

The Role of Machine Learning Algorithms for Diagnosing Diseases

Abstract

Nowadays, machine learning algorithms have become very important in the medical sector, especially for diagnosing disease from the medical database. Many companies using these techniques for the early prediction of diseases and enhance medical diagnostics. The motivation of this paper is to give an overview of the machine learning algorithms that are applied for the identification and prediction of many diseases such as Naïve Bayes, logistic regression, support vector machine, K-nearest neighbor, K-means clustering, decision tree, and random forest. In this work, many previous studies were reviewed that used machine learning algorithms for detecting various diseases in the medical area in the last three years. A comparison is provided concerning these algorithms, assessment processes, and the obtained results. Finally, a discussion of the previous works is presented.

Keywords: *Disease Diagnostics, Machine Learning, Classification Algorithm*

I. INTRODUCTION

Disease diagnosis is the method of determining which disease explains a person's symptoms. Some symptoms and signs are non-specific, and thus the most challenging problem is the diagnosis. Identification of diseases is the most significant point for the treatment of any diseases. Machine learning is the area that can help to predict the disease diagnosis based on the previous training data [1,2]. Many scientists have created various machine learning methods to operate efficiently to diagnose various diseases [3]. Machine learning offers the ability for machines to learn without being specifically programmed. Developing a model by machine learning algorithms can predict an early-stage diagnosis of disease and provide solutions. An early diagnosis and effective treatment are the best way to minimize the death rates induced by any disease. Therefore, most medical scientists are drawn to new predictive model technologies for disease prediction based on machine learning algorithms [4,5,6].

Machine learning models learn from patterns in given training examples without explicit instructions and then use inference to develop useful predictions [7]. Classification methods are widespread in the medical area for identifying and predicting diseases more accurately [8,9,10]. Diseases and

health problems like liver cancer, chronic kidney, breast cancer, diabetes, and heart syndrome have a significant impact on one's health and may lead to death if ignored. By extracting hidden patterns and relationships in the database, the healthcare industry will make successful decisions. Due to progress in machine learning and artificial intelligence, several classifiers and clustering algorithms like K-nearest, Decision Tree, Random Forest, Support Vector Machine (SVM), Naïve Bayes, and other algorithms can give a solution to this situation [11,12,13].

In medical applications, the most famous machine learning technique is classification because it corresponds to problems appearing in everyday life [14]. Classification algorithms first use the training data to build a model and then the resulting model is applied to the test data to obtain a prediction. Different classification methods have been implemented for disease diagnosis and the findings be very promising [15]. These techniques can minimize diagnostic faults and results can be obtained in a short period of time [16].

The rest of this paper is structured as follows. Section II presents machine learning types. Section III presents machine learning algorithms used for disease diagnosis, section IV discussing the previous work. In contrast, section V presents

discussion and comparison. Finally, the conclusion is presented in section VI.

II. MACHINE LEARNING TYPES:

Machine learning (ML) is an artificial intelligence branch that allows the machine to think like such a human being and can make a judgment on their own without human intervention [17,18,19]. It is the process of automatically making machines to learn without being programmed directly. The key aim of Machine Learning is to build a computer program that can access data and use it for learning processes. Many types of machine learning exist, as illustrated in Figure 1 [20]:

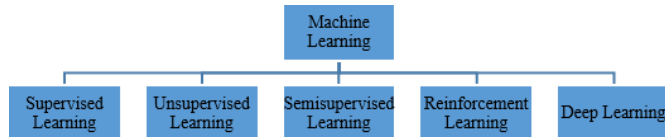


Fig 1. The machine learning types

This paragraph will describe a brief definition of each type of machine learning. Supervised learning means the algorithm learns to predict data from input data, and this type has input and output data. While in unsupervised learning, the algorithm has only input data and learns the inherent structure from the input data [21]. Moreover, semi-supervised learning depends on both techniques (supervised and unsupervised). They utilize both labeled and unlabeled data that mean some input data have label data, and some do not have label data [22,23]. In reinforcement learning, the system attempts to learn through interaction with the environment and reward the desired action and punish undesired ones [24]. The application of these methods is used for a number of medical functions, such as diagnosis of diseases [25,26]. Deep Learning Is a subset of machine learning that consists of multiple layers that holding multiple levels of perception, such that each layer receives information from the previous layer and passes the result to the next layer [27].

III. MACHINE LEARNING ALGORITHMS

There are various machine learning approaches are used to identify the disease. Most of them found in this review paper is listed below:

A. K Nearest Neighbour Algorithm (KNN)

KNN is a simple model and one of the most widely used machine learning processes for classification problems, pattern recognition, and regression. KNN get neighbors among data using Euclidean distance between points of data. This algorithm is used for classification and regression problems [28,29,30]. The value of k (where k is a constant defined by the user) will find all of the similar existing feature cases with the new case and surround all cases to find the new case for a similar category [31]. Therefore, the K value is significant and must be carefully chosen because the system can lead to overfitting if K is very small. There are a few limitations, such as poor performance when there is an extensive training dataset. The computation cost is very high since we need to measure the distance of all

training samples from each query instance [32,33]. Figure 2 represents the basic idea of the KNN algorithm for classifying new data point when the value of k equal to 9. The new point to be classified is (0.6, 0.45), as shown with "X." Two possible classes exist in the large dotted circle in the figure, which contains three objects of the tringle class and five objects the circle class. Using the Euclidean distance, the algorithm will classify the new point "X" to the circle class [34].

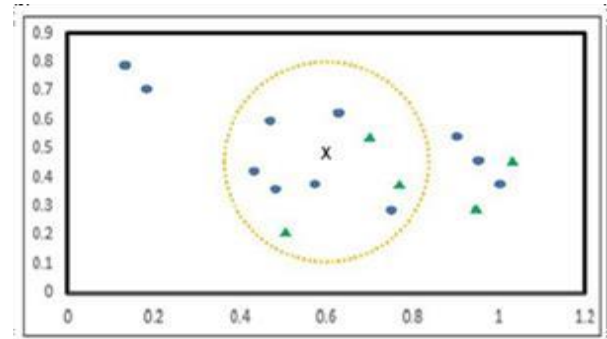


Fig 2. KNN Algorithm for classfing new point

B. K-Means Clustering Algorithm

K-Means is an unsupervised learning algorithm commonly used for the clustering process based on the nearest neighbor. The data can be clustered into k cluster based on the similarity between them. K is an integer number, and for the algorithm to operate, its value must be known. [35,36]. K-mean is the most commonly used algorithm for clustering, and it is capable of detecting the right cluster new data according to the majority of the distance [37,38,39]. The choice of k cluster centroids is initially performed randomly; then, allocate all points to its closest centroids and recomputed centroids for the newly assembled group. K-means are particularly sensitive to noise and outliers, as some are influenced by centroids [40]. One benefit of the K-means method is easy to implement and interpret and effective in computational terms. The drawback of this approach is that K value estimation is difficult. When clusters are globular, efficiency suffers. [41]. Figure 3 illustrates the graphical representation of the K-means algorithm. Two sets of objects are present in the first step. Then they decide the centroids of both sets. The clusters that produced the various dataset clusters are formed again according to the centroid. This process repeats until the best clusters are reached [42].

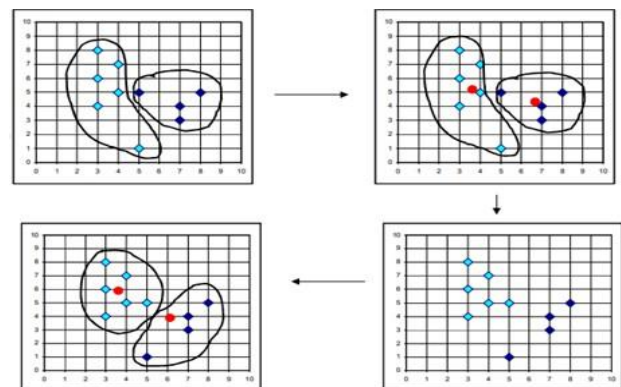


Fig 3. K-means clustering process

C. Support Vector Machine

The Support Vector Machine SVM has shown to be powerful for different classification problems [43]. It attempts to find the best hyperplane among classes by finding the number of points on the class descriptors' edge. The distance between classes is known as the margin. The better accuracy for the classification can be obtained when there is a higher margin. The data points on the border are called support vectors. SVM is used for both regression and classification problems [44,45]. This approach works well for solving a problem in the form of a linear and nonlinear dataset [46,47]. SVM algorithm uses various kernel types like linear radial basis function (RBF), Polynomial, and Sigmoid for a prediction model [48]. SVM operates on high-dimensional space for features and selects the best hyperplane for classifying data points into two classes. It is efficient for smaller and larger datasets that cannot be handled [49]. Figure 4 represents a simple example of the SVM algorithm using hyperplane for diagnosis of diabetes data [50].

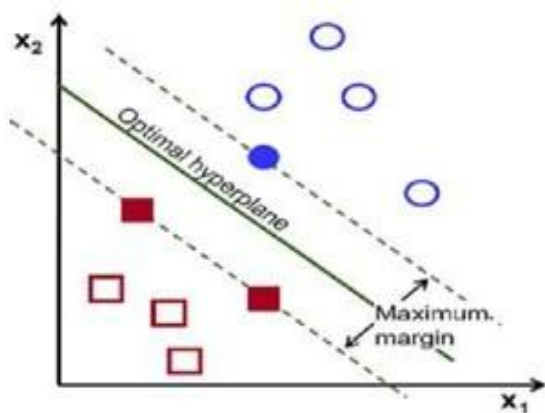


Fig 4. Classification of Diabetes by SVM

D. Naive Bayes algorithm

Naïve Bayes (NB) is a probability and statistical method based on classification algorithm. It is a standard algorithm in a machine learning applications due to its simplicity in enabling all features to contribute equally to the final decision. Computational efficiency is equal to this simplicity, making the NB approach exciting and appropriate for different fields. The NB classification's key component is prior, posterior, and class conditional probability [51]. This method has many benefits, like easy for massive datasets and very useful. It could be used for binary and multiclass problems of classification. A smaller amount of training data is required and can be used for both discrete and continuous data [52]. The application of this algorithm can be used for filtering spam emails and classifying documents [53].

E. Decision Tree Algorithm

The Decision Tree (DT) is an algorithm for supervised machine learning to solve regression and classification issues by repeatedly dividing data depending on a particular variable. The data is divided into the nodes, and the tree's leaf represents the final decisions. The purpose of the decision tree is to construct a model that can be used to predict the target variable

by learning simple decision rules derived from training data [54]. The tree is created by using the training data during a training process. The leaf nodes hold the name of the class, while a decision node is a non-leaf node. [55,56]. The decision tree manages categorical and numerical data. The nonlinear relationship among arguments does not impact the efficiency of the tree [57]. No preprocessing of the data is required. When the tree is repeatedly built, the probability of overfitting may occur [58]. Figure 5 represent a simple decision tree that consists of the root node, one child node, and three leafs node. One application of decision tree is used in the medical field. For example, using DT to detect breast cancer, the leaf nodes of the tree are split into two groups (Benign or Malignant). To decide whether the tumor is benign or malignant based on the Clump Thickness (CT), rules will be defined among the selected data set features. Figure 5 illustrates an example of using the DT algorithm to detect breast cancer [59].



Fig 5. Simple decision tree classifying breast cancer

F. Random Forest Classification

The random forest is an ensemble model that can also be used as a form of a nearest neighbor's predictor. The main idea about ensemble approaches is that a group of the model will form a powerful model. The random forest comes with a standard ML approach decision tree in ensemble terms. In this algorithm, an input is entered at the top, and the data is bucketed into smaller and smaller sets as it traverses through the tree. The random forest brings this notion to the next level by combining trees with an ensemble notion. A random forest classifier's benefit is that its running times are concise, unbalanced, and missing data can be treated. [60,61]. In the random forest, the new dataset or testing data is distributed to all created subtrees. Each decision subtree in the forest is permitted to decide the class of the dataset. Then the model will select the most appropriate class from the majority voting. Figure 6 illustrates the basic idea of the random forest algorithm for diagnosing heart disease [62,7].

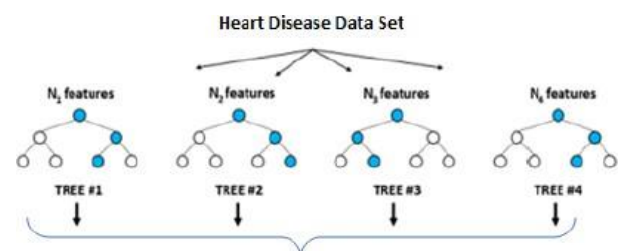


Fig 6. Random forest for classifying dataset of heart disease.

G. Logistic Regression

Logistic regression is a supervised learning algorithm used for solving a problem in the form of binary classification. Logistic regression is a mathematical model used with logistic function to model binary classification, and for logistic regression, there are several more complicated extensions. In essence, logistic regression is a regression model that predicts that a given data item or entry is likely to belong to a given class using the regression model. Logistic regression uses a sigmoid function to model the data, as illustrated in Figure 7 [63,64,65]. Logistic regression has many important points, such as implementation simplicity, computational effectiveness, training-based effectiveness, ease of regularization. For input features, no scaling is needed. Nevertheless, the ability to solve a nonlinear problem and susceptible to overfitting [8].

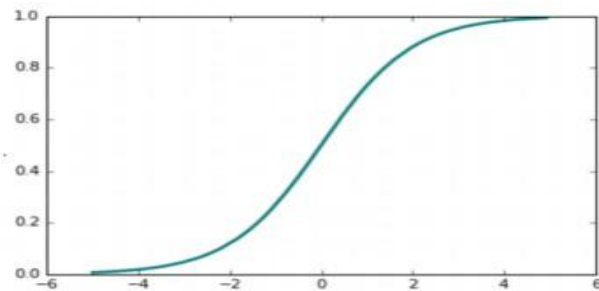


Fig 7. Logistic Function

H. Deep Learning

Deep learning is a subset of ML that involves a number of nonlinear transformations. It uses various algorithms that can learn how to view input data using several layers of processing with complicated structures. There are several deep learning architectures, like recurrent neural networks (RNN), deep auto-encoders, and convolutional neural networks (CNN). These algorithm have been used to different areas like natural language processing, speech recognition, and medicals field [66]. RNN is a deep learning algorithm with internal memory to keep the recent information. In general, memory units in RNN architecture have the links to themselves, which transfer information from the execution in the previous. RNN alters the nature of the current forward process to adjust to the context of current input [67]. The idea of the region of interest in MRI medical images typically distributed across several adjacent slices results in having similarities in successive slices. RNN is able to extract inter-slice contexts from the input slices as a form of sequential data [68]. A convolutional neural network (CNN) is another deep learning algorithm that used to solve complex problems and classification with a large number of datasets. The model consists mainly of four separate layers: convolution, max-pooling, fully connected, and the output layer [69]. Figure8 illustrates a CNN structure for classification of medical images having M classes accepting a sub image of 32×32 from the medical image. A feature map of different sizes is created by each convolutional layer and pooling layers will minimize the size of feature maps to be transferred to the next

layers, then the fully connected layer generate output as a prediction of the appropriate class [70].

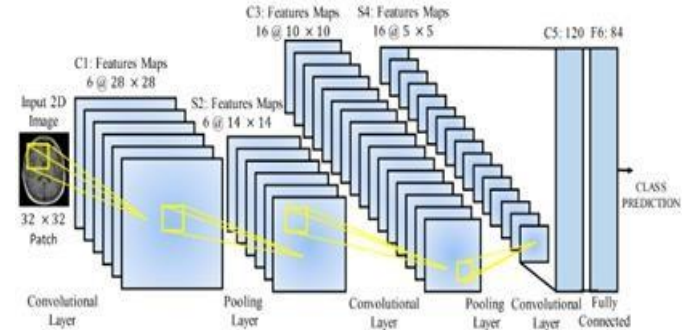


Fig 8. CNN Architecture to classify medical image

After explaining the algorithms mentioned in the above section, each algorithm's advantages and disadvantages are summarized in Table I.

TABLE I. THE ADVANTAGE AND DISADVANTAGE OF THE MENTIONED ALGORITHMS IN SECTION THREE

Alg.	Advantage	Disadvantage
SVM	<ul style="list-style-type: none"> It can manage linear and nonlinear data There is less probability of overfitting Scale-up with high dimensional data 	<ul style="list-style-type: none"> With a large dataset, performance goes down Difficult to select a suitable kernel function Do not work well when a dataset is noisy
Naïve Bayes	<ul style="list-style-type: none"> The easy for massive datasets Handles discrete and continuous data Can be used for both binary and multi-classification Not sensitive to irrelevant features 	<ul style="list-style-type: none"> Computationally-intensive especially, for models including many variables Sometimes, models that are appropriately trained and tuned outperform Naïve Bayes models as they are too simple
KNN	<ul style="list-style-type: none"> A model is cheap and easy to implement Used for both classification and regression. Works smoothly on multiclass problems 	<ul style="list-style-type: none"> Computation is very high Unknown records are relatively expensive to classify High sensitive to irrelevant
K-Mean	<ul style="list-style-type: none"> The ease of implementation More potent than hierarchical clustering when variables are large 	<ul style="list-style-type: none"> The K value estimation is difficult. Performance degrades when clusters are globular Sensitive to outlier and noise
Decision Tree	<ul style="list-style-type: none"> Used for both regression and classification Ease of handling numeric and categorical data 	<ul style="list-style-type: none"> Maybe overfitting occurs during repeated construction of the tree Larger trees get hard to interpret
Random Forest	<ul style="list-style-type: none"> Can be used for regression and classification problem Solve overfitting a problem in the decision tree 	<ul style="list-style-type: none"> Need much time for training Complexity
Logistic regression	<ul style="list-style-type: none"> Computational efficiency Ease of regularization For input features, no scaling is required 	<ul style="list-style-type: none"> Difficult to solve a nonlinear problem. Overfitting
Deep Learning	<ul style="list-style-type: none"> detects automatically features Can be applied on different data types. 	<ul style="list-style-type: none"> Need GPUs for training Highly costly to train because of complex data models

IV. LITERATURE REVIEW

In 2018, Thiruna Vukkarasu K. and et al, [71] used different classification algorithms to predict liver disease. Logistic regression, k-nearest neighbor, and support vector machines are the algorithms used for this work. The accuracy score and the confusion matrix are used to compare these algorithms for classification.

In 2018, Utomo Pujianto et al. [72] proposed a model based on the K-means and support vector machine to process Chronic Kidney Disease (CKD) data. The K-mean algorithm is used for the clustering process to determine the cluster of data, and then data will be classified by the SVM algorithm. Besides the SVM algorithm, different nonlinear kernels are used, such as polynomial kernels, RBF kernels, and sigmoid kernels. The best performance is obtained from the classifier method in two clusters, according to the results obtained.

In 2018, Naresh Khuriwal and Nidhi Mishra [73] proposed an adaptive ensemble voting method for diagnosed breast cancer. The purpose of this study is to compare and clarify how artificial neural networks and logistic algorithms provide the right solution when working with ML methods for breast cancer detection, even if the variables are minimized. In the proposed method, the Wisconsin Diagnosis Breast Cancer dataset is used. Compared to literature related work, it is shown that 98.50 % accuracy from another machine learning algorithm is achieved by the ANN method with a logistic algorithm.

In 2018, Ankita Tyagi and et al. [74] suggested a different machine learning methods for diagnosing and preventing thyroid disease. ML has an essential role in the disease detection process. This work deals with the analysis and classification models used in thyroid disease depend on data obtained from the UCI ML repository. A various machine learning algorithms are applied to identify the predicted risk on a patient's chance of getting a thyroid disease, which are SVM, K.N.N., and Decision Trees.

In 2019, Priyanka Sonar and et al. [75] enhanced an approach to analyzing a system that could predict a patient's diabetic risk level with better accuracy. The predicted model is dependent on several approaches, such as Naive Bayes, Decision Tree, and SVM algorithms. The obtained result shows that the accuracy model with SVM is 77.3%, a decision tree is 85%, and naïve Bayes is 77%.

In 2019, Noor Basha and et al. [76] proposed a heart disease model to detect and analyze heart syndrome with several parameters such as blood pressure, diabetes, heart rate, age, and gender. Data scientists will do predictive work on massive data for early analysis of heart syndrome to save patients' lives. In this analysis, many features are well-thought-out. Many ML algorithms are used to verify the syndrome's efficiency, checked with data prediction, do an evaluation, and detect heart diseases in patients.

In 2019, D. Selvathi and K. Suganya [77] proposed a model based on the ML algorithm to predict diabetic patients using thermographic eye images and incorporate the impact of the eye structure's thermal variation abnormality as a diagnostic imaging tool that is useful for clinical diagnosis by ophthalmologists. Thermal images are preprocessed, and then

texture feature depends on Gray Level Co-occurrence Matrix (GLCM) from gray images. Statistical features are extracted and classified from RGB images using a classifier with different feature combinations. SVM classifier is used to detect diabetic diseased eyes, and their performance is compared.

In 2019, Gokalp Cinarer and Bulent Gursel [78] studied magnetic resonance brain image characteristics such as n/a, gliomatosis, multifocal, and multicentric tumors classification methods were tested. The input images' statistical characteristic was analyzed during the classification process, and the information was regularly categorized into different categories. These data are tested with ML approaches such as random forest, KNN, SVM, and linear discriminant analysis. The best accuracy was found 90% with a support vector machine algorithm compared to all other methods tested in this work.

In 2019, Oumaima Terrada and et al. [79] defined a medical decision support system (MDSS) of atherosclerosis. The suggested model is capable of predicting heart disease using data of clinical. MDSS is focused on ML methods for the prediction of the existence and absence of atherosclerosis diseases, such as clustering algorithm K-medoids and k-means for classification, Artificial Neural Network (ANN), and K-Nearest Neighbor (KNN). The proposed model uses Cleveland's heart diseases a database for training and testing data. For each patient, this data is split into training and testing data for 13 characteristics. The system is tested using several similarity metrics, such as the correlation coefficient of Matthews (MCC).

In 2019, Rahma Atallah and Amjed Al-Mousa [80] presented an ensemble model based on the majority voting capable of predicting the potential presence of human heart disease. Detection is focused on simple, inexpensive medical experiments performed at any local clinic. The purpose of the model is to give precision to the Doctor's diagnosis, and the method is trained on the real-life data of patients. Also, there have more reliable solutions than providing just one method. The patient's model classification relies on the majority vote of multiple ML approaches.

In 2019, Tahira Islam Trishna and et al. [81] introduced an automatic diagnosis model for diagnosing hepatitis disease using several ML algorithms to detect hepatitis viruses on Weka software. Many ML algorithms, such as Bayes, random forest, and k-nearest, were used in the suggested method. Naïve Bayes' algorithm is applied to solve text classification problems while K-nearest neighbor and random forest use the regression and classification problem. Finally, each algorithm's experimental results are to calculate. Each method's accuracy is 93.2%, 98.6%, and 95.8% by using a ten-fold cross for naïve Bayes, random forest, and K-nearest, respectively.

In 2019, Nashat Alrefai [82], the Particle Swarm Optimization (PSO), and an ensemble learning method worked together to select features. The combination of four classifiers was used to explain the ensemble method: Support Vector Machine, K-Nearest Neighbor, Naïve Bayes, and Decision Tree (C4.5) used to generate fitness function used in PSO as an optimal solution to cover all the search space in the shortest time with the guarantee of selecting the best number of

meaningful genes that lead to improving the diagnosis of leukemia cancer.

In 2020, J.Neelaveni and Geetha Devasana [83] predicted a model based on the machine learning algorithm to detect Alzheimer's disease using psychological parameters such as age, the number of visits, MMSE, and education.

In 2020, Oyew O.A and Boyinbod O.K [84] developed an ensemble model that uses several machine learning algorithm Support Vector Machine (SVM), Decision Tree (DT) a Multiplayer Perceptron (MP)) to detect prostate cancer in men. The proposed method is measured by using sensitivity, specificity, and accuracy as efficiency metrics. The obtained test shows that the prediction of 98.09%, the specificity of 99.54%, and the accuracy of 99.06% is a relative enhancement on the current systems.

In 2020, Nikita Banerjee and Subhalaxmi Das [85] suggested a model for identifying lung cancer early on in a step. The model focuses on digital image processing and machine learning methods. Digital image processing is used for preprocessing and extract features from the image. The machine learning classification algorithms, such as Support Vector Machines, Random Forest, Artificial Neural Network (ANN) train on the extracted features. According to the classification results, the prediction is made whether the tumor is benign or malignant.

In 2020, N. Komal Kumar and et al. [86] suggested a model to predicate Cardio Vascular Disease (CVD) disease depend on the machine learning ML classification algorithms like the random forest, decision tree, logistic regression, SVM, and K.N.N. According to the obtained result, the random forest classifier algorithm has higher accuracy than the other classifier algorithm.

In 2020, Mahzabeen Emu and et al. [87] applied multiple ML approaches to enhance a non-invasive technique for detecting the level of liver fibrosis. Only finding the important features predicts the stages and generation of liver fibrosis rules for supporting doctors in patients' care as a clinically non-invasive technique. Also, the efficiency of ML algorithms such as random forest, MLP, and logistic regression are evaluated and compared for the feature sets that are complete and reduced.

In 2020, Dayanand Jamkhandikar and Neethi Priya [88], Algorithms of machine learning were used to model a thyroid patient's prediction and diagnosis. In the suggested model, an attempt is made to analyze naïve Bayes, k-nearest neighbor, and Support Vector Machine (SVM) for multiclass thyroid dataset classification. Different ML algorithms would be able to achieve better accuracy in disease prediction by comparative study.

In 2020, Vidya M and Maya V Karki [89] proposed automatic detection of skin lesions, and the suggested system passes from two steps. The first step, apply feature extraction using many algorithms like ABCD. Rule, GLCM and HOG. In this work, preprocessing is used to enhance skin lesions' appearance and visibility to eliminate artifacts, hair, and skin color. In the segmentation, the Geodesic Active Contour is applied to segment the lesion part separately. The ABCD The technique is applied for selecting features of the border, color. At the same time, GLCM with HOG was applied to extract textural features. In a second step, the extracted features are moved to the classifier algorithm to classify skin lesions between benign and melanoma based on several ML algorithms such as K.N.N., SVM, and Naïve Bayes classifier.

In 2020, Senthil Kumar Brindha and et al. [90] a model proposed to predict Early Gastric Cancer (EGC) factors from Mizo-ethnicity Lifestyle and diet features using supervised machine learning algorithms. Eighty cases and 160 controls have been chosen for this analysis, and a dataset that includes 11 features that are the main risk factors for gastric cancer has been selected for data mining. The learning curves indicate the best fit for the used dataset are logistic regression, naive Bayes, and multilayer perceptron.

In 2020, Zabirul Islam and et al [91], Suggested a Convolutional Neural Network and a Long Short-Term Memory were introduced to diagnose COVID-19 automatically from X-ray images. For deep feature extraction, CNN is used in this framework and long short-term memory LSTM is used for detection using the extracted feature. As a dataset in this method, a series of 4575 X-ray images, including 1525 images of COVID-19, was used. The experimental results show that the accuracy of the proposed system reached 99.4%.

In 2020, Halgurd S. Maghdid and et al [92], suggested an effective convolution neural network with AlexNet model for identifying coronavirus disease (COVID 19). This model uses two types of images X-Ray and CT image. In addition, the images (X-ray and CT) used during this analysis are obtained from different sources and they collectively make the radiographic images used for this work to identify COVID-19 publicly applicable to the research community. The experimental results indicate that the accuracy of the models up to 98% via pre-trained network and 94.1% accuracy by updated CNN.

TABLE II. SUMMARY OF THE WORK PERFORMED BY MOST OF THE RESEARCHERS REVIEWED IN THIS PAPER

Ref	Year	Disease	ML algorithms	Dataset	Results	Finding
[92]	2020	Covid 19	CNN	GitHub and Kaggle	Accuracy =94%	While the proposed model achieved a high detection accuracy of COVID-19, this does not indicate a production-ready solution in particular with the small number of images.
[89]	2020	Skin cancer	SVM and KNN	Collected from ISIC 2017 datasets with 1000 instance	Accuracy SVM=97.8% And KNN= 86.25	SVM classifier outperforms with 97.8 % accuracy and 0.94 AUC for textural features.
[88]	2020	Thyroid disease	Naïve Bayes, KNN, and SVM algorithm	UCI data repository site	Accuracy SVM =82%, Naïve Bayes =83% and KNN=85%	The treatment of thyroid patients is better cost and facilities management.

Ref	Year	Disease	ML algorithms	Dataset	Results	Finding
[87]	2020	liver fibrosis	Random forests, MLP, logistic regression algorithm.	Collected by Ain Shams University, Faculty Medicine at El Demerdash Hospital.	Accuracy random forest= 97.228% MLP= 98% Logistic regression= 97%	MLP obtained the highest accuracy rate of all the machine learning methods used in this study.
[86]	2020	Cardio Vascular	Random Forest, Decision Tree, Logistic Regression, SVM, and KNN algorithms.	NIDDK.	Accuracy Random forest=85.71%, decision tree =74.28%, Logistic Regression=74.28%, SVM=77.14 and KNN=68.57%	The random forest is exceeded by all the classifiers under the Analysis of classifying CVD patients.
[85]	2020	Lung cancer	SVM, Random Forest, ANN algorithms	UCI machine learning repository	Accuracy Random Forest =70%, SVM=80% and ANN=96%.	The model can differentiate between benign and malignant, and it can be seen that both texture and region-based artificial neural networks have more accuracy.
[84]	2020	Prostate cancer	SVM, Decision Tree, and MLP algorithm	github.com	accuracy of model 99.06%	Effective in detecting both non-prostate and prostate cases.
[83]	2020	Alzheimer	SVM, a decision tree algorithm	ADNI database	Accuracy SVM=85% and Decision Tree =83%	The model predicts the disease in the patient and distinguishes between cognitive impairment.
[82]	2019	Leukemia Cancer	SVM, KNN, Naïve Bayes (NB), and Decision Tree (C4.5) algorithms	Microarray dataset	The accuracy of classification on microarray datasets and the result of performance are superior to other individual classifiers.	The accuracy of microarray dataset classification and the result of performance is higher than that of other individual classifiers.
[81]	2019	Hepatitis (A, B, C, and E)	KNN, random forest, and naïve Bayes algorithm	Real data of hepatitis patients	Accuracy KNN=95.8%, random forest=98.6% and naïve Bayes =93.2%	The proposed system is appropriate for hepatitis virus diagnosis.
[80]	2019	Heart	KNN, Random Forest, Logistic Regression, and Ensemble Method algorithms	UCI dataset	Accuracy KNN=87%, Random Forest= 87%, Logistic Regression=87%, Hard Voting Ensemble Method=90%	The ensemble model has achieved higher accuracy and outperforms the accuracy of each classifier.
[79]	2019	Atherosclerosis	ANN, KNN, K-medoids, and K- mean algorithm	Cleveland heart diseases database at the University of California Irvine.	The overall accuracy of the system is 96%	The proposed system could help strengthen the effective atherosclerosis disease support system for medical diagnosis.
[78]	2019	Brain Tumors	SVM, KNN, RF, algorithm	Rembrandt image planes of a total of 33 patients, with an average of 20 images per patient. 30	Accuracy KNN= 87%, SVM=90%, LDA=83% and RF=83%	SVM is better accurate as compared to the other algorithm.
[77]	2019	Diabetic Eye	SVM algorithm	FLIR.com 283 eye thermal images	Accuracy SVM=86. 22%	The findings indicate that the custom IR thermal image system can be used as an effective method for the diagnosis of diabetic thermal eye disease, both non-invasive and non-contact.
[76]	2019	Heart Syndrome	KNN, Naïve Bayes, Decision tree, random forest, and SVM algorithm	Kaggale web site	Accuracy KNN=85% Decision tree=82% Random forest=81% Naïve Bayes=80% SVM=82%	KNN algorithm has a very good Accuracy score compared to other algorithms.
[75]	2019	Diabetes	Naïve Bayes, Decision Tree, and SVM algorithms	Global dataset	Accuracy decision tree = 85%, Naïve Bayes =77% And SVM= 77.3%	The decision tree more accurate than Naïve Bayes and SVM.
[74]	2018	Thyroid	SVM, KNN, Decision Trees algorithm	The UCI machine learning repository	Accuracy KNN=98.62%, SVM==99.63% DT=75.76%	The decreased number of parameters leads to the patient saving both money and time.
[73]	2018	Breast Cancer	ANN and the logistic algorithms	Wisconsin Breast Cancer Database.	The accuracy of the model =98.50%.	ANN is effective in analyzing critical human data and can do pre-diagnosis without special medical knowledge.
[72]	2018	Chronic Kidney	K-Means and SVM algorithm	UCI Machine Learning Website.	K = 2, accuracy= 100%. K = 3, accuracy= 98.86% K = 4, accuracy =98.9% and K = 5, accuracy=96.2%	best kernel method in the SVM is the RBF
[71]	2018	Liver	Logistic Regression, KNN, and SVM algorithms	Indian Liver Patient Dataset (ILPD).	KNN model, Accuracy= 73.97%. Logistic Regression model Accuracy= 73.97%. SVM Accuracy= 71.97%	KNN and Logistic Regression have high accuracy, but the highest sensitivity is logistic regression. Therefore, logistic regression is a more suitable model.

V. DISCUSSION

All explained papers in the previous section are related studies to identify diseases that used various machine learning algorithms in the period from 2018 to 2020. Table 2 presents a summary of nineteen previous works reviewed in this paper includes a type of disease, dataset, methods or algorithms, accuracy, and finding results for each study. Many algorithms are used in reviewed research, such as logistic regression, K nearest neighbor, support vector machine, K-means, decision tree, random forest, and an ensemble model. Also, these algorithms were applied to the standard datasets in a variety of diseases.

The authors in [71] used three different algorithms to diagnose liver diseases. The logistic regression algorithm's accuracy and sensitivity results are being selected as the best algorithm among the others. In [74] and [76], KNN and SVM algorithms were applied to two different datasets and features. The result in [74] showed that the SVM algorithm is better than KNN. On the other hand, in [76] KNN algorithm override SVM by about 3%. In research [76], the decision tree's accuracy is better than the random forest, around 1%. Nevertheless, in [86], the random forest's accuracy exceeds around 12% than the decision tree. Researchers in [73] and [80] used an ensemble model, where the accuracy was 98.50% and 90%. In general SVM and Logistic regression is the only binary classifier whereas decision tree, random forest, KNN, naïve Bayes, and CNN are capable of classifying into more than two classes and K-means for multi clustering. All algorithms work well when the size of the dataset is small, while for a large dataset is better to use deep learning algorithms like CNN. To conclude, the algorithm's accuracy depends on the size of the dataset and the number of features, and the results of the ensemble model are better than using only a single model.

VI. CONCLUSION

Machine learning has emerged with the medical sector for providing tools and analyzing the data related to diseases. Therefore machine learning algorithms play an essential role in achieving the early detection of diseases. This paper provided a review of a different machine learning algorithms for predicting diseases, and standard datasets have been used in various diseases like liver, chronic kidney, breast cancer, heart syndrome, brain tumors, and many other diseases. A list of results discovered by researchers has been tabulated to detect the diseases by various ML algorithms. After comparing nineteen papers for different models that predicted diseases, it illustrated that many algorithms have good accuracy for predicting SVM, K-nearest neighbors, random forest, and the decision tree. Nevertheless, the accuracy of the same algorithm may differ from one dataset to another because many important factors affect the model's accuracy and performance, like datasets, feature selection, and the number of features. Another critical point found in this review is that the model's accuracy and performance can be increased by using a different algorithm to produce one ensemble model.

REFERENCES

- [1] Shaik Razia, P.Swathi Prathyusha, N.Vamsi Krishna, N.Sathya Sumana, "A Review on Disease Diagnosis Using Machine Learning Techniques," *International Journal of Pure and Applied Mathematics*, Volume 117, No. 16, 2017..
- [2] Diyar Qader Zeebaree, Habibollah Haron, Adnan Mohsin Abdulazeez, and Dilovan Asaad Zebari, "Trainable Model Based on New Uniform LBP Feature to Identify the Risk of the Breast Cancer". In 2019 International Conference on Advanced Science and Engineering (I.C.O.A.S.E.) (pp. 106-111). IEEE, April 2019.
- [3] Iswanto Iswanto, E. Laxmi Lydia, K. Shankar, Phong Thanh Nguyen, Wahidah Hashim, Andino, "Identifying Diseases and Diagnosis using Machine Learning", *International Journal of Engineering and Advanced Technology*, Vol. 8, August 2019.
- [4] Reem A. Alassaf, and et al, "Preemptive Diagnosis of Chronic Kidney Disease Using Machine Learning Techniques", *International Conference on innovations in Information Technology (I.T.)*, IEEE, 2018.
- [5] Gopi Battineni, Getu Gamo Sagaro, Nalini Chinatalapudi and Francesco Amenta, "Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis", *Journal of Personalized Medicine*, 2020.
- [6] Diyar Qader Zeebaree, Habibollah Haron, Adnan Mohsin Abdulazeez and Dilovan Asaad Zebari, "Machine learning and Region Growing for Breast Cancer Segmentation", *International Conference on Advanced Science and Engineering*, IEEE, 2019.
- [7] Hossam Meshref, "Cardiovascular Disease Diagnosis: A Machine Learning Interpretation Approach", *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 12, 2019.
- [8] Joel Jacob, Joseph Chakkalakal Mathew, Johns Mathew and Elizabeth Issac, "Diagnosis of Liver Disease Using Machine Learning Techniques", *International Research Journal of Engineering and Technology*, Vol. 05, Issue: 04, Apr-2018.
- [9] Siddhesh Iyer, Shivkumar Thevar, Priyamurgan Guruswamy, Ujwala Ravale, "Heart Disease Prediction Using Machine Learning", *International Research Journal of Modernization in Engineering Technology and Science*, Vol. 02, Issue:07, July 2020.
- [10] Dilovan Asaad Zebari, Diyar Qader Zeebaree, Adnan Mohsin Abdulazeez, Habibollah Haron, and Haza Nuzly Abdull Hamed, "Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images, *IEEE Access*, Vol.8, 2020.
- [11] Sneha Grampurohit and Chetan Sagarnal, "Disease Prediction using Machine Learning Algorithms" *International Conference for Emerging Technology (I.N.C.E.T.)*, Belgaum, India. Jun 5-7, 2020.
- [12] Pahulpreet Singh Kohli and Shriya Arora, "Application of Machine Learning in Disease Prediction", *International Conference on Computing Communication and Automation (I.C.C.A.A.)*, IEEE, 2018.
- [13] Adnan Mohsin Abdulazeez, Baraa Wasfi Salim, Diyar Qader Zeebaree, and Dana Doghramachi, "Comparison of VPN Protocols at Network Layer Focusing on Wire Guard Protocol", *iJIM*, Vol. 14, No. 18, 2020.
- [14] Berina Alić, Lejla Gurbeta and Almir Badnjević, "Machine Learning Techniques for Classification of Diabetes and Cardiovascular Diseases", *6th Mediterranean Conference on Embedded Computing*, 2017.
- [15] Divya Jain and Vijendra Singh, "Feature selection and classification systems for chronic disease prediction: A review", *Elsevir*, 2018.
- [16] Huseyin Polat, Homay Danaei Mehr and Aydin Cetin, "Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods", *published online springer*, 2017.
- [17] Golmei Shaheamlung, Harshpreet Kaur and Mandeep Kaur, "Survey on machine learning techniques for the diagnosis of liver disease", *International Conference on Intelligent Engineering and Management*, IEEE, 2020.
- [18] Diyar Qader Zeebaree, AdnanMohsin Abdulazeez, Dilovan Asaad Zebari, Habibollah Haron and Haza Nuzly Abdull Hamed, "Multi-Level Fusion in Ultrasound for Cancer Detection Based on Uniform LBP Features", *Computers,Materials & Continua*, Vol. 66, No.3, pp. 3363–3382, 2021.

- [19] Diyar Qader Zeebaree, Habibollah Haron, Adnan Mohsin Abdulazeez, "Gene Selection and Classification of Microarray Data Using Convolutional Neural Network", International Conference on Advanced Science and Engineering, IEEE, 2018.
- [20] Aurélien Géron, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow", Kiwisoft S.A.S. All rights reserved, 2019.
- [21] Fahad Kamal Alsheref and Wael Hassan Gomaa, "Blood Diseases Detection using Classical Machine Learning Algorithms, International Journal of Advanced Computer Science and Applications, Vol. 10, No. 7, 2019.
- [22] G. Nagarajan, A.P.Mohan Raju, V. Logeshwaran, K. Nandhakumar and S. Naveenkumar, "Effective Prediction Model for Heart Disease Using Machine Learning Algorithm", International Journal of Engineering Research & Technology (I.J.E.R.T.), ISSN: 2278-0181, Special Issue – 2019.
- [23] Tejaswini Untawale, "A REVIEW ON MACHINE LEARNING TECHNIQUES TO PREDICT DISEASES", International Research Journal of Modernization in Engineering Technology and Science, Vol.02, Issue: 07, July 2020.
- [24] Shruti Katiyar and Shruti Jain, "Predictive Analysis on Diabetes, Liver and Kidney Diseases using Machine Learning", International Journal for Research in Applied Science & Engineering Technology (I.J.R.A.S.E.T.), ISSN: 2321-9653; I.C. Value: 45.98; S.J. Impact Factor: 7.429, Volume 8 Issue V May 2020.
- [25] AJAY SHRESTHA and AUSIF MAHMOOD, "Review of Deep Learning Algorithms and Architectures", IEEE, 2019.
- [26] Yuanyuan Jia, Zhiren Tan and Junxing Zhang, "DKDR: An Approach of Knowledge Graph and Deep Reinforcement Learning for Disease Diagnosis", IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking, 2019.
- [27] Meherwar Fatima and Maruf Pasha, "Survey of Machine Learning Algorithms for Disease Diagnostic", Journal of Intelligent Learning Systems and Applications, 2017.
- [28] Dildar Masood Abdulqader, Adnan Mohsin Abdulazeez, Diyar Qader Zeebaree, "Technology Reports of Kansai University", ISSN: 04532198, Volume 62, Issue 03, April 2020.
- [29] N.D.K.G. Dharmasiri and S. Vasanthapriyan, "Approach to Heart Diseases Diagnosis and Monitoring through Machine Learning and iOS Mobile Application", International Conference on Advances in I.C.T. for Emerging Regions, IEEE, pp. 407 – 412, 2018.
- [30] Adel Al-Zebari and Abdulkadir Sengur, "Performance Comparison of Machine Learning Techniques on Diabetes Disease Detection", IEEE, 2019.
- [31] Animesh Urgiye and Rupali Bhartiya, "Review of Machine Learning Algorithm on Cancer Data Set", International Journal of Scientific Research & Engineering Trends, Volume 6, Issue 6, Nov-Dec-2020.
- [32] Zoltan Geler, Vladimir Kurbalija, Mirjana Ivanovi, Milo's Radovanovi, "Weighted kNN and constrained elastic distances for time-series classification", Expert Systems With Applications, Elsevier, 2020.
- [33] M. SUSHMA SRI, "A Review On Object Tracking Based On K.N.N. Classifier", International Research Journal of Engineering and Technology, Vol 06, Issue: 12, December 2019.
- [34] Nurul Ezzati Md Isal, Amiza Amir, Mohd Zaizu Ilyas, and Mohammad Shahrazel Razalli, "The Performance Analysis of K-Nearest Neighbors (K-NN) Algorithm for Motor Imagery Classification Based on E.E.G. Signal, M.A.T.E.C. Web of Conferences, 2017.
- [35] Rajesh N, T Maneesha, Shaik Hafeez, Hari Krishna, "Prediction of Heart Disease Using Machine Learning Algorithms", International Journal of Engineering & Technology, P 363-366, 2018.
- [36] S. NagaMallik Raj, N. Thirupathi Rao, Venkata Naresh Mandhala and Debnath Bhattacharyya, "Machine Learning Algorithms To Enhance Security In Wireless Network", Journal of Critical Reviews, ISSN- 2394-5125, Vol 7, Issue 14, 2020.
- [37] Diyar Qader Zeebaree, Habibollah Haron, Adnan Mohsin Abdulazeez and Subhi R. M. Zeebaree, "Combination of K-means clustering with Genetic Algorithm: A review", International Journal of Applied Engineering Research, Vol. 12, No. 24, pp. 14238-14245, 2017.
- [38] Dawlat Mustafa Sulaiman, Adnan Mohsin Abdulazeez, Habibollah Haron and Shereen S. Sadiq, "Unsupervised Learning Approach-Based New Optimization K-Means Clustering for Finger Vein Image Localization", International Conference on Advanced Science and Engineering, IEEE, 2019.
- [39] Alan Fuad Jahwar, Adnan Mohsin Abdulazeez, "META-HEURISTIC ALGORITHMS FOR K-MEANS CLUSTERING: A REVIEW", Palarch's Journal of Archaeology of Egypt/Egyptology, 2021.
- [40] N.Valarmathy and S.Krishnaveni, "Performance Evaluation and Comparison of Clustering Algorithms used in Educational Data Mining", International Journal of Recent Technology and Engineering, ISSN: 2277-3878, Vol-7, Issue-6S5, April 2019.
- [41] Susmita Ray, "A Quick Review of Machine Learning Algorithms", International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 14th -16th Feb 2019.
- [42] Shraddha Shukla and Naganna S. , "A Review ON K-means DATA Clustering APPROACH", International Journal of Information & Computation Technology, Vol. 4, No. 17 , pp. 1847-1860, 2014.
- [43] Adnan Mohsin Abdulazeez, Maryam Ameen Sulaiman, Diyar Qader Zeebaree, "Evaluating Data Mining Classification Methods Performance in Internet of Things Applications", Journal Of Soft Computing And Data Mining, Vol.1, NO. 2, pp. 11-25, 2020.
- [44] Amani Yahyaoui, Akhtar Jamil, Jawad Rasheed, Mirsat Yesiltepe, "A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques", IEEE, 2019.
- [45] Priyanka Lodha, Ajay Talele and Kishori Degaonkar , "Diagnosis of Alzheimer's Disease using Machine Learning", IEEE, 2018.
- [46] Zainuri Saringat, Aida Mustapha, R. D. Rohmat Saedudin, Noor Azah Samsudin, "Comparative analysis of classification algorithms for chronic kidney disease diagnosis", Bulletin of Electrical Engineering and Informatics, Vol. 8, No. 4, pp. 1496-1501, December 2019.
- [47] Jaimin Shah and Raj Patel, "Classification techniques for Disease detection using Big-data", 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques, IEEE, 2019.
- [48] Pavithra V and Jayalakshmi V, "Review of Feature Selection Techniques for Predicting Diseases", Proceedings of the Fifth International Conference on Communication and Electronics Systems, IEEE, 2020.
- [49] Shalini M and S. Radhika, "Machine Learning techniques for Prediction from various Breast Cancer Datasets", IEEE, 2020.
- [50] Ahmed Hamza Osman and Hani Moetque Aljahdali, "Diabetes Disease Diagnosis Method based on Feature Extraction using K-SVM", International Journal of Advanced Computer Science and Applications, Vol. 8, No. 1, 2017.
- [51] Aji Prasetya Wibawa and et al, "Naïve Bayes Classifier for Journal Quartile Classification", IJES – Vol. 7, No. 2, 2019.
- [52] Shahadat Uddin, Arif Khan, Md Ekramul Hossain and Mohammad Ali Moni3, "Comparing different supervised machine learning algorithms for disease prediction", B.M.C. Medical Informatics and Decision Making, 2019.
- [53] Raparathi Yaswanth and Y. Md. Riyazuddin, "Heart Disease Prediction using Machine Learning Techniques", International Journal of Innovative Technology and Exploring Engineering, Vol. 9, Issue-5, March 2020.
- [54] Yash Jayesh Chauhan, "Cardiovascular Disease Prediction using Classification Algorithms of Machine Learning", International Journal of Science and Research (I.J.S.R.), ISSN: 2319-7064, 2018.
- [55] Adel Sabry Eesa, Zeynep Orman and Adnan Mohsin Abdulazeez, "A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems", Expert Systems with Applications, Elsevier, pp. 2670-2679, 2015.
- [56] Adel Sabry Eesa, Zeynep O.R.M.A.N., Adnan Mohsin Abdulazeez , "A new feature selection model based on ID3 and bees algorithm for intrusion detection system", Turkish Journal of Electrical Engineering & Computer Sciences, 2015.
- [57] Xiaolu Tian and et al, "Using Machine Learning Algorithms to Predict Hepatitis B Surface Antigen Seroclearance", Hindawi Computational and Mathematical Methods in Medicine Volume, 2019.

- [58] A.Deva Kumar, Josephine Prem Kumar, V.S Prakash and Divya K.S, "Supervised Learning Algorithms: A Comparison", *Kristu Jayanti Journal of Computational Sciences*, Vol. 1, Issue 1, pp. 01-12, 2020.
- [59] Autsuo Higa, "Diagnosis of Breast Cancer using Decision Tree and Artificial Neural Network Algorithms", *International Journal of Computer Applications Technology and Research*, Vol. 7, Issue 01, pp.23-27, 2018.
- [60] S. Venkata Lakshmi, M. K. Meena and N. S. Kiruthika, "Diagnosis of Chronic Kidney Disease using Random Forest Algorithms", *International Journal of Research in Engineering, Science and Management*, Volume-2, Issue-3, March-2019.
- [61] K.VijayaKumar, B.Lavanya, I.Nirmala, S.Sofia Caroline, "Random Forest Algorithm for the Prediction of Diabetes", *Proceeding of International Conference on Systems Computation Automation and Networking*, 2019.
- [62] Kaitlin Kirasich, Trace Smith and Bivin Sadler, "Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets", *S.M.U. Data Science Review*, Vol. 1 | No. 3, 2018.
- [63] Ram MurtiRawat, Shivam Pancha, Vivek Kumar Singh and Yash Panchal, "Breast Cancer Detection Using K-Nearest Neighbors, Logistic Regression and Ensemble Learning", *Proceedings of the International Conference on Electronics and Sustainable Communication Systems*, IEEE, 2020.
- [64] Gulshan Kumar, "A Review On Machine Learning Concept And Its Algorithms", *Alochana Chakra Journal*, ISSN NO: 2231-3990, 2020.
- [65] Dilovan Asaad Zebari, Diyar Qader Zeebaree, Jwan Najeeb Saeed, Nechirvan Asaad Zebari, Adel AL-Zebari, *Image Steganography Based on Swarm Intelligence Algorithms: A Survey*, *Test Engineering & Management*, Vol. 83, pp. 22257 – 22269, 2020.
- [66] Tarek M. Hassan I, Mohammed Elmogy and El-Sayed Sallam, "Diagnosis of Focal Liver Diseases Based on Deep Learning Technique for Ultrasound Images", *published online*, *springer*, 2017.
- [67] Siwei Ma, Xinfeng Zhang, Chuanmin Jia, Zhenghui Zhao, Shiqi Wang, and Shanshe Wang, "Image and Video Compression with Neural Networks: A Review", *IEEE Transactions On Circuits And Systems For Video Technology*, 2019.
- [68] Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He and Paul Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges", *Journal of Digital Imaging*, Springer, 2019.
- [69] Uday Pratap Singh, Siddharth Singh Chouhan, Sukirty Jain, And Sanjeev Jain", *IEEE*, 2019.
- [70] Syed Muhammad Anwar, Muhammad Majid, Adnan Qayyum, Muhammad Awais, Majdi Alnowami and Muhammad Khurram Khan, "Medical Image Analysis using Convolutional Neural Networks: A Review", *Journal of Medical Systems*, 2019.
- [71] Thirunavukkarasu K, Ajay S. Singh, Md Irfan and Abhishek Chowdhury, "Prediction of Liver Disease using Classification Algorithms", *International Conference on Computing Communication and Automation (I.C.C.C.A.)*, IEEE, 2018.
- [72] Utomo Pujianto, Nur A'yuni Ramadhani and Aji Prasetya Wibawa, "Support Vector Machine with Purified K-Means Clusters for Chronic Kidney Disease Detection", *The 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT) 2018*
- [73] Naresh Khuriwal and Nidhi Mishra, "Breast Cancer Diagnosis Using Adaptive Voting Ensemble Machine Learning Algorithm", *IEEE*, 2018.
- [74] Ankita Tyagi, Ritika Mehra and Aditya Saxena, "Interactive Thyroid Disease Prediction System Using Machine Learning Technique", *5th IEEE International Conference on Parallel, Distributed and Grid Computing (PDGC-2018)*, Solan, India, 20-22 Dec 2018.
- [75] Priyanka Sonar and K. JayaMalini, "Diabetes Prediction Using Different Machine Learning Approaches", *Proceedings of the Third International Conference on Computing Methodologies and Communication (I.C.C.M.C. 2019) IEEE*, 2019.
- [76] Noor Basha, Ashok Kumar P S, Gopal Krishna C and Venkatesh P, "Early Detection of Heart Syndrome Using Machine Learning Technique", *4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECOT)*, 2019.
- [77] D. Selvathi and K. Suganya, "Support Vector Machine Based Method for Automatic Detection of Diabetic Eye Disease using Thermal Images", 2019.
- [78] Gokalp Cinaree and Bulent Gursel Emiroglu, "Classification of Brain Tumors by Machine Learning Algorithms", *IEEE*, 2019.
- [79] Oumaima Terrada, Bouchaib Cherradi, Abdelhadi Raihani and Omar Bouattane, "Classification and Prediction of atherosclerosis diseases using machine learning algorithms", *IEEE*, 2019.
- [80] Rahma Atallah and Amjed Al-Mousa, "Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method", *IEEE*, 2019.
- [81] Tahira Islam Trishna, and et al, "Detection of Hepatitis (A, B, C and E) Viruses Based on Random Forest, K-nearest and Naïve Bayes Classifier", *10th I.C.C.C.N.T. 2019 July 6-8, 2019, I.I.T. – Kanpur Kanpur, India*, 2019.
- [82] Nashat Alrefai, "Ensemble Machine Learning for Leukemia Cancer Diagnosis based on Microarray Datasets", *International Journal of Applied Engineering Research* ISSN 0973-4562 Volume 14, Number 21, pp. 4077-4084, 2019.
- [83] J.Neelaveni and Geetha Devasana, "Alzheimer Disease Prediction using Machine Learning Algorithm", *6th International Conference on Advanced Computing & Communication Systems (I.C.A.C.C.S.)*, 2020.
- [84] Oyewo O.A and Boyinbode O.K, "Prediction of Prostate Cancer using Ensemble of Machine Learning Techniques", *I.J.A.C.S.A.) International Journal of Advanced Computer Science and Applications*, Vol. 11, No. 3, 2020.
- [85] Nikita Banerjee Subhalaxmi Das, *Prediction Lung Cancer– In Machine Learning Perspective*, *IEEE*, 2020.
- [86] N. Komal Kumar, G.Sarika Sindhu, D.Krishna Prashanthi, A.Shaeen Sulthana, "Analysis and Prediction of Cardio Vascular Disease using Machine Learning Classifiers", *6th International Conference on Advanced Computing & Communication Systems (I.C.A.C.C.S.)*, 2020.
- [87] Mahzabeen Emu, Farjana Bintay Kamal, Salimur Choudhury and Thiago E. Alves de Oliveira, "Assisting the Non-invasive Diagnosis of Liver Fibrosis Stages using Machine Learning Methods", *IEEE*, 2020.
- [88] Dayanand Jamkhandikar and Neethi Priya, "Thyroid Disease Prediction Using Feature Selection And Machine Learning Classifiers", *The International Journal of analytical and experimental modal analysis*, ISSN NO: 0886-9367, 2020.
- [89] Vidya M and Maya V Karki, "Skin Cancer Detection using Machine Learning Techniques", *IEEE*, 2020.
- [90] Senthil Kumar Brindha and et al, "Data Mining for Early Gastric Cancer Etiological Factors from Diet-Lifestyle Characteristics", *Proceedings of the International Conference on Intelligent Computing and Control Systems*, *IEEE*, 2020.
- [91] Halgurd S. Maghdid, Aras T. Asaad, Kayhan Zrar Ghafoor, Ali Safaa Sadiq and Muhammad Khurram Khan, "Diagnosing COVID-19 Pneumonia from X-Ray and CT Images using Deep Learning and Transfer Learning Algorithms", *arXiv preprint arXiv:2004.00038*, 2020.
- [92] Md. Zabirul Islam, Md. Milon Islam and Amanullah Asraf, "A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images", *Elsevier*, 2020.