Available from: <https://doi.org/10.1002/lrh2.10382>.
- 24) Zaeni IA, Ratri DP, Taufani AR. Classification of the Degree of Dehydration in Diarrhea Using the Naïve Bayes Algorithm. 2021. Available from: <https://doi.org/10.1109/IBITeC53045.2021.9649060>.
- 25) Leo J, Luhanga E, Michael K. Machine Learning Model for Imbalanced Cholera Dataset in Tanzania. *The Scientific World Journal*. 2019;2019:1–12. Available from: <https://doi.org/10.1155/2019/9397578>.
- 26) Campbell AM, Racault MFF, Goult S, Laurenson A. Cholera Risk: A Machine Learning Approach Applied to Essential Climate Variables. *International Journal of Environmental Research and Public Health*. 2020;17(24):9378–9378. Available from: <https://doi.org/10.3390/ijerph17249378>.
- 27) Yanik H, Kaloğlu AH, Değirmenci E. Detection of escherichia coli bacteria in water using deep learning: A faster r-cnn approach. 2020. Available from: <https://doi.org/10.31803/rg-20200524225359>.
- 28) Salguero JSL, Rondón RM, TJV, Gil JAC, Galvis CAN, OML, et al. Automatic detection of Cryptosporidium in optical microscopy images Using YOLOv5x: A comparative study. Available from: <https://doi.org/10.1139/bcb-2023-0059>.
- 29) Kumar Y, Kaur I, Mishra S. Foodborne Disease Symptoms, Diagnostics, and Predictions Using Artificial Intelligence-Based Learning Approaches: A Systematic Review. *Archives of Computational Methods in Engineering*. 2023;p. 1–26. Available from: <https://doi.org/10.1007/s11831-023-09991-0>.
- 30) Karanth S, Patel J, Shirmohammadi A, Pradhan AK. Machine learning to predict foodborne salmonellosis outbreaks based on genome characteristics and meteorological trends. *Current Research in Food Science*. 2023;6:100525–100525. Available from: <https://doi.org/10.1016/j.crf.2023.100525>.
- 31) Ali N, Kirchhoff J, Onoja PI, Tannert A, Neugebauer U, Popp J, et al. Predictive Modeling of Antibiotic Susceptibility in *E. Coli* Strains Using the U-Net Network and One-Class Classification. *IEEE Access*. 2020;8:167711–167720. Available from: <https://doi.org/10.1109/ACCESS.2020.3022829>.
- 32) Ayalew AM, Salau AO, Tamyalew Y, Abeje BT, Woreta N. X-Ray image-based COVID-19 detection using deep learning. 2023. Available from: <https://doi.org/10.1007/s11042-023-15389-8>.
- 33) Feng S, Hategeka C, Grépin KA. Addressing missing values in routine health information system data: an evaluation of imputation methods using data from the Democratic Republic of the Congo during the COVID-19 pandemic. *Population Health Metrics*. 2021;19(1). Available from: <https://doi.org/10.21203/rs.3.rs-422960/v1>.
- 34) Thakur K, Kaur M, Kumar Y. A Comprehensive Analysis of Deep Learning-Based Approaches for Prediction and Prognosis of Infectious Diseases. *Archives of Computational Methods in Engineering*. 2023;30(7):4477–4497. Available from: <https://doi.org/10.1007/s11831-023-09952-7>.
- 35) Kumar Y, Koul A, Mahajan S. A deep learning approaches and fastai text classification to predict 25 medical diseases from medical speech utterances, transcription and intent. *Soft Computing*. 2022;26(17):8253–8272. Available from: <https://doi.org/10.1007/s00500-022-07261-y>.
- 36) Kaur S, Kumar Y, Koul A, Kamboj SK. A Systematic Review on Metaheuristic Optimization Techniques for Feature Selections in Disease Diagnosis: Open Issues and Challenges. *Archives of Computational Methods in Engineering*. 2023;30(3):1863–1895. Available from: <https://doi.org/10.1007/s11831-022-09853-1>.
- 37) Kumar Y, Koul A, Sisodia PS, Shafi J, Verma K, Gheisari M, et al. Heart Failure Detection Using Quantum-Enhanced Machine Learning and Traditional Machine Learning Techniques for Internet of Artificially Intelligent Medical Things. *Wireless Communications and Mobile Computing*. 2021;2021:1–16. Available from: <https://doi.org/10.1155/2021/1616725>.
- 38) Koul A, Bawa RK, Kumar Y. Artificial intelligence techniques to predict the airway disorders illness: a systematic review. 2023. Available from: <https://doi.org/10.1007/s11831-022-09818-4>.
- 39) Singh J, Sandhu JK, Kumar Y. An Analysis of Detection and Diagnosis of Different Classes of Skin Diseases Using Artificial Intelligence-Based Learning Approaches with Hyper Parameters. *Archives of Computational Methods in Engineering*. 2023;p. 1–28. Available from: <https://doi.org/10.1007/s11831-023-10005-2>.
- 40) Bansal K, Bathla RK, Kumar Y. Deep transfer learning techniques with hybrid optimization in early prediction and diagnosis of different types of oral cancer. *Soft Computing*. 2022;26(21):11153–11184. Available from: <https://doi.org/10.1007/s00500-022-07246-x>.
- 41) Koul A, Bawa RK, Kumar Y. An Analysis of Deep Transfer Learning-Based Approaches for Prediction and Prognosis of Multiple Respiratory Diseases Using Pulmonary Images. *Archives of Computational Methods in Engineering*. 2023;p. 1–27. Available from: <https://doi.org/10.1007/s11831-023-10006-1>.
- 42) Khanna NN, Maindarkar MA, Viswanathan V, Fernandes JFE, Paul S, Bhagawati M, et al. Economics of Artificial Intelligence in Healthcare: Diagnosis vs. Treatment. *Healthcare*. 2009;10(12):2493–2493. Available from: <https://doi.org/10.3390/healthcare10122493>.
- 43) Modi K, Singh I, Kumar Y. A Comprehensive Analysis of Artificial Intelligence Techniques for the Prediction and Prognosis of Lifestyle Diseases. *Archives of Computational Methods in Engineering*. 2023;30(8):4733–4756.

- 44) Chang H, Choi JYY, Shim J, Kim M, Choi MY. Benefits of Information Technology in Healthcare: Artificial Intelligence, Internet of Things, and Personal Health Records. *Healthcare Informatics Research*. 2023;29(4):323–333. Available from: <https://doi.org/10.4258/hir.2023.29.4.323>.
- 45) Stewart C. Artificial intelligence (AI) in Healthcare Market Size Worldwide 2021–2030. 2021. Available from: <https://www.statista.com/statistics/1334826/ai-in-healthcare-market-size-worldwide>.
- 46) Alowais SA, Alghamdi SS, Alsuehaby N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Medical Education*. 2023;23(1):689–689. Available from: <https://doi.org/10.1186/s12909-023-04698-z>.
- 47) Zeng D, Cao Z, Neill DB. Artificial intelligence-enabled public health surveillance—from local detection to global epidemic monitoring and control. *Artificial Intelligence in Medicine*. 2021;p. 437–453. Available from: <https://doi.org/10.1016/B978-0-12-821259-2.00022-3>.
- 48) Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthcare Journal*. 2021;8(2):e188–e194. Available from: <https://doi.org/10.7861/fhj.2021-0095>.
- 49) Naik N, Hameed BMZ, Shetty DK, Swain D, Shah M, Paul R, et al. Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility? *Frontiers in Surgery*. 2022;9. Available from: <https://doi.org/10.3389/fsurg.2022.862322>.
- 50) Kumar Y, Singla R. Effectiveness of Machine and Deep Learning in IOT-Enabled Devices for Healthcare System. In: Internet of Things. Springer International Publishing. 2022;p. 1–19. Available from: https://doi.org/10.1007/978-3-030-81473-1_1.