Al-Rafidain Journal of Computer Sciences and Mathematics (RJCM)

www.csmj.mosuljournals.com

Intelligent Prediction of Human Health Risks Based on Medical History: A Review

Mais Irreem A kamal[1,*] and Fawziya M. Ramo[2]

[1,2] Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Iraq
Emails: maisirreem_alsaigh@uomosul.edu.iq, fawziyaramo@uomosul.edu.iq

Article information

Article history:
Received :11/3/2024
Accepted :29/5/2024
Available online :15/12/2024

Abstract

The main access to modern healthcare using artificial technologies is related to the medical topic and prediction of risks to human health. Enhance patient medical care using intelligent prediction models such as machine learning, like gradient boosting trees, supervised machine learning and logistic regression which have a great importance in detecting diseases by analyzing medical images and diagnosing chronic diseases, in addition to use deep learning models like deep neural networks, recurrent neural networks, and long short term memory to predict many disease like depression risk, lung cancer, heart diabetic and kidney diseases. Enhance healthcare provider insights using intelligent prediction models to predict future health conditions, treatment outcomes, and disease progression. Moreover, the contribution of the intelligent prediction model helps the healthcare professional identify potential risks and intervene proactively by analyzing patient historical data for early diseases detection like using. Ultimately, the combination of medical history, intelligent prediction, and healthcare data analysis will empower healthcare providers with valuable tools to improve patient outcomes in efficient healthcare organizations.

Keywords:
Machine learning, deep learning, health care, prediction, medical history.

Correspondence:
Author: Mais Irreem A kamal
Email: maisirreem_alsaigh@uomosul.edu.iq

I. Introduction

This research briefly explores the impact of intelligent algorithms such as machine learning and deep learning algorithms in healthcare and disease prediction. By enabling data-driven insights, these algorithms are reshaping healthcare practices. The use of machine learning models, their adaptability, and their ability to handle structural data play an essential role in predicting diseases, improving treatment plans, and simplifying administrative tasks. In contrast, deep learning algorithms can handle unstructured data such as electronic health records and medical images, which greatly contribute to accurate diagnosis of diseases. This summary highlights the integration and diversity of these important approaches to enhancing patient care, interest in enhancing healthcare operations, and a new era of patient data analysis and personalized healthcare solutions.

Healthcare fields refer to improving the mental, physical and emotional health of an individual. It includes a wide range of activities and services designed to diagnose, prevent and treat a patient's disease or injury. Health care services include medical care, rehabilitation services, dental care, mental health services, and many other areas. The meaning of health care is usually any provision made by health care professionals such as nurses, doctors, dentists, and health professionals. These healthcare professionals work in a variety of settings, such as private practices, clinics, hospitals, and community health centers [1]. The health care system is supported by a wide range of organizations and agencies, such as government agencies, insurance companies, and nonprofit organizations. Access to health care is a fundamental right, and many governments around the world have introduced programs and policies to ensure that all people have access to promising health care services. Which depends on a variety of factors, including income, cultural beliefs, and geography[2].

The meaning of health care refers to the provision of high-quality medical care to individuals and communities in order to maintain, promote and restore health. It includes many services such as diagnosis and treatment of diseases, preventive care, rehabilitation, and palliative care should be provided to those suffering from chronic diseases [3], see Fig. 1. Healthcare systems can vary from country to country, but they include the integration of providers Services from the public and private sectors, such as clinics, hospitals, and doctors' offices[4]. Health care is fundamental to maintaining the mental, physical and social well-being of individuals, and is an essential part of a human right recognized by the United Nations[5].

Machine learning plays an important role in healthcare revolutions in many ways, for example in treatment and diagnosis. Machine learning uses patient data analysis and provides accurate and faster diagnosis. Furthermore, it suggests many plans for personalized treatment based on the patient's medical history. Genetics and any other factors. Machine learning can be used to monitor patients remotely and adjust treatment plans. Also in the field of medical imaging, machine learning can help radiologists diagnose medical images, such as CT scans, , drug discovery and manufacturing, smart health records, medical treatment, disease identification and diagnosis, medical imaging, disease records and disease prediction.

Deep learning has emerged as an important benefit in healthcare to its diverse and remarkable applications. as cause to its ability to learn complex patterns automatically from heterogeneous and large dataset, it make a huge revolution in medical imaging analysis, enabling early and accurate disease diagnosis by using deep learning techniques like convolutional neural networks, recurrent neural networks and long short-term memory, which are used to analyzing time-series data, predicting disease progression and facilitating patient monitoring. Deep learning has capability to natural language processing and improved unstructured data analysis, like medical literature, electronic health records, that lead to more personalized and efficient healthcare decision-making. The deep learning benefits in healthcare fields extend to genomics, drug discovery, and precision medicine, make the potential to reduce costs, accelerate research, and improve patient outcomes.

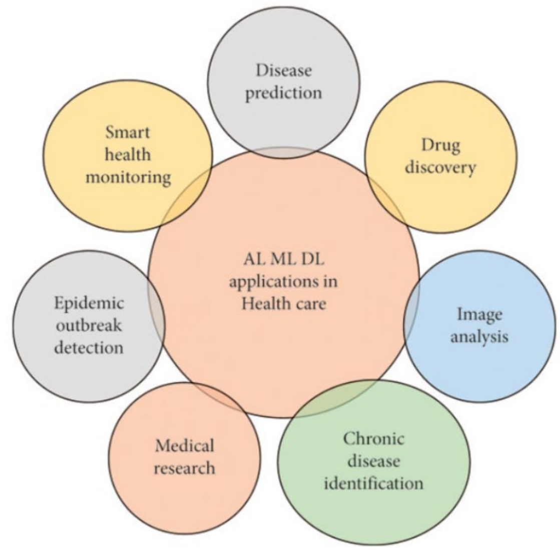

Fig 1. Application of Artificial Intelligent, Machine learning and Deep learning in Healthcare(Bordoloi *et al.*, 2022)

II. Artificial Intelligent and Health Care

It is important to know that many artificial intelligent algorithms have potential to improve healthcare in a number of ways, it's not a replacement for human healthcare professionals. Machine learning and Deep Learning should be used as a tool to assist healthcare professionals in providing the best possible care to patients. However, the use of machine learning in this field raises many privacy and ethical concerns. It's crucial that Machine learning is developed and deployed in a transparent and responsible manner to ensure its benefits patients and healthcare providers without causing harm or infringing on individual rights. See Fig. 2.

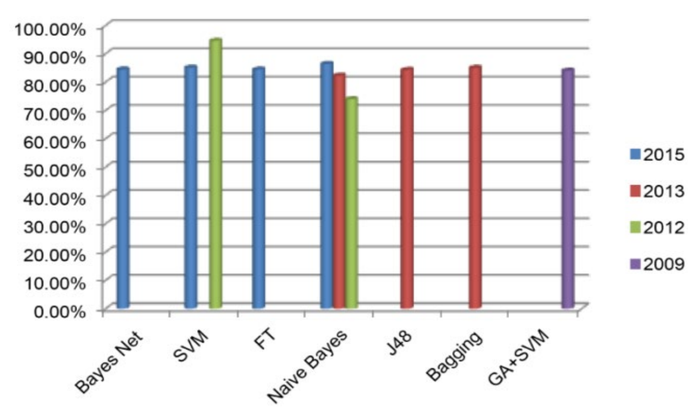

Fig 2. The most popular machine learning algorithms used in the Medical literature(Fatima and Pasha, 2017).

Artificial intelligent models are only as good as the data they are trained on, it is crucial to have high-quality and diverse datasets in healthcare. Additionally, it's important to validate these models on new data before they are used in clinical practice to ensure their accuracy and reliability.

Machine learning and deep learning techniques used to solve many different tasks in healthcare approach, examples of these tasks are:

Clustering: it can help the researcher to group together the similar medical cases to analyze the patterns that are founded and conduct research in the future.

Classification: by the aid of its algorithms which can help to label and determine the type of disease or medical case [23].

Prediction: it can make a diseases prognosis on how the future events will unfold by using current data.

Recommendations: without the need to actively search, algorithms can offer necessary medical information.

Anomaly detection: by using it in healthcare, it will be able to find any things that stand out from common patterns and determine whether they require any actions to be done or not.

III. Basic Model Utilize Artificial Intelligent Techniques in Healthcare

The basic model that use artificial intelligent techniques in healthcare consists of four stage, See Fig. 3:

Stage 1: (Formation of research questions): Identify the specific problem or question you want to solve using machine learning or deep learning techniques in healthcare.

Stage 2: (Collection of Data): Collect relevant data from various sources, including electronic medical records, imaging data, clinical trials, and other relevant data sources[44]. The data preprocessing is done by cleaning, normalizing, and then transforming it into a suitable format to analysis it.

Stage 3: (data preprocessing, and extract the data characteristics): This stage identify the most important variable in the data as a feature selection which will be used for machine learning algorithm or deep learning algorithm. This step will reduce the dimension of the data and preserving its most important features.

Stage 4: (train model and perform evaluation): following steps are done in this stage:

Machine learning or deep learning algorithm choosing: By depending on the data and problem definition, the most appropriate machine learning or deep learning algorithms must apply, such as decision trees, support vector machines, random forests, convolutional neural networks, recurrent neural networks [27].

Algorithm training and validating: Divide the dataset into two sets training and validation. first train the algorithm on the training set and then test it on the validation set. cross-validation techniques is used to optimize the used algorithm's hyper parameters.

Model evaluation: Performance evaluation by using any metrics such as accuracy, precision, F1 score, AUC-ROC and recall [17].

Model deployment: after the satisfactory of the model's performance, it must deployed into the healthcare system. This step may require additional considerations such as regulatory compliance, ethical concerns, and cyber security.

Monitor and update the model: Monitor the model's performance over time and update it as needed based on new data or changes in the problem or the healthcare system.

A different types of medical data used in artificial intelligent techniques[19] in diseases prediction like electronic health records (EHRs), medical images, laboratory results, vital signs, medication records, medical history, treatment plans and progress notes, biometric data and wearable devices and health applications, theses data need many steps to convert this row data to a useful one to make an accurate diseases predation, the steps are:

Data Source: In healthcare, data can come from various sources, like electronic Health, records (EHRs), medical claims data, clinical trial data, clinical trial data, medical device data, patient-generated data and public health data.

There are digital versions of patients' medical records, including their medical history, diagnoses, medications, laboratory results, and imaging studies. The medical claims data are generated when healthcare providers submit claims to insurance companies for reimbursement. The clinical trial data are generated during clinical trials, which are studies that test new treatments or interventions in humans. It can provide valuable information about the safety and efficacy of new treatments. Medical devices such as heart monitors, pacemakers, and insulin pumps can generate data that can be used to monitor a patient's health and track treatment progress.

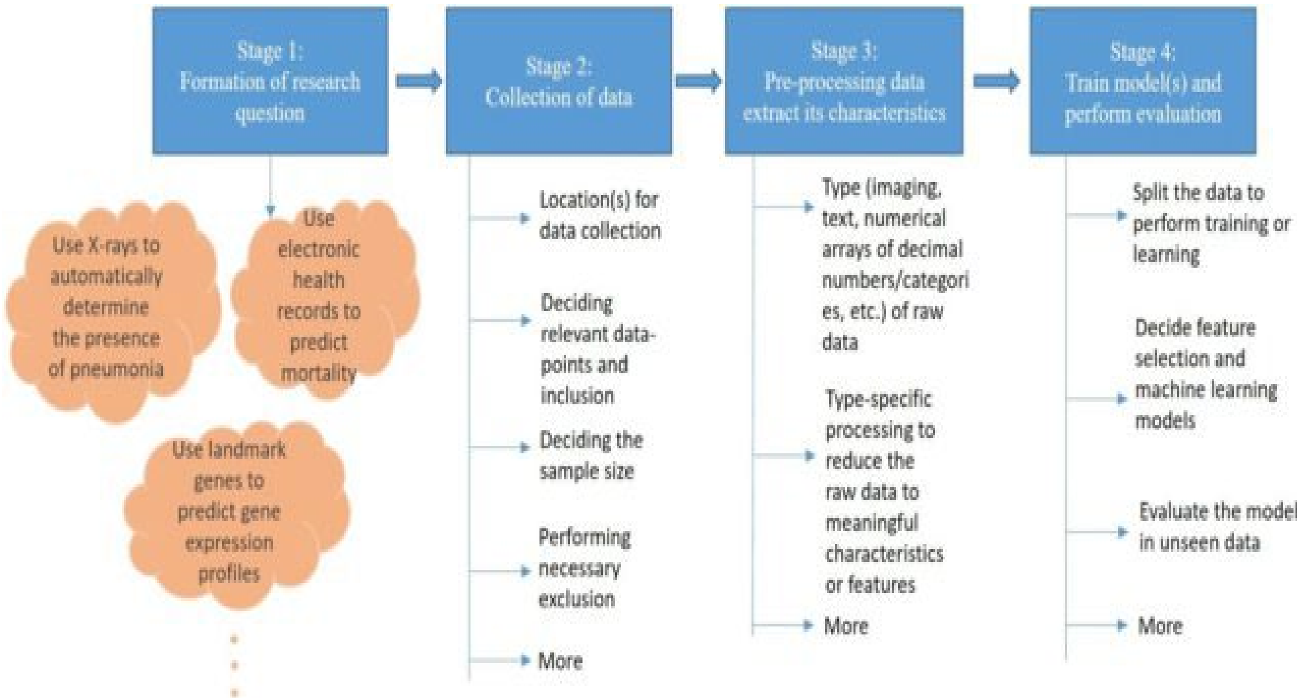

Fig 3. The basic model that use Artificial intelligent techniques in Healthcare(Saha *et al.*, 2020)

IV. Machine Learning Approaches in Healthcare

Machine learning plays an important role in modern healthcare by benefit from data-driven insights to enhance patient care, diagnosis, and operational efficiency. It enables early disease detection and personalized treatment plans through the analysis of vast patient datasets, including medical records and imaging. machine learning algorithms facilitate drug discovery, making it faster and more cost-effective, while also optimizing clinical trials. Electronic health records benefit from machine learning's ability to automate administrative tasks, improving accuracy and reducing paperwork. Predictive analytics aid in resource allocation and early intervention, while telemedicine and remote monitoring systems use machine learning to provide continuous patient care. Additionally, machine learning contributes to fraud detection and efficient healthcare management. In essence, machine learning is transforming healthcare by making it more data-driven, efficient, and patient-centric. In this section, we discuss existing research efforts that predict health risks or health diseases using various machine learning techniques. A thorough overview of the use of deep learning and machine learning methods in healthcare predictive analytics is given in this study. It investigates how well different models—such as random forests, decision trees, and logistic regression—predict illness risks from electronic health records (EHRs). The paper shows how deep learning models may automatically extract important features, improving prediction accuracy for a variety of diseases, and emphasizes the significance of handling imbalanced datasets[46]. With a focus on predictive analytics, this systematic study highlights the transformational potential of artificial intelligence in patient care by examining its application in disease detection. The study discusses a variety of artificial intelligent methods for analyzing massive medical datasets for patient risk assessment, medication discovery, and disease prediction. These methods include deep learning and machine learning. One way to improve patient outcomes and diagnostic accuracy is to integrate artificial intelligent with current healthcare systems[47]. This study introduces a wearable smart sensor system that monitors and predicts health risks by analyzing human perspiration data. The system uses advanced sensing technology combined with artificial intelligent algorithms to provide real-time health status updates and risk assessments. By correlating sweat analytes with medical history, the system aims to offer personalized health monitoring and early warnings for potential health issues, thereby improving preventive healthcare strategies[48]. Walaa Gouda and Rabab Yasin , in 2022, aimed to assess a computed tomography (CT) to help analyst the covid-19 and the disease severity prediction, there were 1020 patient enrolled 81.7 %males and 18.3 % women, the severity score of the lung and crazy paving and consolidation are used to indicate the sensitivity in the qualitative method.The study showed a clear discrepancy in sensitivity and specificity score between the groupe which used CT indicators or not, that reach to 84% sensitivity and

81% specificity score for the optical score, and 90% sensitivity and 81% specificity score for the high opacity, they therefore announce that the use of pneumonia assessment is a major advance in CT for COVID-19, as it provides a rapid and accurate tool for assessing sensitivity and helping to predict critical, high-risk cases[16].

Shengtao Dongm, et al, in 2022, The interactive electronic medical system is used to illustrate the diagnosis of diseases, by diagnosing 237 people with symptoms out of 1957 people. Whereas, stepwise logistic regression analysis showed that female sex, hypothyroidism, hyperlipidemia and diabetes were independent risk factors for the whole group.

The study depends on the gender recognition is distinguish between the woman were the diabetes mellitus in the female cases were significantly associated with the symptomatic RCCT (rotator cuff calcific tendinitis), and the male were diabetes, hyperlipidemia, hypothyroidism and mellitus have been diagnosed. By the aid of artificial intelligence many musculoskeletal disorders have been diagnosis and treatment that lead to early prediction and treatment[23].

Chui S Chu, et al., in 2020, design a model the predict the natural history of oral squamous cell carcinoma (OSCC) tumor behavior is the best way to make treatment plans and developing optimal strategies. In this paper many machine learning algorithms have been employed to improve clinical this prediction.

A review of 19-year period which includes a 467 OSCC patients to construct a detailed clinic pathological database. an 34 features from it have been used in four different machine learning algorithms, linear regression (LR), support vector machine (SVM), decision tree (DT) and k-nearest neighbors (KNN) models. These predictive approaches assessed by area under the curve (AUC) calculation and receiver operating characteristic (ROC), the decision tree (DT) model which use 34 features prove it is the most successful predicted model in oral squamous cell carcinoma (OSCC) tumor by achieving 41% sensitivity, 70% accuracy and 84% of high specificity[12].

Hiroshi Kanegae, et al., in 2020, design a prediction system used machine learning techniques new risk prediction development and validation, the dataset were collected from japan employees from an employee's annual healthy checkup between (2005-2016)for a 18258 individuals.

The researchers in this paper used two machine learning methods XBoost and ensemble, and used one traditional method the logistic regression, the XBoost model was the best predicted model of the systolic blood pressure. The researchers developed a high accuracy future hypertension predictive model using a machine learning methods, this will help in identify the risk individuals and non-pharmacological intervention to prevent any hypertension future development, this machine learning methods has better performance and a high accuracy in both classification and prediction[24].

David G. Kiely, et al., in 2019, design an approach of a productive model by utilization of the healthcare resources and identify the patient at high risk of this disease. The National Health Service in England provides hospital episode statistics to give full cover for the patients' data, in addition to linking the national pulmonary hypertension service in Sheffield. To develop the predictive model a supervised machine learning algorithm (gradient boosting trees) has been used, where the successive tree aims to reduce the error of the previous tree by using boosting. This model provide a low cost screening that facility the early diagnosis and to improve the diagnostic rates and that will enhance the patient outcomes[25].

PAPASTEFANOU, I., et al., in 2020, established a method for predicting women's risk of having SGA newborns, a method was created to predict women's risk of having a newborn with SGA, based not only on maternal medical history and demographics, but using a logistic regression model that uses the combination of maternal factor and vital signs in the first trimester of pregnancy. This method allows a good prediction of patient-specific risks for different pre-specified gestational lengths and birth weight.

A dataset of 124,443 women underwent routine ultrasound screening for singleton pregnancies, and the proposed model considers a priori risk of SGA using Bayes' theorem to combine biomarkers and maternal factors to develop new methods for predicting SGA risk[26].

Researchers have viewed machine learning algorithms[3] to predict different diseases[17], such as stepwise logistic regression analysis, joint Gaussian distribution, linear regression (LR), support vector machine (SVM), decision tree (DT) and k-nearest neighbors (KNN) models, extreme Gradient Boosting (XGBoost) and logistic regression models, to predict different diseases with a variant degree of accuracy and sensitivity, see table (1), to early prediction and treatment of many musculoskeletal disorders (Symptomatic rotator cuff calcific tendinitis), neonates prediction of the natural history of oral squamous cell carcinoma (OSCC) tumor, hypertension prediction, idiopathic pulmonary arterial hypertension, diabetes disease in addition to predict the COVID-19 pneumonia disease[15] [27], diseases prediction fields have been developed in many ways by using machine learning technology especially gradient boosting trees that had an accuracy reach to 99.99% specificity, 14.10% sensitivity in predicting idiopathic pulmonary arterial hypertension[25], in addition to that using machine learning predictive models like support vector machines and logistic regression were the most commonly used in clustering between many others models. These models are highly applicable in prediction chronic diseases diabetes[28], and by using Deep neural network with stacked auto-encoders in determining chemotherapy response achieve best result with 97%, 90%

and 98% in accuracy, sensitivity and specificity value [29].

Machine learning models have emerged as powerful tools in disease prediction, revolutionizing healthcare practices. These models can analyze vast and complex datasets, including patient records, genetic information, lifestyle factors, and environmental data, to identify patterns and risk factors associated with various diseases. By processing this information, machine learning algorithms can provide accurate predictions about the likelihood of disease onset or progression, enabling healthcare providers to take proactive measures. Whether it predicts the risk of heart disease, cancer, diabetes, or infectious outbreaks, these models offer valuable insights that help healthcare professionals tailor preventive strategies and personalized treatment plans, ultimately improving patient outcomes and the overall efficiency of healthcare delivery.

V. Deep Learning Approaches in Healthcare

Deep learning, a subset of machine learning, plays a pivotal role in healthcare by harnessing the power of neural networks to extract intricate patterns and insights from complex medical data. Its ability to automatically learn hierarchical representations from vast datasets has transformed various aspects of the healthcare industry. Deep learning models excel in medical image analysis, enabling the accurate detection and diagnosis of conditions like cancer and radiological abnormalities through techniques such as convolutional neural networks (CNNs). Moreover, recurrent neural networks (RNNs) and long short-term memory (LSTM) networks are used for time-series data analysis, making them invaluable in patient monitoring and predicting disease progression. The utilization of deep learning has also significantly advanced natural language processing (NLP) tasks, such as extracting critical information from electronic health records and improving clinical decision support systems. As deep learning continues to evolve, it promises to enhance diagnostics, treatment personalization, and overall healthcare outcomes, making it a transformative force in the medical field. There are many researches discussed diseases using various deep learning techniques to predict health risks or health, like Baha Ihnaini, et al., in 2021, proposed a smart healthcare recommendation application to early diabetes prediction based on data fusion perspectives and deep machine learning, by the aid of the data fusion the irrelevant burden of the system have been eliminated and improve the data collected from EHRs and they use machine learning-based diabetes prediction. And by using of features dataset with 2768 cases and by use eight features for each one of them, by using deep machine learning model, the system performance to predict the diabetes disease more accurately that reaches to 99.6%[34].

Monica Krishnamoorth, et al., in 2021, used the ICD coded (electronic medical record) which are gathered between 2018-2019 in Africa, and by using convolution neural network (CNN) it will be possible to 12 diseases risk prediction, the researchers first transfer the patient EMR to multi step series data and they used a software framework (OPTUNA) to automating the data and find the hyper parameters which are number of neurons, activation function, batch size and learning rate. This paper accomplished an accuracy of 80.73 % in 12 classes of high risk diseases predicting, like certain infections and parasitic diseases, diseases of the ear and mastoid process, diseases of the circulatory systems, diseases of the blood and metabolic disorders and blood forming nutritional, mental and behavioral and diseases of the nervous system diseases of the eye, adnexa [35].

Jiaheng xie, et al., in 2021, designed an approach to predict the hospital readmissions to many diseases, by using two trajectories the first one based on representation and the second on based on long short term memory (LSTM) depending on medical history of the patients. The researchers made an accurate framework to predict the patient readmission so it will be possible to make a priority to high risk patients to negative consequences reduce, the papers result have been tested and evaluated and it show a high performance and reliability by using recalls, the F1 scores and precisions[36].

Papastefanou, I. et al., in 2021, developed a model that first to predict the patient risk for SGA neonates, based on the maternal history and characteristics. two characteristic have been studied the patient birth weight and gestational age at the time of delivery and second to compare the performance of the predicted model with other previous models. The data were collected from the king's College Hospital, London and Medway Maritime Hospital, Gill Ingham, UK, between 2006 and 2016, at (11-13) weeks gestation, where the medical history and the maternal characteristics have been recorded and combined screening for aneuploidies has been performed. The paper approach used the birth weight and the gestational age in a single continuous model, which used to provide a personalized Gaussian distribution. the proposed model was fitted by using Markov chain Monte Carlo techniques[26].

Aggarwal Y, et al., in 2020, try to recognize a diabetes disease, because diabetes mellitus may cause stroke, myocardial infarction, atherosclerosis and other complications. It features of undirected cause of nonlinear heart rate variability (HRV) parameters which are used in this paper to diagnose of diabetes by two machine learning tools which are artificial neural network (ANN) and support vector machine (SVM).

The datasets collected from 526 digital lead-I electrocardiogram was recorded from male of 10–12 week and from Streptozotocin induced diabetic rats. the accuracy of using ANN is about 86% but it reaches to 90% by using SVM, so it passable to early predict the diabetes disease by using the HVR with the machine learning tools and develop an early diabetes pre diagnoses prediction system[37].

Nai-Hua Lai, et al., in 2020, start to building a model to controlling the ATDH, where an ATDH is the early prevention by high risk discrimination. The ATDH is associated with age, sex, hepatitis A, B and C, alcohol consumption and nutritional status. The hepatotoxicity (ATDH) that induced by anti-tuberculosis (TB) drugs causes a serious and unpredicted reactions result, the patient usually requires to stop taking the TB until the symptoms are disappear. Stopping the TB may cause extensive drug resistant (XDR) or multiple drug resistant (MDR).

Machine learning technology used by the researcher to detect diagnosis and prognosis this disease and its treatment strategies. Three different machine learning technology used in this research to predict ATDH and a comparison between them to determine the accurate method by using the TB patients clinical data and patients genomic data, to determine the best predictive model of the ATDH disease. The three machine learning methods are artificial neural network support vector machine and random forest, which are used to predict the ATDH, the ANN which are combined with genetic and non-genetic risk factors is the best model in this prediction[27].

Horvath, L. ,et al., in 2020, Developed a model to detect a mycobacterium infection, and the manual microscopy is considered as the first detection manner, but is causes a workload on the laboratory staff, so combining the automated microscopy method with the artificial intelligent tools to design and implement an application that can reduce the staff workload and increase the diagnostic result quality. The data set have been collected in 2019 from university hospital Heidelberg, as 6500 sample. By using neural network and special hardware devices that can be do both scan and analysis in a parallel way the time that spend to detect the diseases is reduce from 5-15 minute in the manual way to 2.5 minute by the aid of artificial intelligent.

In addition to the reduce the manual slide handling time the combing of human review and automated analysis and classification can improve the detection sensitivity and that can be used to correct the given DNN classified prediction, so it seems the most promising reliable diagnosis way[11].

Wenxing Hong, et al, in 2019, combined the technology of deep learning with the health care big data analysis to pertinence improvement and the comprehensiveness of the medical examination which used to inform the patients with the potential diseases to target them by doing medical examinations. An algorithm name medical history based on the potential predictions which are a novel deep learning, it is a hyped algorithm, by using the patient medical history can predict the potential diseases in a more accurate way, to reduce the delay treatments. The algorithm combine the advantage of deep neural networks (DNN) and the Factorization Machine (FM), it considered the low and the high order relations in diseases features to improve the model comprehensiveness. The proposed algorithm implemented on real data base to the potential phenotype based prediction by observed the gene phenotypes[38].

Kadam Vinay R,, et al., in 2019, design a new model for predicting three diseases heart, diabetic and kidney has been designed by using artificial neural network and probabilistic modeling. One proper dataset has been built by using three different resources Pime datasets, UCI repository and Kaggle datasets, the dataset are split into the training dataset and the testing dataset. The probabilistic model and deep learning technique used in the diseases prediction, the ANN used in prediction to get a prediction accuracy for 95% for heart disease and 98% for kidney disease and 72% for diabetes disease and that will make a big effect in reduces the cost of treatment and the risk on patient life[39].

Syed Nawaz, et al., in 2020, have been viewed that cardiovascular and coronary artery disease is one of the most serious causes of death in India and the whole world, the researchers analyzed several attributes related to this disease such as age, gender, cholesterol and blood pressure from the Kaggle data set.

Using three different machine learning techniques, K-Nearest Neighborhood (KNN), support vector machine (SVM) and decision trees (DT) did not provide accurate performance due to the large data set. The researchers used an artificial neural network (ANN) and flow tensor keras to improve the accuracy of disease prediction, recall and precision used to evaluate the techniques used. Accuracy of SVM, KNN, DT, and ANN are 81.97, 67.2, 81.97, and 85.24. Therefore the artificial neural network has a higher level of accuracy in predicting a heart disease than other algorithms in the Kaggle dataset[40].

The choice of the best deep learning technique or model for disease prediction in healthcare depends on the nature of the data and the specific disease being targeted. However, several approaches have demonstrated exceptional performance. Convolutional Neural Networks (CNNs) are widely regarded as effective for image-based disease prediction, such as identifying tumors in medical images like CT scans and mammograms. Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks, excel in handling sequential data, making them valuable for time-series predictions like disease progression or patient monitoring. Transformer-based models, originally designed for natural language processing, have shown great promise in analyzing electronic health records and unstructured clinical notes. Ensemble methods, which combine multiple deep learning models, can further enhance prediction accuracy by leveraging the strengths of various architectures. Transfer learning, where pre-trained models are fine-tuned on healthcare data, has also emerged as a powerful strategy for disease prediction when labeled data is limited. Ultimately, the optimal choice of technique or model depends on the specific healthcare task, the quality and quantity of available data, and the computational resources at hand.

Many researches used deep learning algorithms to predict different diseases, see table (2), such as predicting diabetes prediction many classes of high risk diseases, diseases of the blood and blood forming nutritional and metabolic disorders, mental and behavioral, the patient readmission, recognize a diabetes, early prevention by high risk discrimination, detect a mycobacterium infection[32], the potential phenotype based by observed the gene phenotypes, heart, diabetic and kidney cardiovascular and coronary artery disease, early prediction of the hospital length of stay, lung cancer, the Risk of depression, chronic diseases, skin disease and mortality in patients in the ICU[3].

VI. Results and Discussion

The paper have been reviewed many artificial learning models deals with health care and patient disease detection and prediction, choice between machine learning and deep learning algorithms for disease prediction depends on various factors, including the complexity of the task and the available data. In general, deep learning algorithms, particularly deep neural networks, have shown remarkable success in handling complex and high-dimensional data, making them well- suited for tasks like image analysis and natural language processing in healthcare. Deep learning excels in tasks such as medical image classification and interpretation, where intricate patterns are crucial for accurate diagnosis, it reach to accuracy about 99.6% in diabetes prediction and 97.24% in monitoring heartbeat rates. On the other hand, traditional machine learning algorithms, like support vector machines (SVMs) and random forests, can be effective in dealing with structured data and when interpretability and feature selection are essential it reaches to 99% accuracy and 99.99% specificity, 14.10% sensitivity in Idiopathic pulmonary arterial hypertension prediction. The choice often involves a trade-off between model complexity, data availability, computational resources, and the need for interpretable results. In practice, a combination of both machine learning and deep learning techniques, known as hybrid models, can yield the best results by leveraging the strengths of each approach for comprehensive disease prediction in healthcare.

VII. Conclusion

In conclusion, choosing artificial intelligent model to deal with health risk disease prediction depending on patient medical history is a an important issue, if it was machine learning algorithms or deep learning algorithms, it does not mean that one is completely better than the other, but rather the appropriate tool between them should be chosen for the specific healthcare task. Deep learning algorithms shine for their ability to handle unstructured and complex data, and in tasks such as natural language processing and medical image analysis, providing remarkable accuracy in disease diagnosis and prognosis. In contrast, the use of machine learning algorithms remains valuable, especially when the data is structured and feature selection and interpretability are critical. Choosing the best approach involves a combination of machine learning and deep learning techniques, using hybrid models that leverage the strengths of each, to simultaneously achieve accurate and comprehensive disease prediction while ensuring interpretability and resource efficiency. Ultimately, in the dynamic landscape of healthcare, choosing the right and most promising algorithm must align with the available data and the unique requirements of the healthcare problem and foster a collaborative, data-driven approach to improve patient care and outcomes.

1. Acknowledgments

The authors would like to express their appreciation to all the reviewers who will spend their effort and valuable time to review the manuscript, and we appreciate their valuable comments and their suggestions, which will help us in improving the quality of the manuscript.

Table 1. Research that used machine learning.

Index	First author	Date	Type of disease	Dataset	Technologies used	Accuracy
1	Shengtao Dong (Dong *et al.*, 2022)	2022	early prediction and treatment of many musculoskeletal disorders (Symptomatic rotator cuff calcific tendinitis)	clinical data hospital have been collected between 2008 to 2020, and shoulder symptoms questionnaire	stepwise logistic regression analysis	52%–79%
2	PAPASTEFANOU, I. (Papastefanou, Wright and Nicolaides, 2020)	2021	SGA neonates	from the king's College Hospital in London and Medway Maritime Hospital (2006-2016)	joint Gaussian distribution	Reach to 75.5%
3	Chui S Chu (Chu *et al.*, 2020)	2020	the natural history of oral squamous cell carcinoma tumor prediction	467 OSCC patients treated over a 19-year period to construction of a clinico-athological database	linear regression, support vector machine, decision tree and k-nearest neighbors models	70%
4	Hiroshi Kanegae(Kanegae *et al.*, 2020)	2020	Hypertension prediction	employee's annual healthy checkup between (2005-2016)for a 18258 individuals	extreme gradient boosting and logistic regression models	0.976 in the derivation and 0.876 in validation sets
5	David G. Kiely(Kiely *et al.*, 2019).	2019	Idiopathic pulmonary arterial hypertension	National Health Service in England provides hospital	gradient boosting trees	99.99% specificity, 14.10% sensitivity
6	Jobeda Jamal Khanam(Khanam and Foo, 2021)	2021	Diabetes disease	Pima Indian Diabetes (PID) dataset	Naïve bayes, support vector machines, linear regression, random forest, K-nearest neighbor, decision tree, and neural network	88.6% for NN
7	abnormality detection in medical images(Chest X-ray images (CXR)) (Chandra *et al.*, 2020)	2020	Automatic detection in medical images(Chest X-ray images (CXR))	Montgomery set (Candemir et al., 2014Jaeger et al., 2014), and Shenzhen set (Candemir et al., 2014; Jaeger et al., 2014).	SVM classifier	95.60 ± 5.07%
8	Walaa Gouda(Gouda and Yasin, 2020)	2020	predict the COVID-19 pneumonia disease severity	SPSS Inc. Released 2015. IBM SPSS statistics for	AI-Rad Companion	81.8% sensitivity and 81.9% specificity

				windows, version 23.0,Armonk, NY: IBM Corp.		
9	Gopi BattinFeni (Battineni *et al.*, 2020)	2020	Chronic Disease Diagnosis	conducted in January 2020 and resulted in 453 documents	Supervised machine learning SVM, logistic regression	99%
10	Muhammad Arsalan(Arsalan *et al.*, 2019)	2019	Diabetic and Hypertensive Retinopathy	40 RGB fundus camera images	dual-stream feature network (DS-Net)	91%

Table 2. Research that used deep learning.

Index	First author	Date	Type of disease	Dataset	Technologies used	Accuracy
1	Baha Ihnaini(Ihnaini *et al.*, 2021).	2022	diabetes prediction	eight features of 2768 cases	deep machine learning model	99.6%
2	Monica Krishnamoorth(Krishnamoorthy *et al.*, 2021).	2021	predicting 12 class of high risk diseases of blood forming nutritional and metabolic disorders, mental and behavioral	EMR gathered in Africa between (2018-2019)	convolution neural network (CNN)	80.73 %
3	JIAHENG XIE(Xie *et al.*, 2022).	2021	predict the patient readmission	United States Department of Health and Human Services (HHS)	long short term memory (LSTM)	88.4%
4	Aggarwal(Aggarwal *et al.*, 2020).	2020	recognize a Diabetes	dataset_E.csv	artificial neural network (ANN) and support vector machine (SVM)	86% by using ANN 90% by using SVM
5	Nai-Hua Lai (Lai *et al.*, 2020).	2020	early prevention by high risk discrimination	Wan Fang Hospital	artificial neural network(ANN), support vector machine(SVM) and random forest(RF)	88.0% for ANN 78% for SVM 73% for RF
6	Horvath, L(Horvath *et al.*, 2020).	2020	detect a mycobacterium infection	university hospital Heidelberg, as 6500 sample	Deep neural networks (DNN)	97.4%,
7	WENXING HONG(Hong *et al.*, 2019).	2019	potential phenotype prediction based on observed the gene phenotypes	real-world datasets	deep neural networks (DNN), factorization machine(FM)	improve the prediction accuracy
3	Kadam Vinay R.(Kadam Vinay R, K.L.S.Soujanya and Abstract—, no date).	2019	predicting heart, diabetic and kidney	resources Pime datasets, UCI repository and kaggle datasets	artificial neural network and probabilistic modeling	95% for heart disease 98% for kidney

						disease 72% for diabetes
4	Syed Nawaz Pasha(Pasha *et al.*, 2020).	2020	Cardiovascular and coronary artery disease	kaggle dataset	support vector machines, k-nearest neighbor, decision trees, artificial neural network	85.24%
5	G. Damiani(Damiani *et al.*, 2020)	2020	Prediction in Squamous Cell Carcinoma of the Scalp Treated	Photoradiotheray Unit, Policlinico Milan, Italy	artificial neural networks analysis(ANNsA)	85.7%
6	Zhixu Hu(Hu *et al.*, 2022)	2022	prediction of the hospital length of stay	hospital discharge records for the urban areas of Chengdu, China, with 10.7 million records from 678 hospitals between (2015-2019)	decision Tree, linear support vector machine and deep neural network	96%
7	Marvin Chia-Han Yeh(Marvin Chia-Han Yeh1, 2, MD, PhD; Yu-Hsiang Wang3, MD; Hsuan-Chia Yang4, 5, PhD; Kuan-Jen Bai6, 7, 8, MD; Hsiao-Han Wang1, 2, 4, 9*, MD; Yu-Chuan (Jack) Li1, 2, 4, 5, 9*, MD, 2021)	2021	Lung cancer	electronic medical records(EMR)	convolution neural network (CNN)	93%
8	Sumaiya Tarannum Noor(Noor, Sumaiya Tarannum, Syeda Tasmiah Asad, 2021)	2021	the Risk of Depression	heartbeat rates in a timeseries manner, 5000 patients that contain 140 timesteps	Recurrent Neural Networks (RNN) and long short term memory (LSTM)	97.24%
9	Ramesh Nadarajah(Nadarajah *et al.*, 2021)	2021	Chronic diseases	Clinical Practice Research Datalink-GOLD (CPRD-GOLD) dataset	Adapted convolution neural network, Recurrent Neural Networks (RNN) and Transformer	at least 5% compared with other models
10	Parvathaneni Naga Srinivasu(Srinivasu *et al.*, 2021)	2021	Skin Disease	Kaggle dataset	Deep Learning Neural Networks (DNN) with Mobile Net V2 and LSTM	85.34**%**
11	Hans-Christian Thorsen-Meyer(Hans-Christian Thorsen-Meyer, Annelaura B Nielsen, Anna P Nielsen, Benjamin Skov Kaas-Hansen, Palle Toft, Jens Schierbeck and Piotr J Chmura, Marc Heimann, Lars Dybdahl, Lasse Spangsege, Patrick Hulsen, Kirstine Belling, Søren Brunak, 2020)	2020	mortality in patients in the ICU	data from patients in four ICUs in Denmark, between 2011 and 2016	Recurrent Neural Networks (RNN) consisting of long short term memory	The predictive performance improved over the Time course of an ICU stay

References

[1] Steinfeldt, J. et al. Medical history predicts phenome-wide disease onset. doi:10.1101/2023.03.10.23286918.

[2] Lasantha, D., Vidanagamachchi, S. & Nallaperuma, S. Deep learning and ensemble deep learning for circRNA-RBP interaction prediction in the last decade: A review. Engineering Applications of Artificial Intelligence vol. 123 at https://doi.org/10.1016/j.engappai.2023.106352 (2023).

[3] Janghel, R. R. & Rathore, Y. K. Deep Convolution Neural Network Based System for Early Diagnosis of Alzheimer's Disease. IRBM 42, 258–267 (2021).

[4] Damiani, G. et al. Artificial neural networks allow response prediction in squamous cell carcinoma of the scalp treated with radiotherapy. J. Eur. Acad. Dermatology Venereol. 34, 1369–1373 (2020).

[5] Ristevski, B. & Chen, M. Big Data Analytics in Medicine and Healthcare. J. Integr. Bioinform. 15, (2018).

[6] Diwakar, M. et al. Latest trends on heart disease prediction using machine learning and image fusion. Mater. Today Proc. 37, 3213–3218 (2020).

[7] Javaid, M., Haleem, A., Pratap Singh, R., Suman, R. & Rab, S. Significance of machine learning in healthcare: Features, pillars and applications. Int. J. Intell. Networks 3, 58–73 (2022).

[8] Amalahu, V. O. MASTER THESIS for MSc degree in the specialty 121 Software Engineering Title: SOFTWARE METHOD FOR DETECTING DISEASES BASED ON PATIENTS' HETEROGENEOUS DATA SOFTWARE METHOD FOR DETECTING DISEASES BASED ON PATIENTS' HETEROGENEOUS DATA View project. doi:10.13140/RG.2.2.20920.26889.

[9] Noor, Sumaiya Tarannum, Syeda Tasmiah Asad, M. M. K. Predicting the Risk of Depression Based on ECG Using RNN. Comput. Intell. Neurosci. 2021, (2021).

[10] Fatima, M. & Pasha, M. Survey of Machine Learning Algorithms for Disease Diagnostic. J. Intell. Learn. Syst. Appl. 09, 1–16 (2017).

[11] Horvath, L. et al. Machine-assisted interpretation of auramine stains substantially increases through-put and sensitivity of microscopic tuberculosis diagnosis. Tuberculosis 125, (2020).

[12] Ferizi, U. et al. Artificial Intelligence Applied to Osteoporosis: A Performance Comparison of Machine Learning Algorithms in Predicting Fragility Fractures From MRI Data. J. Magn. Reson. Imaging 49, 1029–1038 (2019).

[13] Chu, C. S., Lee, N. P., Adeoye, J., Thomson, P. & Choi, S. W. Machine learning and treatment outcome prediction for oral cancer. J. Oral Pathol. Med. 49, 977–985 (2020).

[14] Khan, A. & Zubair, S. An Improved Multi-Modal based Machine Learning Approach for the Prognosis of Alzheimer's disease. J. King Saud Univ. - Comput. Inf. Sci. 34, 2688–2706 (2022).

[15] Gouda, W. & Yasin, R. COVID-19 disease: CT Pneumonia Analysis prototype by using artificial intelligence, predicting the disease severity. Egypt. J. Radiol. Nucl. Med. 51, (2020).

[16] Saha, A., Tso, S., Rabski, J., Sadeghian, A. & Cusimano, M. D. Machine learning applications in imaging analysis for patients with pituitary tumors: a review of the current literature and future directions. Pituitary 23, 273–293 (2020).

[17] Garapati, S. L. & Garapati, S. Application of Big Data Analytics: An Innovation in Health Care. International Journal of Computational Intelligence Research vol. 14 http://www.ripublication.com (2018).

[18] Dong, S. et al. Risk Factor Analysis for Predicting the Onset of Rotator Cuff Calcific Tendinitis Based on Artificial Intelligence. Comput. Intell. Neurosci. 2022, (2022).

[19] Ihnaini, B. et al. A Smart Healthcare Recommendation System for Multidisciplinary Diabetes Patients with Data Fusion Based on Deep Ensemble Learning. Comput. Intell. Neurosci. 2021, (2021).

[20] Krishnamoorthy, M., Hameed, M. S. A., Kopinski, T. & Schwung, A. Disease Prediction Based on Individual's Medical History Using CNN. in Proceedings - 20th IEEE International Conference on Machine Learning and Applications, ICMLA 2021 89–94 (Institute of Electrical and Electronics Engineers Inc., 2021). doi:10.1109/ICMLA52953.2021.00022.

[21] Xie, J., Zhang, B., Ma, J., Zeng, D. & Lo-Ciganic, J. Readmission Prediction for Patients with Heterogeneous Medical History: A Trajectory-Based Deep Learning Approach. ACM Trans. Manag. Inf. Syst. 13, (2022).

[22] Papastefanou, I., Wright, D. & Nicolaides, K. H. Competing-risks model for prediction of small-for-gestational-age neonate from maternal characteristics and medical history. Ultrasound Obstet. Gynecol. 56, 196–205 (2020).

[23] Aggarwal, Y., Das, J., Mazumder, P. M., Kumar, R. & Sinha, R. K. Heart rate variability features from nonlinear cardiac dynamics in identification of diabetes using artificial neural network and support vector machine. Biocybern. Biomed. Eng. 40, 1002–1009 (2020).

[24] Lai, N. H. et al. Comparison of the predictive outcomes for anti-tuberculosis drug-induced hepatotoxicity by different machine learning techniques. Comput. Methods Programs Biomed. 188, (2020).

[25] Hong, W., Xiong, Z., Zheng, N. & Weng, Y. A Medical-History-Based Potential Disease Prediction Algorithm. IEEE Access 7, 131094–131101 (2019).

[26] Kanegae, H. et al. Highly precise risk prediction model for new-onset hypertension using artificial intelligence techniques. J. Clin. Hypertens. 22, 445–450 (2020).

[27] Kiely, D. G. et al. Utilising artificial intelligence to determine patients at risk of a rare disease: idiopathic pulmonary arterial hypertension. Pulm. Circ. 9, (2019).

[28] Kadam Vinay R, K.L.S.Soujanya, P. S. Disease Prediction by Using Deep Learning Based on Patient Treatment History 746.

[29] Pasha, S. N., Ramesh, D., Mohmmad, S., Harshavardhan, A. & Shabana. Cardiovascular disease prediction using deep learning techniques. in IOP Conference Series: Materials Science and Engineering vol. 981 (IOP Publishing Ltd, 2020).

[30] Chandra, T. B., Verma, K., Singh, B. K., Jain, D. & Netam, S. S. Automatic detection of tuberculosis related abnormalities in Chest X-ray images using hierarchical feature extraction scheme. Expert Syst. Appl. 158, (2020).

[31] Zhou, X. et al. Deep learning-based polygenic risk analysis for Alzheimer's disease prediction. Commun. Med. 3, 49 (2023).

[32] Papastefanou, I., Wright, D. & Nicolaides, K. H. Competing-risks model for prediction of small-for-gestational-age neonate from maternal characteristics and medical history. Ultrasound Obstet. Gynecol. 56, 196–205 (2020).

[33] Khanam, J. J. & Foo, S. Y. A comparison of machine learning algorithms for diabetes prediction. ICT Express 7, 432–439 (2021).

[34] Battineni, G., Sagaro, G. G., Chinatalapudi, N. & Amenta, F. Applications of machine learning predictive models in the chronic disease diagnosis. J. Pers. Med. 10, (2020).

[35] Arsalan, M., Owais, M., Mahmood, T., Cho, S. W. & Park, K. R. Aiding the diagnosis of diabetic and hypertensive retinopathy using artificial intelligence-based semantic segmentation. J. Clin. Med. 8, (2019).

[36] Kadam Vinay R, K.L.S.Soujanya, P. S. & Abstract—. Disease Prediction by Using Deep Learning Based on Patient Treatment History 746.

[37] Hu, Z., Qiu, H., Wang, L. & Shen, M. Network analytics and machine learning for predicting length of stay in elderly patients with chronic diseases at point of admission. BMC Med. Inform. Decis. Mak. 22, 1–15 (2022).

[38] Marvin Chia-Han Yeh1, 2, MD, PhD; Yu-Hsiang Wang3, MD; Hsuan-Chia Yang4, 5, PhD; Kuan-Jen Bai6, 7, 8, MD; Hsiao-Han Wang1, 2, 4, 9*, MD; Yu-Chuan (Jack) Li1, 2, 4, 5, 9*, MD, P. Artificial intelligence⇓based prediction of lung cancer risk using nonimaging electronic medical records: Deep learning approach. J. Med. Internet Res. 23, 1–13 (2021).

[39] Nadarajah, R. et al. Predicting patient-level new-onset atrial fibrillation from population-based nationwide electronic health records: Protocol of FIND-AF for developing a precision medicine prediction model using artificial intelligence. BMJ Open 11, (2021).

[40] Srinivasu, P. N. et al. Classification of skin disease using deep learning neural networks with mobilenet v2 and lstm. Sensors 21,

(2021).
[41] Hans-Christian Thorsen-Meyer, Annelaura B Nielsen, Anna P Nielsen, Benjamin Skov Kaas-Hansen, Palle Toft, Jens Schierbeck, T. S. & Piotr J Chmura, Marc Heimann, Lars Dybdahl, Lasse Spangsege, Patrick Hulsen, Kirstine Belling, Søren Brunak, A. P. Exploiting temporal relationships in the prediction of mortality. The Lancet Digital Health vol. 2 e152–e153 at https://doi.org/10.1016/S2589-7500(20)30056-X (2020).
[42] Biswas, M. et al. Two-stage artificial intelligence model for jointly measurement of atherosclerotic wall thickness and plaque burden in carotid ultrasound: A screening tool for cardiovascular/stroke risk assessment. Comput. Biol. Med. 123, (2020).
[43] Davis, D. A., Chawla, N. V., Blumm, N., Barabási, A. L. & Christakis, N. Predicting individual disease risk based on medical history. in International Conference on Information and Knowledge Management, Proceedings 769–778 (2008). doi:10.1145/1458082.1458185.
[44] Dicuonzo, G., Galeone, G., Shini, M. & Massari, A. Towards the Use of Big Data in Healthcare: A Literature Review. Healthcare (Switzerland) vol. 10 at https://doi.org/10.3390/healthcare10071232 (2022).
[45] Mokal, S. R., Chorghe, D. R., Madhavi, N. J., Kelaskar, P. V. & Jadhav, P. S. N. Disease Prediction using Python. Int. J. Res. Appl. Sci. Eng. Technol. 11, 2082–2086 (2023).
[46] Wang J, Hu X. (2024). Healthcare Predictive Analytics Using Machine Learning and Deep Learning Techniques: A Survey. Journal of Electrical Systems and Information Technology.
[47] Smith A, Liu B, Zhao C. (2024). Artificial Intelligence in Disease Diagnosis: A Systematic Review. Springer.
[48] Kim S, Park J, Lee H. (2024). Wearable Smart Sensor System for Monitoring and Intelligent Prediction of Health Risks. IEEE Xplore