



## Delving into Machine Learning's Influence on Disease Diagnosis and Prediction

Basu Dev Shivahare<sup>1</sup>, Jagendra Singh<sup>2</sup>, Vinayakumar Ravi<sup>6</sup>, Radha Raman Chandan<sup>3</sup>, Tahani Jaser Alahmadi<sup>7\*</sup>, Prabhishek Singh<sup>2</sup> and Manoj Diwakar<sup>4,5</sup>

<sup>1</sup>Department of CSE, Galgotias University, Greater Noida, India

<sup>2</sup>School of Computer Science Engineering & Technology, Bennett University, Greater Noida, India

<sup>3</sup>Department of CS, School of Management Science, Varanasi, Uttar Pradesh, India

<sup>4</sup>Department of CSE, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India

<sup>5</sup>Graphic Era Hill University, Dehradun, India

<sup>6</sup>Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia

<sup>7</sup>Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

### Abstract:

**Introduction/ Background:** Medical diagnoses have increasingly depended on digitized images obtained through cutting-edge technology. These algorithms offer a promising avenue to transform diagnostic processes in healthcare, with their application scope continually widening due to ongoing advancements. This paper explores machine learning's role in clinical analysis and prediction, examining various studies that apply these techniques in clinical diagnosis, focusing on their use in analyzing images and providing accurate diagnoses.

**Materials and Methods:** This study employs a comparative analysis approach, utilizing diverse machine learning algorithms like SVM, K-nearest neighbors, Random Forests, and Decision Trees to analyze digitized medical images and patient records. We extracted data from several medical databases, ensuring a varied and comprehensive dataset. We also evaluated the impact of different data characteristics on the algorithms' effectiveness, aiming to understand the variability in their diagnostic precision across various conditions.

**Results:** The results indicate that machine learning algorithms, particularly SVM, K-nearest neighbors, Random Forests, and Decision Trees, demonstrate significant accuracy in diagnosing diseases from digitized images and medical records. SVM and Random Forests showed particularly high effectiveness in clinical diagnosis, suggesting their robustness across different medical conditions and datasets. These findings underscore the potential of machine learning to enhance diagnostic precision and predict illnesses early, aligning with the growing trend of technology-driven medical diagnostics.

**Discussion:** The findings reinforce the pivotal role of machine learning in transforming medical diagnostics. The variability in algorithm performance highlights the necessity for tailored approaches, considering dataset specifics and the medical condition being diagnosed. This study underscores the potential of machine learning to enhance diagnostic accuracy, yet it also emphasizes the need for continuous refinement and understanding of the underlying factors affecting algorithm performance. Future research should focus on optimizing these algorithms within diverse clinical settings to fully harness their diagnostic capabilities.

**Conclusion:** This study highlights the transformative potential of machine learning in medical diagnostics, demonstrating how various algorithms can effectively analyze digitized images and patient records to diagnose diseases. While the performance of these algorithms varies based on dataset characteristics, the overall high accuracy underscores machine learning's promise in healthcare. As the field continues to evolve, machine learning is poised to become an integral part of clinical diagnosis, enhancing the accuracy and efficiency of medical evaluations and treatments.

**Keywords:** Machine learning, Algorithms, Disease diagnosis, Image-based disorder, Human and disease, Health system.

© 2024 The Author(s). Published by Bentham Open.

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: <https://creativecommons.org/licenses/by/4.0/legalcode>. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



Received: January 03, 2024

Revised: March 10, 2024

Accepted: March 20, 2024

Published: ?? ??, 2024



Send Orders for Reprints to  
[reprints@benthamscience.net](mailto:reprints@benthamscience.net)

\*Address correspondence to this author at the Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; E-mail: [tjalahmadi@pnu.edu.sa](mailto:tjalahmadi@pnu.edu.sa)

Cite as: Shivahare B, Singh J, Ravi V, Chandan R, Alahmadi T, Singh P, Diwakar M. Delving into Machine Learning's Influence on Disease Diagnosis and Prediction. *Open Public Health J*, 2024; 17: e18749445297804. <http://dx.doi.org/10.2174/0118749445297804240401061128>

## 1. INTRODUCTION

The diagnostic process stands as the linchpin in understanding the essential medications integral to a medicine's societal functionality. Within the medical system, diagnosis holds significant sway, orchestrating the management of illnesses by outlining treatment options, forecasting potential outcomes, and serving as an educational cornerstone. Achieving a comprehensive diagnosis becomes pivotal in implementing suitable and effective therapy [1]. While diagnostic testing and imaging advancements have notably streamlined the diagnostic process, the utilization of human scientific judgment remains imperative in ensuring precise diagnoses, particularly in today's rapidly evolving technological landscape [2].

At the core of medical practice, diagnosis functions as the cornerstone for categorizing medications necessary for a medicine's operation within a societal framework. It plays a vital role in guiding treatment strategies, predicting potential outcomes, and establishing a knowledge base. A thorough diagnosis forms the bedrock for implementing appropriate and efficient therapies [3]. Despite significant strides in diagnostic testing and imaging technologies, the integration of human scientific judgment remains indispensable for accurate diagnoses, especially in the contemporary era marked by swift technological advancements [4].

The integration of machine learning in healthcare began in the 1970s, marking a significant milestone with the establishment of the global journal, *Artificial Intelligence in Medicine*, in 1980. Over the following two decades, the field of disease detection witnessed the incorporation of traditional machine learning techniques such as the Support Vector Machine, Naive Bayes, and several artificial neural networks. The emergence of AlexNet in 2012 signified the advent of deep learning, as neural networks demonstrated exceptional performance, ushering in a new era for this domain [5].

In the last decade, there has been a remarkable increase in resources directed towards the development and implementation of artificial intelligence (AI) in the healthcare sector. Referenced studies [6-8] highlight the convergence of AI and machine learning technologies within healthcare, fostering the creation of software, platforms, automated systems, and devices capable of monitoring and enhancing individuals' health status. This

convergence has not only expanded the technological landscape but also transformed the possibilities for healthcare applications, driving advancements in diagnostics, treatment strategies, and personalized health monitoring.

Analysis of clinical data can swiftly identify a condition, expediting the patient's journey toward recovery [9]. Traditional methods of diagnosing illnesses often prove time-consuming, labor-intensive, and costly. Researchers [10] have demonstrated the potential of machine learning-based diagnostic techniques, which offer efficiency in both time and cost. These machine-learning approaches have not only proven successful in identifying common illnesses but have also shown proficiency in diagnosing rare diseases [11]. Evidence presented by authors [12] underscores the significance and robustness of employing AI and ML approaches to address healthcare challenges. Machine Learning (ML), a subset of Artificial Intelligence (AI), utilizes algorithms to learn from data, distinguished by its capacity to autonomously improve performance through modifications during the learning process [13]. ML has showcased its efficacy across various domains, including but not limited to robotics, education, travel, and particularly in healthcare, where its techniques primarily focus on illness diagnosis.

### 1.1. Purpose of the Study

This paper's overarching goal is to talk about the use of machine learning methods and algorithms for clinical diagnosis and prognosis. The authors of this publication provide important insights for those working in the field of disease diagnostics using machine learning. They can use this data to pick the most applicable and cutting-edge machine learning methods. As a result, the diagnostic procedure is more likely to discover and classify diseases efficiently and accurately. This research summarises the state of the art of machine learning algorithms for disease detection and prognosis. This study did a thorough analysis of existing literature about the use of machine learning algorithms in the medical field in the context of disease diagnosis. In this study, we examine the similarities and differences between various algorithms, evaluation techniques, and final results [9].

### 1.2. Milestones of AI

The advent of artificial intelligence (AI) has ushered in a new era of innovation, significantly impacting various sectors, including healthcare. Within this realm, machine

learning (ML), a subset of AI, has become a pivotal tool in revolutionizing disease diagnosis and prediction. This introduction delves into the milestones of AI, acquaints us with its distribution, and outlines the advances and challenges within the context of healthcare. Machine learning's journey in healthcare began as a part of the broader evolution of AI, which traces back to the mid-20th century. The conceptual foundation laid by pioneers like Alan Turing set the stage for AI's development, with the first neural network for computers being created in the 1950s. However, it was not until the advent of more powerful computing resources and the availability of large datasets that machine learning began to significantly influence healthcare [7, 8].

### 1.3. ML's Applications in Healthcare

Today, ML's applications in healthcare are extensive and growing. From predictive analytics that forecast disease outbreaks to algorithms that detect patterns in medical images for early disease diagnosis, machine learning is reshaping how we approach health and medicine. For instance, ML algorithms have been instrumental in diagnosing diseases from various medical imaging techniques, such as MRI and CT scans, by identifying patterns imperceptible to the human eye. The distribution of machine learning in healthcare is global, with both developed and developing countries harnessing its potential. Advanced healthcare systems use machine learning to enhance precision medicine, while in resource-limited settings, it offers tools for improving access to diagnostics and treatment planning [12].

### 1.4. Significant Advances in ML

Significant advances in ML, such as deep learning, have propelled the field forward. Deep learning algorithms, which mimic the human brain's neural networks, have achieved remarkable success in diagnosing

conditions such as diabetic retinopathy and various cancers. These algorithms continually learn from vast amounts of data, improving their diagnostic accuracy over time. However, the integration of machine learning in healthcare is not without challenges. Ethical concerns, such as data privacy and the potential for algorithmic bias, pose significant hurdles. Moreover, the "black box" nature of some ML algorithms, where decision-making processes are not transparent, raises questions about accountability in clinical settings. As we delve into machine learning's influence on disease diagnosis and prediction, it is crucial to recognize its roots in the broader AI landscape, appreciate its global distribution and advances, and address the challenges it faces. By doing so, we can harness the full potential of machine learning to transform healthcare, making it more predictive, personalized, and accessible [13, 14].

## 2. MACHINE LEARNING IN DISEASE DIAGNOSIS

### 2.1. Utilizing Machine Learning for Disease Diagnosis in Clinical Settings

The decision tree approach is often regarded as the most popular data modelling technique [14] due to its simplicity and intuitiveness. Better techniques like support vector machines (SVM) and artificial neural networks (ANN) have become widely used in recent years [15]. Predicting protein function and gene expression using machine learning algorithms is only one of several areas where this technique is being studied [16]. In contrast to methods that rely on the protein's sequence or structure, those that employ machine learning to predict protein function don't require any prior knowledge of homology or parameters derived from homology. Therefore, there has been a rise in studies targeted at creating effective machine-learning strategies for predicting protein function in the context of disease diagnosis.

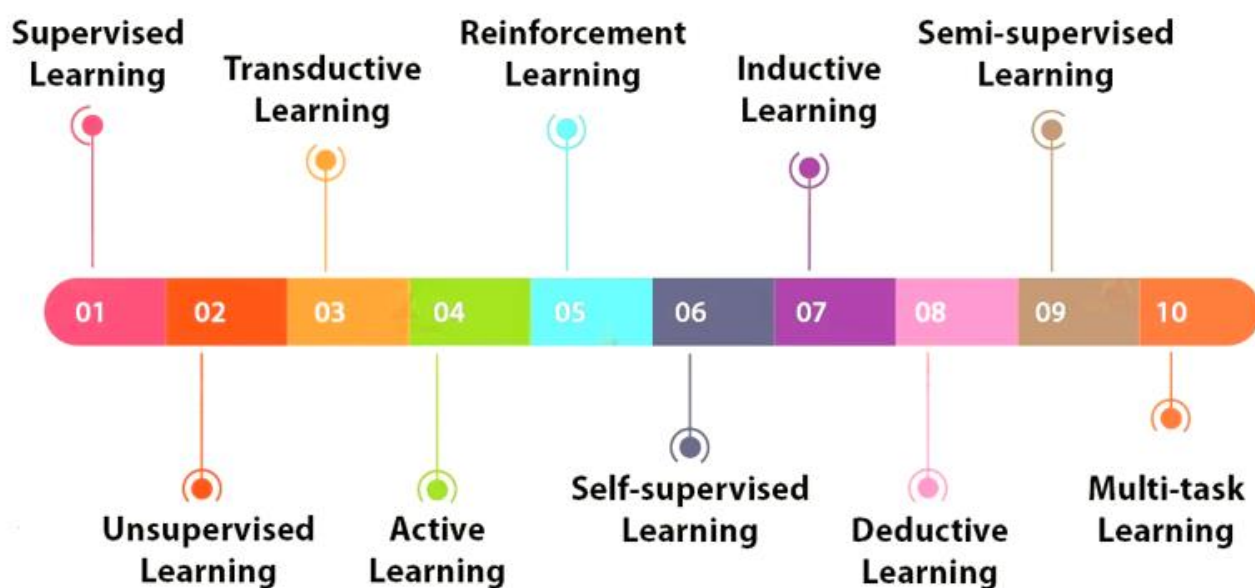


Fig. (1). Various approaches to machine learning.

Fig. (1) illustrates the diverse methodologies within machine learning, showcasing a range of approaches from supervised and unsupervised learning to reinforcement learning, each with distinct mechanisms and applications in data analysis and prediction. This figure provides a visual taxonomy of machine learning techniques, highlighting their unique characteristics and the contexts in which they are optimally employed. The process of medical diagnosis, often considered empirical yet not fully understood as a cognitive task [17], undergoes segmentation when applying computer systems, particularly Machine Learning (ML). It encompasses several stages to streamline the complexity. Commencing with data collection, which spans various formats—clinical data from interviews, demographics, imaging, speech, patient history, and even physiological sounds [18] though this list is not exhaustive. Subsequent stages involve data processing, entailing data preparation tasks like handling missing values, reducing data dimensions, and managing noisy data [19]. Following this, the process moves to identifying the target variable and predictors. Then, models are trained using this data, providing them with the necessary information for subsequent diagnostic applications.

### 2.3. Advantages of Using Machine Learning to Diagnose Disease

The presence of several overlapping structures and instances, as well as distractions, fatigue, and limits in the human visual system, might create challenges. In such situations, the availability of a “second opinion” can be advantageous [20]. The use of computer-aided design (CAD) systems has been fostered in the context of diagnostic procedures. Computer-aided diagnosis (CAD) is a theoretical framework that assigns equal importance to both doctors and computers, with the aim of supporting clinicians in making optimal clinical choices and implementing best practises [21]. Furthermore, the use of electronic health record (EHR) systems has been prevalent in the healthcare industry as a means to aid in clinical decision-making. This is mostly attributed to the rising complexity seen among patients, the occurrence of high diagnostic mistakes, and the abundance of data accessible for analysis [22].

ML approaches play a crucial role in unravelling intriguing linkages within data, aided by the accessibility of intelligent tools for data analysis [23]. It can be used to either bolster or cast doubt on the conclusions reached by clinicians [24]. Clinical decision-making and the automatic tuning of bedside devices have both benefited from the introduction of machine learning (ML) based systems that constantly examine the expanding influx of data streams for patterns. As a result of these developments, patient treatment outcomes have improved, and overall treatment costs have decreased significantly [25]. According to the cited articles [26], the potential of machine learning (ML) to provide optimal therapeutic support has not been realised so far. This may be attributed to the lack of transparency in ML algorithms and analytics. Additionally,

the issues of data quality and generalizability of machine learning models are significant concerns [27].

### 2.4. Primary Weaknesses in these Regions

Some of the weaknesses are listed as follows:

- One of the primary weaknesses in these regions is the lack of comprehensive digital infrastructure to support AI and ML technologies. Limited access to quality data is another significant challenge, as machine learning's efficacy heavily depends on the availability of large, annotated, and high-quality datasets. In many developing areas, the data that exists is often fragmented, non-digitized, or of poor quality, hampering the development and deployment of effective ML-based diagnostic tools.
- Moreover, the shortage of local expertise in AI and machine learning poses a substantial barrier. Developing regions often lack the educational programs and research institutions necessary to cultivate talent in these cutting-edge fields, leading to a dependency on external expertise that may not fully grasp the local healthcare context and needs.
- Another critical concern is the ethical considerations surrounding the deployment of AI in healthcare, including issues of privacy, consent, and bias. These issues are magnified in regions without stringent regulations or frameworks governing the use of AI, potentially leading to mistrust and reluctance among healthcare providers and patients to adopt these technologies.

In summary, while machine learning holds immense potential to revolutionize disease diagnosis and prediction, its global impact is uneven due to significant weaknesses in certain regions. Addressing these challenges requires a concerted effort involving infrastructure development, education, data management, and ethical governance to ensure that the benefits of AI in healthcare are accessible to all, irrespective of geographic location.

## 3. EMPLOYING MACHINE LEARNING ALGORITHMS FOR DISEASE DIAGNOSIS AND PREDICTION

To detect the diseases, a variety of algorithms based on machine learning are applied. The majority of them are given below from this review study:

### 3.1. Support Vector Machine

Use a support vector machine (SVM) for more help. SVM has been proven to be useful for many classification problems. It searches for the optimal hyperplane by enumerating the number of nodes lying on the border between two classes. Margin refers to the value disparity between two groups. The success rate of a categorization system rises when there is a larger margin of safety. Boundary information is represented *via* support vectors. SVM can be used to tackle both regression and classification problems [28]. Both linear and nonlinear data sets benefit from this approach. The SVM method utilizes several different kernel types for use in a prediction model. These include linear radial basis function (RBF), polynomial, and sigmoid kernels. SVM

utilizes a high-dimensional space for attributes and selects the optimum hyperplane to categories data points into two groups. It is effective on large and small datasets that are normally challenging to analyses. Support vector machine (SVM) analysis of hyperplane data for diabetes detection is depicted in a relatively basic manner in Fig. (2) [29]. Fig. (2) depicts the process of classification using a Support Vector Machine (SVM), illustrating how this machine learning model constructs a hyperplane or set of hyperplanes in a high-dimensional space to categorize data points into distinct classes. The figure visually demonstrates the concept of maximizing the margin between different classes, which is fundamental to the SVM's effectiveness in classifying complex datasets.

### 3.2. Naive Bayes Algorithm

Naive Bayes (NB) is a classification algorithm used in statistics and probability, see Fig. (3). It is a common technique in machine learning applications because of its

ease of use and the fact that all factors can be considered equally when making a final call. The NB method is intriguing because of its simplicity and its versatility due to its computing efficiency. The NB classification relies on three key concepts: prior, posterior, and class conditional probability [30]. In addition to its many other merits, this method is particularly well-suited to and easy to apply to huge datasets. It could be applied to problems of binary and multi-class classification. Less data is required for training purposes, and both discrete and continuous data can be used. The application of this technique could be used to classify papers and eliminate junk mail [31]. Fig. (3) illustrates the Naïve Bayes Classifier approach, a probabilistic machine learning model that applies Bayes' theorem with the assumption of independence among predictors to predict the category of a given sample. The figure visually represents how this classifier calculates the probability of each class and the conditional probability of each feature within those classes to make predictions.

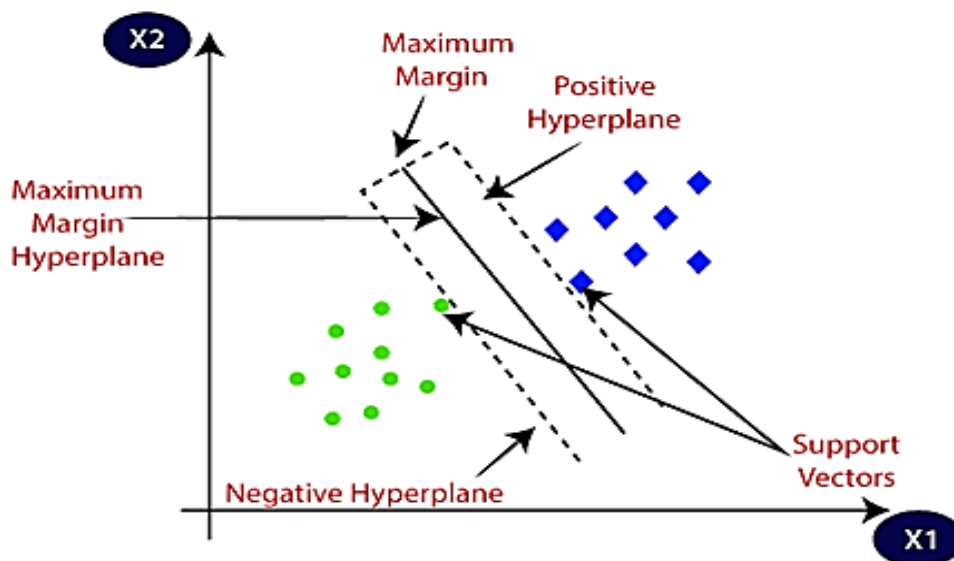
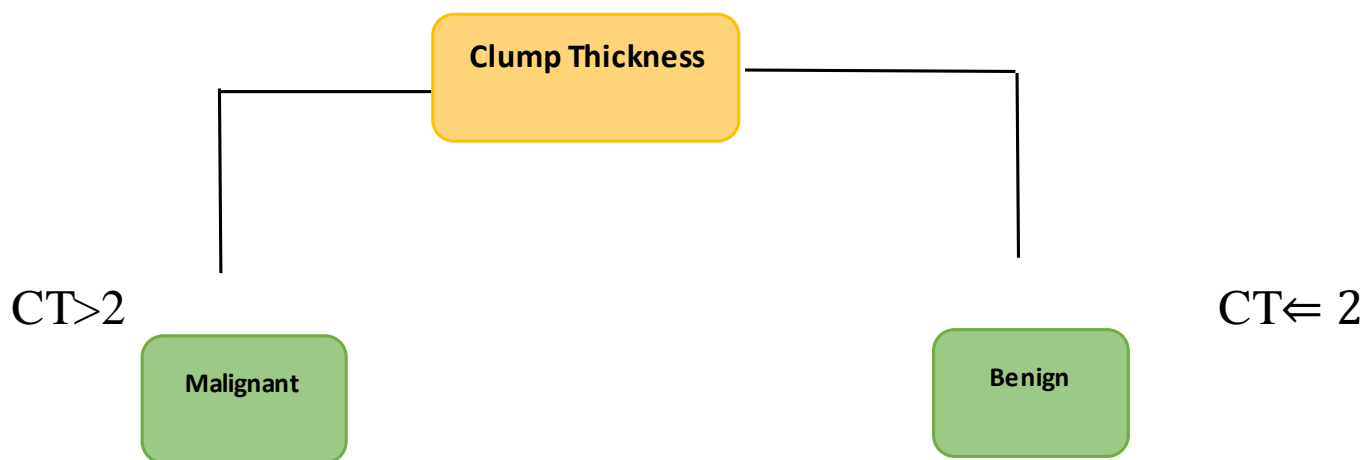


Fig. (2). Classification using SVM classifier.



Fig. (3). Naïve bayes classifier approach.



**Fig. (4).** Classification using a simple decision tree approach.

### 3.3. Decision Tree Algorithm

In order to tackle regression and classification issues, the Decision Tree (DT) uses periodic splitting of data based on a given variable. Nodes represent data structures, while the tree's leaf represents evaluations, (Fig. 4). The purpose of the decision tree is to construct a model for forecasting the dependent variable from training data by learning simple decision rules [32]. The tree is learned by constructing it from the training data. The name of the class is stored in a decision node, which is a non-leaf node. Both numerical and categorical information can be processed using the decision tree. The effectiveness of the tree remains unaffected by the nonlinear connection between arguments [33, 34]. No preliminary data preparation is required. When the tree is built again and over again, overfitting might occur.

The decision tree shown in Fig. (4) consists of a single leaf node, one child node, and the tree's parent node. Decision trees have found use in the medical field. If we use DT to diagnose breast cancer, for instance, we might assign a benign or malignant label to each of the tree's leaf nodes. Clump thickness (CT) will be used to establish whether or not the tumour is malignant by developing rules among the necessary data set parts. Fig. (4) shows the DT algorithm in action in a breast cancer diagnostic setting [35]. Fig. (4) showcases the classification process using a simple decision tree approach, highlighting how this method splits data into subsets based on the value of input features, creating a tree-like model of decisions. The figure demonstrates the hierarchical structure of decision nodes and leaf nodes, which represent the questions and outcomes, respectively, guiding the user through the logical steps of data classification.

### 3.4. Random Forest Classification

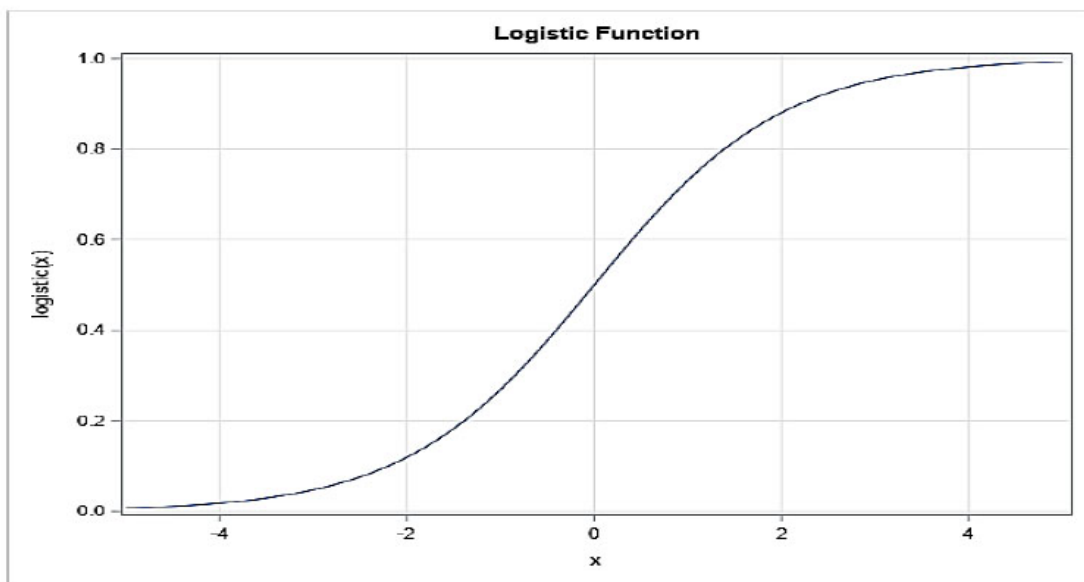
The random forest ensemble model can also be used to develop a predictor based on the nearest neighbors. An essential principle of ensemble methods is that when multiple models are combined, the resulting model is more

robust than any of the individual models. In addition to the random forest, the ensemble also includes a traditional ML technique decision tree. This procedure takes an initial input and branches out into finer and finer subsets of the data. The random forest expands on this concept by merging trees with the idea of an ensemble. A random forest classifier's strengths lie in its flexibility in dealing with missing data, its speedy execution, and its inherent inequity [36]. In a random forest, the newly generated subtrees travel with the testing data or the new dataset. Any decision subtree in the forest can be used to choose the dataset's classification. After a vote is taken, the model will select the most popular option. The fundamental idea behind the random forest algorithm for cardiac illness detection is depicted in Fig. (5) [37]. Fig. (5) illustrates the use of Random Forest for categorizing datasets, depicting how this ensemble learning method constructs multiple decision trees during training and outputs the class that is the mode of the classes of individual trees. This figure visually represents the aggregation of predictions from various decision trees to enhance classification accuracy and mitigate overfitting, characteristic of the Random Forest approach.

### 3.5. Logistic Regression

Logistic relapse is a regulated learning procedure used to resolve issues of double categorization. The logistic capability and logistic relapse, which have various more perplexing expansions, are utilized to show a twofold order in science. In its most essential structure, logistic relapse is a sort of relapse model that predicts the probability that a given data of interest or section has a place with a particular class [38]. Logistic regression employs the sigmoid function (Fig. 6) to model the data. Logistic regression's regularization simplicity, computational efficiency, and application ease are three of its most important features. Scaling input features is not required. However, overfitting and the ability to solve nonlinear problems [35].

**Fig. (5).** Use of random forest for categorizing datasets.



**Fig. (6).** Logistic function.

### 3.6. Deep Learning

As a subfield of machine learning, deep learning is distinguished by its use of multi-level layered nonlinear transformations. The Recurrent Neural Network (RNN) can use the input slices, which can be seen as a sort of sequential data, to infer relationships between slices [39]. Convolutional neural networks (CNNs) are frequently employed in deep learning applications for data analysis and complicated classification tasks. It has four layers

total: a convolutional layer, a max-pooling layer, a fully linked layer, and an output layer. A 32x32 pixel section of the original medical image is used as input for the CNN model. In deep neural networks, feature maps are created by convolutional layers and passed on to the following layers, where their size is reduced by pooling layers. In the end, the completely connected layer generates a prediction for the correct class, as described in another study [37].

**Table 1. Article that discussed the diagnosis of diabetes disease using machine learning.**

| Authors   | Diseases Discussed   | ML Algorithm         | Dataset                      | Performance Evaluation        |
|---|--|----------------------|------------------------------|-------------------------------|
| Fitriyani, N.L.; <i>et al.</i> (2019) [38]        | Hypertension and diabetes  | DPM                  | Privately owned              | 96.74% Accuracy               |
| Fernández-Edreira, D., <i>et al.</i> (2021) [35]. | Type 1 diabetes  | RF                   | DIABIM-MUNE                  | 0.80 - AUC                    |
| Ali, A.; Alrubei, <i>et al.</i> (2020) [39]       | Classification of Diabetes   | KNN                  | Privately owned 4900 samples | 99.9% Accuracy                |
| Tsao, H.Y.; Chan, P.Y.; Su, E.C.Y. (2018) [37]    | Predict diabetic retinopathy and identify biomedical characteristics that can be interpreted | SVM, DT, ANN, and LR | Privately owned              | 79.5% SVM Accuracy, 0.839-AUC |

**Table 2. Article that discussed the diagnosis of COVID-19 disease using machine learning.**

| Authors                                    | Diseases Discussed                                | ML Algorithm | Dataset         | Performance Evaluation |
|--|---|--------------|-----------------|------------------------|
| Song, Y.; Zheng, <i>et al.</i> (2021) [40] | Disease detection for COVID-19                    | CNN          | Mixed dataset   | 86% Accuracy           |
| Jin, C.; Chen, W <i>et al.</i> (2020) [41] | Disease detection for COVID-19                    | CNN          | Cohen's dataset | 94.1% Accuracy         |
| Ahsan, M.M. <i>et al.</i> (2020) [42]      | Image segmentation and COVID-19 disease detection | CNN          | Cohen's dataset | 95.38% Accuracy        |

#### 4. DIAGNOSIS OF DISEASES USING MACHINE LEARNING (DDML)

Machine learning (ML) techniques have found extensive application among scholars and professionals in diagnosing various illnesses. This section presents an overview of several notable machine learning-based disease diagnostic techniques recognized for their importance and criticality. The discussion highlights significant conditions such as heart disease, renal illness, breast cancer, diabetes, Parkinson's disease, Alzheimer's disease, and COVID-19 [40]. Meanwhile, other illnesses are briefly addressed within the category of other diseases, see Table 1.

##### 4.1. Parkinson's Disease

One medical problem that has been extensively discussed in the ML literature is Parkinson's disease. It is a chronic neurological condition with a sluggish progression rate. Damage to or loss of dopamine-producing neurons in specific brain regions impairs basic cognitive and motor functions. Several other methods based on ML have been suggested. In order to create accurate Parkinson's disease diagnostic systems, researchers like [41] employed algorithms like KNN, SVM, NB, and RF. In their computational results, RF is shown to have the highest performance (90.26% accuracy), whereas NB is shown to have the lowest performance (69.23% accuracy). Near-perfect accuracy was reached on both the training and testing sets when a deep CNN-based model was suggested [42]. However, no issues with overfitting were mentioned throughout the experiment. Finally, the ultimate classification and regression, which is now generally anticipated, especially in CDSS, is not well interpreted by the experimental data. For Parkinson's telemonitoring speech data also turned to DL-based methods at UCI [43]. In their trial, DNN was able to correctly identify Parkinson's illness in 81.67% of cases.

##### 4.2. COVID-19

Another pestilence, SARS-CoV-2 (or Coronavirus), has emerged as quite possibly of humankind's most noteworthy danger. Regardless of the worldwide race to get an immunization to showcase in light of the pestilence, it remained exorbitant to by far most individuals for its whole span. As a result of its high transmission rates and immunization-related opposition, the clever Coronavirus Omicron strain is causing more concern. As of now, Ongoing Reverse Transcription-Polymerase Chain Reaction (RT-PCR) is the highest quality level for deciding if a patient has contracted Coronavirus. The analyst supported the utilization of Machine Learning and Artificial Intelligence in conjunction with other demonstrative strategies like chest X-beams and Computed Tomography (CT) during the plague. Chen *et al.* (2020) made a UNet++ model with CT information from 51 Coronavirus and 82 non-coronavirus patients, and it accomplished an exactness of 98.5%. The authors surveyed ten different DL models utilizing just 108 Coronavirus patients and 86 non-COVID-19 patients, and they had the option to accomplish a general exactness of almost 100%. Utilizing a tremendous dataset of 453 CT examination pictures, Wang *et al.* (2020) [44] made a commencement-based model that accomplished 73.1% precision. Nonetheless, data about the degree and nature of the model's organizational activity was restricted. Authors [42] proposed the COVNet model and accomplished 96% exactness utilizing an enormous dataset involving 4356 chest CT pictures of Pneumonia patients, of which 1296 were confirmed Coronavirus occurrences. The summary of existing studies is shown in Table 2.

##### 4.3. Alzheimer's Disease

Sixty percent to seventy percent of persons who are diagnosed with dementia [43] have Alzheimer's disease, a brain condition that usually starts slowly but develops over time. Alzheimer's disease causes cognitive decline, which



manifests in a variety of ways. The steady decline in body functioning is reflected in a prognosis of three to nine years from the time of diagnosis. On the other hand, getting a diagnosis early may help you take preventative measures and start appropriate treatment as soon as possible, both of which increase your chances of survival. Over the years, the use of machine learning and deep learning to identify people with Alzheimer's disease has shown encouraging results. A CNN model for early detection and categorization of Alzheimer's disease was suggested by Ahmed *et al.* (2020) [44]. The suggested model obtained 99% accuracy within the dataset consisting of 6628 MRI pictures. The deep feature-based models provided by Nawaz *et al.* (2020) [45] have an accuracy of 99.12%.

#### 4.4. Other Diseases

ML and DL have been used for the diagnosis of a wide range of additional disorders in addition to the one that was discussed before. Two primary factors for this rise in utilisation are the proliferation of large amounts of data and the rise in the processing capability of computers. For instance, Mao *et al.* (2020) [41] classified diseases based on eye movement by using Decision Tree (DT) and Random Forest (RF). When Nosseir and Shawky (2019) [39] developed automated skin disease classification systems using KNN and SVM, they found that KNN achieved the greatest performance by attaining an accuracy of 98.22%. This was the best performance that was seen utilising KNN. Khan *et al.* (2020) [37] were able to classify multimodal brain tumours using CNN-based methods such as VGG16 and VGG19. The experiment was conducted using three picture datasets that were readily accessible to the public: BraTs2015, BraTs2017, and BraTs2018. The results showed an accuracy of 97.8%, 96.9%, and 92.5%, respectively.

### 5. UTILIZING MACHINE LEARNING FOR DIAGNOSING IMAGE-BASED DISORDERS/ ILLNESSES

Machine learning algorithms are prevalent across various domains. Particularly in the realm of AI, the

utilization of Machine Learning (ML) for medical image analysis has gained rapid traction. Diverse fields, including computer vision, computer-assisted diagnostics, and image processing for disease identification, have seen significant advancements through ML applications. The introduction of cutting-edge imaging techniques in medical imaging—such as computed tomography (CT), positron emission tomography (PET), tomosynthesis, magnetic resonance imaging (MRI), and tomography—has substantially evolved the field. Consequently, there is an escalating demand for cutting-edge machine learning methods to streamline medical imaging data analysis. Machine Learning (ML) encompasses a set of algorithms specifically designed to automatically recognize patterns within datasets, facilitating predictive analysis or informed decision-making in ambiguous situations. Distinguishing itself from other approaches, ML heavily relies on data, leaving minimal room for human judgment. These algorithms assimilate new data and juxtapose it with their training data to make predictions.

Modern machine-learning methodologies have been employed for disease prediction or identification. Natural Language Processing (NLP) techniques have been applied to scrutinize electronic health records (EHRs) to extract relevant information useful in disease prediction or diagnosis. Explainable artificial intelligence (AI) methodologies, such as SHAP (SHapley Additive exPlanations), have been introduced to interpret the predictions derived from machine learning (ML) models. This interpretability is crucial in the medical domain, where transparent decision-making holds paramount importance. Generative models, notably Generative Adversarial Networks (GANs), have been developed to generate synthetic medical images. These synthetic images serve to enrich existing datasets, particularly in the context of lung disease, ultimately enhancing model performance. These strategies can be integrated to improve model performance, with the choice of method contingent upon the nature of the data and the specific situation at hand.

**Table 3. Comparison of machine learning algorithms, its evaluation procedures and resulting outcomes.**

| Authors   | Type of Disease | Machine Learning Algorithms/ Evaluation Procedures | Dataset  | Resulting Outcomes                               | Finding   |
|---|-----------------|--|--|--|---|
| Md. Zabirul Islam, <i>et al</i> (2020) [43]         | COVID-19        | CNN  | GitHub and Kaggle                                    | 94% Accuracy                                     | The suggested model did well in terms of detection accuracy (CAVID-19), but this is not indicative of a production-ready solution, especially given the limited quantity of images. |
| Vidya M and Maya V Karki, (2020) [44]               | Skin cancer     | SVM and KNN  | Collected from ISIC 2017 datasets with 1000 instance | 97.8% Accuracy SVM And 86.25 KNN                 | For textural features, the SVM classifier performs exceptionally, with an accuracy of 97.8% and an AUC of 0.94.   |
| Dayanand Jamkhandikar and Neethi Priya. (2020) [45] | Thyroid disease | Naïve Bayes, KNN, and SVM algorithm                | UCI data repository site                             | 82% Accuracy of SVM, 83% Naive Bayes and 85% KNN | Better cost and facility management are used in the treatment of thyroid patients.  |

| Authors                                      | Type of Disease | Machine Learning Algorithms/ Evaluation Procedures                          | Dataset  | Resulting Outcomes   | Finding   |
|--|-----------------|---|--|--|---|
| Mahzabeen Emu, <i>et al</i> (2020) [46]      | liver fibrosis  | Random forests, MLP, logistic regression algorithm.                         | Collected by Ain Shams University, Faculty of Medicine at El Demerdash Hospital. | 97.228% Accuracy random forest, 98% MLP and 97% Logistic regression  | In this research, MLP had the greatest accuracy rate of all the machine learning techniques.  |
| N. Komal Kumar, <i>et al</i> (2020) [47]     | Cardio Vascular | Random Forest, Decision Tree, Logistic Regression, SVM, and KNN algorithms. | NIDDK.   | 85.71% Accuracy of Random Forest, 74.28% decision tree, 74.28% Logistic Regression, 77.14 SVM and 68.57% KNN     | All of the classifiers under the Analysis of categorizing CVD patients outperform the random forest.  |
| Nikita Banerjee Subhalaxmi Das, (2020) [48]  | Lung cancer     | SVM, Random Forest, ANN algorithms  | UCI machine learning repository  | 70% Accuracy Random Forest, 80% SVM and 96% ANN  | The model can distinguish between benign and malignant tumours, and it can be demonstrated that both artificial neural networks with a focus on texture and regions have higher levels of accuracy. |
| Oyewo O.A and Boyinbode O.K (2020) [49]      | Prostate cancer | SVM, Decision Tree, and MLP algorithm                                       | github.com   | 99.06% accuracy of model   | The detection efficacy extends to both non-prostate and prostate patients.  |
| J.Neelaveni and Geetha Devasana, (2020) [50] | Alzheimer       | SVM, a decision tree algorithm  | ADNI database  | 85% SVM Accuracy and 83% Decision Tree   | The model forecasts the patient's condition and differentiates between cognitive impairment and dementia.   |
| Nashat Alrefai (2019) [51]                   | Leukemia Cancer | SVM, KNN, Naïve Bayes (NB), and Decision Tree (C4.5) algorithms             | Microarray dataset   | The classification accuracy and performance on microarray datasets are better than other individual classifiers. | Both the accuracy of the classification of microarray datasets and the performance they provide are much greater than those of other individual classifiers.  |

## 6. COMPARISON OF MACHINE LEARNING ALGORITHMS, ITS EVALUATION PROCEDURES AND RESULTING OUTCOMES

The above Table 3 provides an overview of the previous research studies. The main focus of each study, the dataset that was used, the techniques or algorithms that were employed, the accuracy levels that were achieved, and the conclusions that were drawn from the study results.

## 7. DISCUSSION

All the earlier sections encompassed articles detailing studies dedicated to diagnosing illnesses utilizing various machine-learning techniques. Table 3 offers a comprehensive summary of the prior works referenced in this article. This summary highlights each study's focus on the illness examined, the dataset utilized, the techniques or algorithms employed, the achieved accuracy levels, and the resulting conclusions. A multitude of algorithms such as logistic regression, K-nearest neighbors, support vector machines, k-means, decision trees, random forests, and ensemble models were analyzed across diverse studies. These methods were applied to commonly used datasets for various illnesses. For instance, in another study [39], researchers diagnosed liver diseases using three distinct algorithms, selecting logistic regression due to its reliability and precision in producing findings. KNN and SVM algorithms were employed to analyze data from different datasets and feature sets [40]. Meanwhile, KNN was marginally superior to SVM by about 3% [41]. Additionally, research showed a marginal accuracy difference of about 1% between the decision tree and the random forest, whereas N. Komal Kumar, *et al.* 2020

demonstrated that the random forest is approximately 12% more accurate than the decision tree.

Researchers using an ensemble model [45] and by Rahma Atallah and Amjed Al-Mousa, (2019) achieved success rates of 98.50% and 90%, respectively. K-means is utilized for multi-clustering, while SVM and logistic regression act as binary classifiers. Conversely, decision trees, random forests, KNN, naive Bayes, and CNN have the capacity to classify into more than two classes. Furthermore, no other clustering algorithm can match the functionality of K-means. The discussion of this research can be listed as follows:

- The section reviews articles that use various machine learning techniques to diagnose illnesses, summarizing the focus illness, datasets, algorithms, accuracy, and conclusions.
- A wide array of algorithms, including logistic regression, K-nearest neighbors (KNN), support vector machines (SVM), k-means, decision trees, random forests, and ensemble models, were explored in these studies.
- The algorithms were applied to well-known datasets for diagnosing different illnesses, demonstrating the versatility and applicability of machine learning in medical diagnostics.
- Specific findings include the use of logistic regression for its effectiveness in liver disease diagnosis, and a comparison between KNN and SVM showing KNN's slight superiority in accuracy by about 3%.
- Another comparison indicates a minimal accuracy difference between decision trees and random forests, with one study highlighting that random forests outperform decision trees by approximately 12%.

## 8. GAPS IN AI'S APPLICABILITY AND RELIABILITY

While artificial intelligence (AI) has made significant strides in the medical field, particularly in disease diagnosis and prediction, there are considerable gaps that need to be addressed before AI can be considered a comprehensive replacement for human judgment in all areas of medicine.

### 8.1. Data Quality and Availability

AI systems require vast amounts of high-quality, diverse, and well-annotated data to learn effectively. However, there is a significant gap in the availability of such data, especially from underrepresented groups or regions, leading to potential biases in AI-driven diagnoses.

### 8.2. Complexity of Medical Conditions

Some medical conditions are incredibly complex and nuanced, requiring a holistic understanding of the patient's history, symptoms, and other factors that AI might not fully integrate or interpret accurately.

### 8.3. Interdisciplinary Knowledge

Medicine often requires interdisciplinary knowledge, combining insights from various fields (*e.g.*, psychology, sociology, genetics). AI systems, primarily trained on specific datasets, may lack this comprehensive approach.

### 8.4. Adaptability and Continuous Learning

The medical field is continuously evolving with new research findings and treatment methods. AI systems need constant updates and training to stay current, which can be a significant challenge.

## 9. ETHICAL AND RESPONSIBILITY CONCERNS

### 9.1. Responsibility for Incorrect Diagnoses

The risk of incorrect diagnoses and the associated liability is a major concern. Determining the responsibility whether it falls on the AI developers, the healthcare providers, or the technology itself is a complex issue that has yet to be fully resolved.

### 9.2. Ethical Use of Data

The ethical implications of using patient data to train AI systems are immense. Issues of consent, privacy, and data security are paramount, especially when dealing with sensitive health information.

### 9.3. Bias and Inequality

AI systems can perpetuate and even exacerbate existing biases and inequalities if they're trained on biased data sets or if their deployment overlooks the diverse needs of different populations. Ensuring fairness and equity in AI-driven medical diagnoses is a significant challenge that needs to be addressed.

In summary, while AI has the potential to augment and, in some cases, transform medical diagnosis, it is not yet in a position to replace human oversight entirely. The technology must overcome substantial gaps in data quality, complexity understanding, adaptability, and

ethical governance to ensure that AI's integration into healthcare enhances patient outcomes without compromising on safety, fairness, and privacy.

## CONCLUSION

Leveraging machine learning algorithms in medical diagnostics stands as a significant opportunity to seamlessly integrate computer-based technologies into healthcare systems. These methods aid healthcare practitioners in early illness detection, particularly in regions like India where limited medical resources, high mortality rates, and a shortage of physicians exist, with just one doctor for every 1700 individuals. It is important to recognize that technology cannot substitute the extensive training and experience essential to becoming a physician. Nonetheless, it can handle routine and time-consuming diagnostic tasks, enabling doctors to focus on more intricate clinical procedures. Recently, machine learning has emerged as a potent tool for analyzing disease-related data in the medical field, with a significant emphasis on early disease detection. This research aims to provide a comprehensive overview of the current landscape of machine learning techniques in disease prognosis. Various diseases, including liver illness, chronic kidney disease, breast cancer, cardiac syndrome, and brain tumors, are among those typically represented in the datasets used.

The study compiles and summarizes the outcomes of multiple machine-learning algorithms utilized by researchers for disease diagnosis. Several algorithms, like SVM, K-nearest neighbors, Random Forests, and Decision Trees, exhibit remarkably high prediction accuracy, showcasing consistent performance and reliability across different articles using various predictive models. However, the precision of each method can vary based on the specific dataset used. The accuracy and performance of these models are highly sensitive to critical factors such as dataset quality, methods for selecting features, and the quantity of features included. The findings underscore the significance of employing ensemble learning techniques that amalgamate multiple strategies to enhance model accuracy and performance, marking a substantial advancement in the realm of using machine learning for disease prediction.

## PROSPECTS FOR THE FUTURE

We anticipate that our review will be beneficial for both novice and experienced researchers and practitioners in the field of DDML. Further investigations are warranted into the limitations highlighted in the paper's discussion section. Future work in DDML might focus on advancing large-scale data integration, encompassing numerical, aggregate, and image data. Additionally, there's potential for exploration in multiclass classification with highly imbalanced and missing data, alongside delving into the explanation and interpretation of multiclass data classification using XAI (Explainable Artificial Intelligence). Moving forward, exploring additional parameters in building ML models could refine model selection, potentially reducing execution time. For

instance, leveraging CNN (Convolutional Neural Networks) may yield better outcomes for image data. Data normalization can be crucial for impartial results, while the use of deep learning and ensemble models holds promise. Simultaneous diagnosis for patients with multiple diseases could also be an area of further exploration.

## LIST OF ABBREVIATIONS

|      |   |                                 |
|------|---|---------------------------------|
| ML   | = | Machine Learning                |
| EHRs | = | electronic health records       |
| NLP  | = | Natural Language Processing     |
| GANs | = | Generative Adversarial Networks |

## CONSENT FOR PUBLICATION

Not applicable.

## FUNDING

None.

## CONFLICT OF INTEREST

The authors declare no conflict of interest financial or otherwise.

## ACKNOWLEDGEMENTS

Declared none.

## REFERENCES

- Jutel A. Sociology of diagnosis: A preliminary review. *Social Health Illn* 2009; 31(2): 278-99. <http://dx.doi.org/10.1111/j.1467-9566.2008.01152.x> PMID: 19220801
- Md NE. The Performance Analysis of K-Nearest Neighbors (K-NN) algorithm for motor imagery classification based on E.E.G. *Signal. MATEC Web Conf* 2017; 104(01024): 6.
- Zeebaree DQ, Haron H, Abdulazeez AM, Zeebaree SRM. Combination of K-means clustering with Genetic Algorithm: A review. *Int J Appl Eng Res* 2017; 12(24): 14238-45.
- Jahwar AF. Meta-heuristic algorithms for k-means clustering: A review. *Palarch's J Archaeol Egypt/Egyptol* 2021; 17(7): 1567.
- Shukla S, Naganna S. A Review ON K-means DATA Clustering APPROACH. *Int J Inf Comput Technol* 2014; 4(17): 1847-60.
- Yahyaoui A. A decision support system for diabetes prediction using machine learning and deep learning techniques. 2019 1st International Informatics and Software Engineering Conference (UBMYK) 06-07 November 2019, Ankara, Turkey, 2019.
- Osman AH, Aljahdali HM. Diabetes Disease Diagnosis Method based on Feature Extraction using K-SVM. *Int J Adv Comput Sci Appl* 2017; 8 1.
- Singh R. Collaborative filtering based hybrid music recommendation system Proceedings of the 3rd International Conference on Intelligent Sustainable Systems, ICISS 2020. 03-05 December 2020, Thoothukudi, India, 2020. <http://dx.doi.org/10.1109/ICISS49785.2020.9315913>
- Yaswanth R. Heart disease prediction using machine learning techniques. *Int J Innov Technol Exp Eng* 2020; 9(5)1
- Chauhan YJ. Cardiovascular Disease Prediction using Classification Algorithms of Machine Learning. *Int J Sci Res* 2018; 20182319
- Tian X. Using machine learning algorithms to predict Hepatitis B surface antigen seroclearance. *Comput Math Methods Med* 2019; 20196915850 <http://dx.doi.org/10.1155/2019/6915850>
- Bohat VK. Neural network model for recommending music based on music genres. 2021 International Conference on Computer Communication and Informatics, ICCCI 2021. 27-29 January 2021, Coimbatore, India, 2021
- Venkata Lakshmi S, Meena M K, Kiruthika N S. >Diagnosis of chronic kidney disease using random forest algorithms *CMBEBIH* 2017; 62589
- Kumar KV. Random forest algorithm for the prediction of diabetes. *Proceeding of International Conference on Systems Computation Automation and Networking*.
- Meshref H. Cardiovascular Disease Diagnosis: A Machine Learning Interpretation Approach. *Int J Adv Comput Sci Appl* 2019; 10 12. <http://dx.doi.org/10.14569/IJACSA.2019.0101236>
- Jacob Joel, Mathew Joseph Chakkalalal, Mathew Johns, Issac Elizabeth. Diagnosis of liver disease using machine learning techniques. *Int Res J Eng Technol* 2018; 5(4): 1.
- Hesamian MH. Deep learning techniques for medical image segmentation: Achievements and challenges. *J Dig Imag* 2019; 32: 582-96.
- Aggarwal R, Tiwari S, Joshi V. Exam proctoring classification using eye gaze detection. 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC). 20-22 October 2022, Trichy, India, 2022. <http://dx.doi.org/10.1109/ICOSEC54921.2022.9951987>
- Ahsan MM, Luna SA, Siddique Z. Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. *Healthcare (Basel)* 2022; 10(3): 541. <http://dx.doi.org/10.3390/healthcare10030541> PMID: 35327018
- Ahsan M, Mahmud M, Saha P, Gupta K, Siddique Z. Effect of data scaling methods on machine Learning algorithms and model performance. *Technologies (Basel)* 2021; 9(3): 52. <http://dx.doi.org/10.3390/technologies9030052>
- Ansari AQ, Gupta NK. Automated diagnosis of coronary heart disease using neuro-fuzzy integrated system. *Proceedings of the 2011 World Congress on Information and Communication Technologies*. 11-14 December 2011, Mumbai, India, 2011. <http://dx.doi.org/10.1109/WICT.2011.6141450>
- Rubin J, Abreu R, Ganguli A, Nelaturi S, Matei I, Sricharan K. Recognizing abnormal heart sounds using deep learning 2017.
- Charleonnana A, Fufaung T, Niyomwong T, Chokchueypattanakit W, Suwannawach S, Ninchawee N. Predictive analytics for chronic kidney disease using machine learning techniques. *Proceedings of the 2016 Management and Innovation Technology International Conference*. 12-14 October 2016, Bang-Saen, Chonburi, Thailand, 2016 <http://dx.doi.org/10.1109/MITICON.2016.8025242>
- Aljaaf AJ, Al-Jumeily D, Haglan HM, et al. Early prediction of chronic kidney disease using machine learning supported by predictive analytics. *Proceedings of the 2018 IEEE Congress on Evolutionary Computation (CEC)*. 08-13 July 2018, Rio de Janeiro, Brazil, 2018. <http://dx.doi.org/10.1109/CEC.2018.8477876>
- Yadav A, Kumar S. A Review of Physical Unclonable Functions (PUFs) and Its Applications in IoT Environment *Ambient Communications and Computer Systems*. Berlin, Heidelberg: Springer Link 2022; pp. 1-13. [http://dx.doi.org/10.1007/978-981-16-7952-0\\_1](http://dx.doi.org/10.1007/978-981-16-7952-0_1)
- Miranda GHB, Felipe JC. Computer-aided diagnosis system based on fuzzy logic for breast cancer categorization. *Comput Biol Med* 2015; 64: 334-46. <http://dx.doi.org/10.1016/j.combiomed.2014.10.006> PMID: 25453323
- Zheng B, Yoon SW, Lam SS. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Syst Appl* 2014; 41(4): 1476-82. <http://dx.doi.org/10.1016/j.eswa.2013.08.044>
- Rajendran K, Jayabalan M, Thiruchelvam V. Predicting breast cancer via supervised machine learning methods on class imbalanced data. *Int J Adv Comput Sci Appl* 2020; 11(8): 54-63. <http://dx.doi.org/10.14569/IJACSA.2020.0110808>

- [29] Abdel-Nasser M, Rashwan HA, Puig D, Moreno A. Analysis of tissue abnormality and breast density in mammographic images using a uniform local directional pattern. *Expert Syst Appl* 2015; 42(24): 9499-511.  
<http://dx.doi.org/10.1016/j.eswa.2015.07.072>
- [30] Sharma S, Khanna P. Computer-aided diagnosis of malignant mammograms using Zernike moments and SVM. *J Digit Imaging* 2015; 28(1): 77-90.  
<http://dx.doi.org/10.1007/s10278-014-9719-7> PMID: 25005867
- [31] Moon WK, Chen IL, Chang JM, Shin SU, Lo CM, Chang RF. The adaptive computer-aided diagnosis system based on tumor sizes for the classification of breast tumors detected at screening ultrasound. *Ultrasonics* 2017; 76: 70-7.  
<http://dx.doi.org/10.1016/j.ultras.2016.12.017> PMID: 28086107
- [32] Naz H, Ahuja S. Deep learning approach for diabetes prediction using PIMA Indian dataset. *J Diabetes Metab Disord* 2020; 19(1): 391-403.  
<http://dx.doi.org/10.1007/s40200-020-00520-5> PMID: 32550190
- [33] Yahyaoui A, Jamil A, Rasheed J, Yesiltepe M. A decision support system for diabetes prediction using machine learning and deep learning techniques. *Proceedings of the 2019 1st International Informatics and Software Engineering Conference (UBMYK)*. 06-07 November 2019, Ankara, Turkey, 2019  
<http://dx.doi.org/10.1109/UBMYK48245.2019.8965556>
- [34] Raza SM, Sajid M. Vehicle routing problem using reinforcement learning: Recent advancements. *Advanced Machine Intelligence and Signal Processing*. Berlin, Heidelberg: Springer Link 2022; pp. 269-80.  
[http://dx.doi.org/10.1007/978-981-19-0840-8\\_20](http://dx.doi.org/10.1007/978-981-19-0840-8_20)
- [35] Fernández-Edreira D, Liñares-Blanco J, Fernandez-Lozano C. Machine Learning analysis of the human infant gut microbiome identifies influential species in type 1 diabetes. *Expert Syst Appl* 2021; 185: 115648.  
<http://dx.doi.org/10.1016/j.eswa.2021.115648>
- [36] Ali A, Alrubei MAT, Hassan LFM, Al-Ja'afari MAM, Abdulwahed SH. Diabetes Diagnosis based on KNN. *IJUM Engineering Journal* 2020; 21(1): 175-81.  
<http://dx.doi.org/10.31436/iiumej.v21i1.1206>
- [37] Tsao HY, Chan PY, Su ECY. Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms. *BMC Bioinformatics* 2018; 19(S9) > : 283.  
<http://dx.doi.org/10.1186/s12859-018-2277-0> PMID: 30367589
- [38] Fitriyani NL, Syafrudin M, Alfian G, Rhee J. Development of disease prediction model based on ensemble learning approach for diabetes and hypertension. *IEEE Access* 2019; 19: 144777-89.  
<http://dx.doi.org/10.1109/ACCESS.2019.2945129>
- [39] Ali A, Alrubei MAT, Hassan LFM, Al-Ja'afari MAM, Abdulwahed SH. Diabetes Diagnosis based on KNN. *IJUM Eng J* 2020; 21(1): 175-81.  
<http://dx.doi.org/10.31436/iiumej.v21i1.1206>
- [40] Song Y, Zheng S, Li L, *et al.* Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. *IEEE/ACM Trans. IEEE/ACM Trans Comput Biol Bioinformatics* 2021; 18(6): 2775-80.  
<http://dx.doi.org/10.1109/TCBB.2021.3065361> PMID: 33705321
- [41] Jin C, Chen W, Cao Y, *et al.* Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. *Nat Commun* 2020; 11(1): 5088.  
<http://dx.doi.org/10.1038/s41467-020-18685-1> PMID: 33037212
- [42] Ahsan MM, Alam E. Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and non-COVID-19 patients. *Symmetry (Basel)* 2020; 12: 1526.  
<http://dx.doi.org/10.3390/sym12091526>
- [43] Islam Z, Asraf A. A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Inform Med Unlocked* 2020; 20: 100412.
- [44] Vidya M, Maya V. Skin cancer detection using machine learning techniques. *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*. 02-04 July 2020, Bangalore, India, 2020.
- [45] Priya N, Jamkhandikar D. Thyroid disease prediction using feature selection and machine learning classifiers. *Int J Analyt Exp Modal Analys* 2020; 20: 0886.9367
- [46] Emu M. Assisting the non-invasive diagnosis of liver fibrosis stages using machine learning methods. *Annu Int Conf IEEE Eng Med Biol Soc* 2020; 2020: 5382-7.
- [47] Komal Kumar N, Sarika Sindhu G, Krishna Prashanthi D, Shaheen Sulthana A. Analysis and Prediction of Cardio Vascular Disease using Machine Learning Classifiers 6th International Conference on Advanced Computing & Communication Systems (ICACCS).  
<http://dx.doi.org/10.1109/ICACCS48705.2020.9074183>
- [48] Banerjee N. Machine learning techniques for prediction of lung cancer. *Int J Recent Technol Eng* 2020; 8(6): 2277-3878.
- [49] Oyewo O. Prediction of Prostate Cancer using Ensemble of Machine Learning Techniques. *Int J Adv Comput Sci Appl* 2020; 11 3.
- [50] Neelaveni J. Alzheimer Disease Prediction using Machine Learning Algorithm. 6th International Conference on Advanced Computing & Communication Systems.
- [51] Alrefai N. Ensemble machine learning for leukemia cancer diagnosis based on microarray datasets. *Int J Appl Eng Res* 2019; 14(21): 4077-84.

**DISCLAIMER:** The above article has been published, as is, ahead-of-print, to provide early visibility but is not the final version. Major publication processes like copyediting, proofing, typesetting and further review are still to be done and may lead to changes in the final published version, if it is eventually published. All legal disclaimers that apply to the final published article also apply to this ahead-of-print version.