Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

A Study of Classification Model in Medical Health Applications

Raja Intan Sariah Raja Mahmood[1], Mokhairi Makhtar[2], Hafizan Juahir[3], Nadiana Ariffin[4], Aceng Sambas[5], Rosaida Rosly[6*], Yousef A. Baker El-Ebiary[7]

[1]*Politeknik Kuala Terengganu and Faculty of Informatics & Computing, Universiti Sultan Zainal Abidin, Malaysia.*
[2]*Faculty of Informatics and Computing, and Artificial Intelligence for Sustainability and Islamic Research, Universiti Sultan Zainal Abidin,Terengganu, Malaysia. mokhairi@unisza.edu.my*
[3]*East Coast Environmental Research Institute, Universiti Sultan Zainal Abidin, Malaysia. hafizanjuahir@unisza.edu.my*
[4]*Politeknik Kuala Terengganu and East Coast Environmental Research Institute, Universiti Sultan Zainal Abidin, Malaysia.*
[5]*Faculty of Informatics and Computing, and Artificial Intelligence for Sustainability and Islamic Research, Universiti Sultan Zainal Abidin,Terengganu, Malaysia. acengsambas@unisza.edu.my*
[*6]*Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, Terengganu, Malaysia. rosaida@umt.edu.my*
[7]*Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia. yousefelebiary@unisza.edu.my*

Medical health sector is one the largest industry worldwide. Nowadays, the emergence of coronavirus (COVID-19) presents an important alert to the healthcare sector. This threatening virus is a concern because this virus is spreading fast and has caused many deaths. The goal of this study is to analyze the classification algorithms that have been utilized for various diseases in past research. This will be considered and explored further for future analysis of COVID-19 that could help researchers manage COVID-19 infections using the discussed data mining techniques. In this study, there are 60 studies have been included from a variety of sources such as iGATE’s, Elsevier, ResearchGate, Springer, BMC Public Health Journal, J. Health Engineering, International journal computing Technology and other etc. With data mining classification algorithms, the researcher can easily do the classification, prediction, clustering and data filtering from various data sources, especially in the healthcare system. The classification algorithms such as Naïve Bayes, support vector machine (SVM), multilayer perceptron (MLP), J48, K-nearest neighbor (KNN), decision tree, random forest and logistic regression were listed as the most applied algorithms in the past research. The naïve bayes, support vector machine (SVM), random forest, multilayer perceptron (MLP) and J48 algorithms were identified as the top five most utilized algorithms. Therefore, this study indicates that these classification algorithms are suitable for identifying, classifying and predicting COVID-19. In conclusion, this review paper will guide the researcher for future research and other research community as well on the upcoming development of machine learning for the medical health sector, especially COVID-19.

Keywords: Mers-Cov, COVID-19, Data Mining, Deep Learning, Classification Algorithm, Accuracy.

1. Introduction

1.1 Background

Today, the healthcare sector is one of the fastest-growing industries globally. Database in the healthcare sector is massive and keeps evolving daily, containing a large amount of collected data [1]. In 2020, heart disease, diabetes, kidney diseases, stroke and cancer were listed as the top global causes of death [2]. At the end of 2019, all healthcare sectors worldwide were facing a threat from coronavirus, which spread fast and caused many deaths. The coronavirus (COVID-19) disease was declared an international alert in 2020 [3]. COVID-19 was first discovered in early December 2019 and was officially declared a pandemic on the 11th of March 2020 [4]. The number of infected people has been increasing, with over 4.4 million cases reported in August 2021, bringing the cumulative number of global cases to over 206 million [4]. The arising of COVID-19 is triggered by severe acute respiratory syndrome (SARS-Cov-2) and keeps spreading globally [3]. The infected individual may have a few COVID-19 symptoms such as fevers, dry cough, tiredness, headaches, aches and pains or loss of taste [5]. Apart from that, others may experience serious symptoms such as difficulty breathing, pneumonia and organ failure that can cause death [6].

1.2 Motivation

As the patient population grows, so does the medical database. Efficient diagnosis and prognosis of the diseases are required to reduce the burden on health care while providing the best possible care to patients [7]. Therefore, data mining can help detect the early symptoms of COVID-19 and other diseases. Data mining techniques are widely applied and used in various contexts and fields, especially in the healthcare system [8]. We can easily classify, predict, cluster, and filter data with data mining algorithms. Predictive models that combine different variables or characteristics to determine a person's risk of becoming infected or experiencing adverse effects can help medical staff test patients in allocating limited health care resources [7].

1.3 Objective

The main focus of this study is to review previous research in various diseases conducted by other researchers to identify the appropriate machine learning algorithm that will be applied for the next research, which is analyzing, classifying and predicting COVID-19.

1.4 Major Challenge

As this study's results will be considered for future research, there are a few obstacles researchers had to face, such as minimum sources of journals or articles related to data mining involving COVID-19. Other than that, it is also difficult to find the algorithm with the best accuracy value because different datasets show different accuracy values when using the same algorithm.

2. Data Mining

In the 1990s, a new technology called data mining was introduced. Today, data mining plays an important role in helping many researchers to predict diseases [9]. This technology includes a variety of new concepts such as databases, artificial intelligence, machine learning and others [10]. Note that there are various meanings of data mining have been laid forth by various researchers [11]. Data mining is one of the techniques of knowledge discovery in a database to gather meaningful information, and other researchers have developed and applied several data mining techniques [12].

Data mining includes the analysis and prediction of data and extends beyond the collection and management of data. To manage the data, several data mining algorithms can be used, which are the prediction and classification methods. Much of the research work in data mining has gone into improving predictive accuracy by applying data mining techniques.

2.1 Prediction Model

A prediction model, or predictive analytics, is a statistical analysis that uses data mining, machine learning and algorithm. When using historical datasets, prediction models will help identify behaviour patterns and trends to predict future scenarios [13]. Apart from that, organizations and groups in various fields such as education, medicine, research and consultation-based company are practicing predictive models and forecasting, especially in decision-making and planning [13].

2.2 Classification Algorithm

Classification is a crucial method in which data are divided into a certain number of classes [14][15]. It is also used to categorize data into categories and classes [16]. Some of the classification algorithms, such as support vector machine (SVM), k-nearest neighbor (KNN), decision tree, random forest and Naïve Bayes algorithms, generally utilize the base classifiers [11]. In supervised learning, the methods generally used include artificial neural network (ANN), SVM, decision tree, Naïve Bayes, J48, logistics regression and others [17]. Note that a particular classifier may be much better compared to others for a particular dataset, but another classifier can perform much better for several other datasets [17].

2.3 Deep Learning

Deep learning becomes part of a machine learning algorithm influenced by the structures and functions of ANN [18]. Nowadays, deep learning is gaining much popularity due to its superiority in terms of accuracy when trained with a huge amount of data, especially involving critical domains such as medical applications, banking, and information retrieval. Apart from that, deep learning plays an important role in creating solutions for massive data, as data comes in many forms, which can be varied and evolve. Therefore, deep learning comprises multiple levels of features [19]. The output is produced when the features are found and created on numerous levels.

Each level represents abstract features discovered from the features depicted in the previous level, thereby causing the level of abstraction to increase at each level. This type of learning enables one to discover and represent higher-level abstractions. One of the deep learning algorithms is neural networks, which are characterized by many layers consisting of features that produce the output [19] [20]. Other than that, deep learning can help solve many machine learning problems, especially in data classification. It can be applied to any data type, such as sound, text, images, time series, and video. The deep learning architectures have performed significantly better than modern methods and achieved high-level performances [6]. Regardless, the issues related to constructing multi-classifiers based on deep learning, as suggested by [21], will be considered.

3. Research Strategy

This paper contains articles and journals on coronavirus (COVID-19), Mers-Cov, heart diseases, diabetes, coronary artery diseases, breast cancer, hepatitis, thalassemia, liver and kidney diseases. Various data mining techniques are used to identify information from the clinical dataset.

In past studies, various classification algorithms and methods have been used to obtain the best result (refer to Table 1). In this review, the researcher has referred to thirty-six research articles and journals related to medical health issues to obtain results, as in Table 2. In general, more than 40 studies have been included in different medical science and computing journal. Figure 1 shows the detailed analysis process with several selection criteria that are also considered to ensure that the result is reliable and reproducible and to minimize error (refer to Table 1 for the Inclusion and Exclusion Criteria).

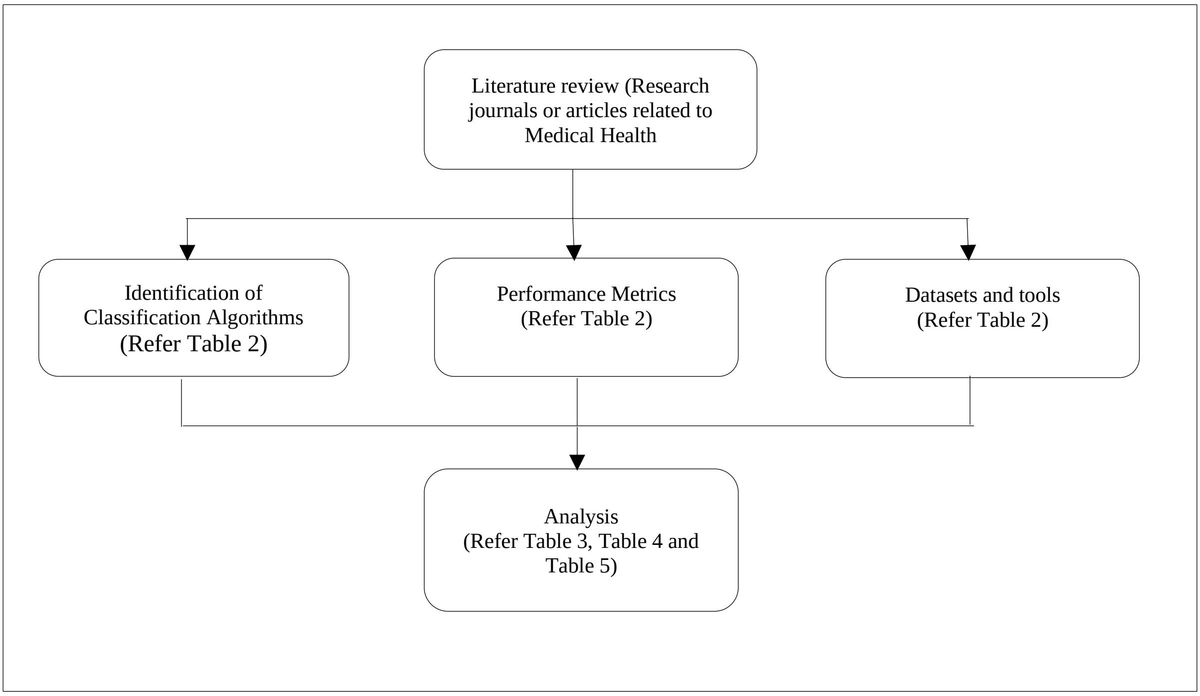

Figure 1. the Analysis Process Approach

Table 1. Inclusions and Exclusions Criteria

Criteria	Inclusion	Exclusion
Year	Research and publications from 2015 to 2022	Research and publications before 2015
Topic	Articles showing content related to Machine Learning	Articles are not related to Machine Learning
Issue	Sources related to Medical Health issues	Sources are not related to Medial Health issue
Language	Articles written in English	Article are not written in English

4. Result and Discussion

This section explains the classification algorithms implemented in previous studies, which cover diseases such as coronavirus (COVID-19), Mers-Cov, heart diseases, diabetes, coronary artery diseases, breast cancer, hepatitis disease, thalassemia, liver diseases and kidney diseases. Table 2 demonstrates the comparative study on the machine learning algorithms that were applied in classifying and diagnosing various diseases. Each of the classification algorithms used has the ability to outperform and produce accurate results. At the same time,

Table 3 shows the analysis of the frequency of use and usage percentage of classification algorithms for various diseases, and Table 4 represents the analysis of the frequency of use and usage percentage for COVID-19 and Mers-Cov disease only. Table 5 shows the overall analysis of classification algorithms that became the choices of many researchers.

There are many types of classification algorithms that have been used in order to predict prediction accuracy. In this research, 51 types of classification algorithms were identified. Each of the classification algorithms has a different performance ability. The details summary of the performance ability of each classification algorithm is shown in the Performance Metrics Column in Table 2. Tables 3 and 4 demonstrated the analysis of the frequency of use and usage percentage for all algorithms used in the previous study. While Table 5 concluded the most applied algorithms for various diseases and COVID-19, Mers-Cov. All the listed classification algorithms in Tables 3 and 4 are beneficial to be used in identifying, classifying, predicting and applying for future research since each algorithm has the ability to outperform and produce the best results. However, the algorithms listed in Table 5 are more appropriate to be applied in the next research. It is because of its popularity and the choice of many researchers (refer to Tables 2 and 3 for details).

This research shows that the Naïve Bayes, SVM, random forest, MLP and J48 algorithms were utilized more than other methods. These five algorithms have become researchers' choices in predicting diseases (refer to Table 3 for more details). The Naïve Bayes has been widely applied in many healthcare system issues [22], for example, in predicting COVID-19, heart diseases, diabetes, cancers and other diseases. Other than that, [12] [23] stated that Naïve Bayes is simple, easy to use, shows the best accuracy result, and the response is fast when used in large databases. Therefore, it gave satisfactory results because it can outperform more sophisticated classification methods. According to [12], prediction accuracy in the SVM is high as it also enrols a larger set of patterns and can dynamically update training patterns whenever there is a new pattern during classification [12]. Meanwhile, [24] expressed that random forest can produce better accuracy results. A random forest is also known as an ensemble classifier that contains many decision trees and returns class as output [25]. In deep learning, MLP is identified as one of the most commonly used neural network models [26]. Therefore, J48 is a decision tree-based algorithm. It is also becoming the choice of many researchers as it was originally developed from the C4.5 algorithm [27].

In recent studies, the Waikato Environment for Knowledge Analysis (WEKA) has been the choice of many as a data mining tool refer to Column Dataset and Tools in Table 2. Apart from that, WEKA is one of the most powerful data mining tools as it offers a large number of classification algorithms. WEKA is open-source software and machine learning tool introduced by the University of Waikato, New Zealand [28]. It is a portable and dependent platform, fully implemented in the Java programming language and runs on almost any modern computing platform [29]. Furthermore, the best feature is that it offers many different algorithms, is easy to use for people who are not data mining experts, and offers flexible facilities for script experiments [30].

Table 2. Comparative study on the Machine Learning Algorithms applied in classified and diagnosing various diseases.

References	Diseases	Machine Learning algorithm	Performance Metrics	**Datasets and Tool (instances, attributes)**
[31]	Heart Disease	J48 KNN Naïve bayes Sequential Minimal Optimization (SMO)	J48 Accuracy - 83.732% KNN Accuracy – 82.775% Naïve Bayes Accuracy – 81.818% SMO Accuracy – 82.775% Note: After comparison, results show that the best classification accuracy achieved by J48 algorithm.	9 instances 8 attributes Tool - WEKA
[32]	Heart Disease	Neural network SVM KNN Naïve Bayes Decision Tree	% accuracy is not applicable but writer explains decision tree has better accuracy as compared to other classifiers	303 instances 76 attributes Tool – Not stated
[33]	Coronary Artery Disease (CAD)	ANN SVM	ANN Positive Predictive Value (PPV) – 0.798 Sensitivity – 88.01 Specificity – 73.64 SVM Positive Predictive Value (PPV) – 0.871 Sensitivity – 92.32 Specificity – 74.42 Note: SVM algorithm predicts the CAD with higher PPV, Sensitivity and Specificity (Previous study showed that the use of the SVM algorithm predicts the diseases with the higher accuracy)	1324 instances 25 attributes Tool – Not stated

[29]	Breast Cancer	Naïve Bayes RBF Network J48	Naïve Bayes Accuracy – 97.36% Sensitivity – 97.4% Specificity – 93.41 % RBF Network Accuracy – 96.77% Sensitivity – 97.07% Specificity – 96.23% J48 Accuracy – 93.41% Sensitivity – 93.4% Specificity – 90.37% Note: Naïve Bayes algorithm is identified as the best prediction model for this research	683 instances 11 attributes Tool – Not stated
[34]	Heart Disease	KNN Naïve Bayes J48 JRip SVM AdaBoost Stochastic Gradient Decent (SGD) Decision Table	KNN Accuracy – 99.7073% Kappa – 0.9941 Naïve Bayes Accuracy – 83.122% Kappa – 0.6611 J48 Accuracy – 98.0488% Kappa – 0.961 Jrip Accuracy – 97.2683% Kappa – 0.9454 SVM Accuracy – 84.1951% Kappa – 0.6825 AdaBoost Accuracy – 84.2927% Kappa – 0.6857 SGD Accuracy – 84.3902% Kappa – 0.6825 Decision Table Accuracy – 93.6585% Kappa – 0.8734	1025 instances 14 attributes Tool - WEKA

			Note: The results show the best results in classification accuracy are KNN, J48 and JRip.	
[35]	Heart Disease	Naïve Bayes Decision Tree KNN	Tool: IHDPS (Intelligent Heart Disease Prediction System) Naïve Bayes Accuracy – 86.53% Decision Tree Accuracy – 89% KNN Accuracy – 85.53% Tool: Weka Naïve Bayes Accuracy – 96.5% Decision Tree Accuracy – 99.2% KNN Accuracy – 88.3% Note: The outcome reveals that Decision Tree has the highest accuracy value for both tools IHDPS and Weka	909 instances 13 attributes Tool – IHDPS, WEKA
[36]	Heart Disease	Random Forest Decision Tree Naïve Bayes	Random Forrest Precision – 81% Recall – 80.9% F-Measure – 80.9% Roc Area – 86.4% PRC Area – 84.8% TP Rate – 80.9% FP Rate – 19.2% Decision Tree Precision – 77% Recall – 77% F-Measure -77% Roc Area – 81.9% PRC Area – 77.4% TP Rate – 77% FP Rate – 24.1% Naïve Bayes	270 instances 13 attributes Tool – Not stated

			Precision – 31.3% Recall – 56% F-Measure – 40.2% Roc Area – 48.2% PRC Area – 49.8% TP Rate – 56% FP Rate – 56% Note: The results show that the Random Forest is identified as the best model for prediction of heart disease compare to Decision Tree and Naïve Bayes	
[37]	Heart Disease	Naïve Bayes SVM Random Forest KNN	Tools: Weka Naïve Bayes Precession – 83.7% Recall – 83.7% SVM Precision – 84% Recall – 83.65% Random Forest Precision – 81.8% Recall – 81.9% KNN Precision – 75.3% Recall – 75.2% Tools: Orange Naïve Bayes Precision – 82.4% Recall – 80.6% SVM Precision – 81.7% Recall – 70.5% Random Forest Precision – 77.9% Recall – 73.4% KNN Precision – 58% Recall – 54.7% Note: Compare to Orange tool and WEKA, WEKA has the	303 instances 76 attributes Tool – WEKA, Orange

			best results for Precession and recall	
[38]	Hepatitis Disease	Naïve Bayes Decision Table J48	Naïve Bayes Accuracy – 84.5% Decision Table Accuracy – 76.12% J48 Accuracy– 83.9% Note: the results show that, Naïve Bayes algorithm has the highest Accuracy.	Not stated (the exact value) Datasets are taken from UCI Machine Learning Repository Tool – Not stated
[39]	Thalassemia	Naïve Bayes MLP	Naïve Bayes Accuracy – 94.12% TP Rate (Recall/Sensitivity) – 93.8% FP Rate – 1.1% Correct Classification – 49 out of 51 Precision – 91.6% MLP Accuracy – 100% TP Rate (Recall/Sensitivity) – 0% FP Rate – 100% Correct Classification – 51 out of 51 Precision – 100% Note: The MLP algorithm are proven to be an effective algorithm to predict and diagnosis Thalassemia	51 instances 16 attributes Tool – WEKA
[40]	Diabetes	Naïve Bayes C4.5	Naïve Bayes (categorical data type– gender, Diabetes compilation disease and family history) C4.5 (ordinal data type – Age, BMI, Blood pressure, Duration of Diabetes sufferers and Blood glucose level) Note: The accuracy for Naïve Bayes and C4.5 is 68%	158 instances 15 attributes Tool – WEKA

[41]	Heart Disease	Decision Tree Naïve Bayes KNN	Decision Tree Sensitivity – 92.1% Specificity – 8.5% Accuracy – 92.2% Naïve Bayes Sensitivity – 84.2% Specificity – 16.5% Accuracy – 84.2% KNN Sensitivity – 100% Specificity – 0% Accuracy – 100% Note: KNN has the highest Accuracy. For prediction the Decision Tree performs well when compared to KNN and Naïve Bayes	303 instances 14 attributes Tool – WEKA
[42]	Liver Disease	C4.5 SVM FT Random Forest Logistic Regression	C4.5 Accuracy – 70.84% Sensitivity – 96.63% Specificity – 6.59% SVM Accuracy – 71.01% Sensitivity – 99.04% Specificity – 1.19% FT Accuracy – 69.13% Sensitivity – 89.18% Specificity – 19.16% Random Forest Accuracy – 73.07% Sensitivity – 90.14% Specificity – 30.54% Logistic Regression Accuracy – 72.04% Sensitivity – 91.11% Specificity – 24.55% Note: LR has the best accuracy compared to C4.5, SVM, FT and Random Forrest.	583 instances 10 attributes Tool – Not stated

[43]	COVID-19	No specific algorithm stated	The best MCC and F1 value for COVID-19 checkers are MCC - 85% F1 (Precision and Recall) – 92%	460 instances 10 COVID-19 checkers are used to screen CoVID-19 Symptoms Tool – Not stated
[44]	Diabetes	MLP BayesNet JRip C4.5 Fuzzy Lattice Reasoning (FLR)	MLP Accuracy - 75% Positive Recall – 37.25% Error Rate – 27.33% BayesNet Accuracy - 85% Positive Recall – 50% Error Rate – 36% JRip Accuracy - 86% Positive Recall – 11.9% Error Rate – 36% C4.5 Accuracy - 86% Positive Recall – 38% Error Rate – 28% FLR Accuracy - 75% Positive Recall – 37.25% Error Rate – 27.33% Note: C4.5 and JRip had the highest Accuracy which is above 85%. Thus this work concludes that, C4.5 and JRip are the most suitable algorithms for prediction diabetes patients.	1024 instances 26 attributes Tool – Not stated
[8]	Kidney Disease	Probabilistic Neural Networks (PNN) MLP SVM RBF	PNN Accuracy-96.7% MLP Accuracy-51.5% SVM Accuracy-60.7%	361 instances 25 variables Tool – DTREG Predictive Modelling System

			RBF Accuracy-87% Note: This study recommends that the Probabilistic Neural Networks (PNN) algorithm is the best algorithm that can be used to predict chronic kidney disease.	
[45]	Diabetes	MLP BayesNet Naïve Bayes J48graft FLR JRip Fuzzy Inference System(FIS) Adaptive Neuro-Fuzzy Inference System (ANFIS)	Tools: WEKA Multilayer Perceptron (MLP) Accuracy – 79.19% BayesNet/ Naïve Bayes Accuracy – 78.98% J48graft/C4.5 Accuracy – 81.33 % FLR Accuracy – 51.43% JRip Accuracy – 80.91% Tools: TANAGRA MLP Accuracy – 83.85% BayesNet/ Naïve Bayes Accuracy – 100% J48graft/C4.5 Accuracy – 90.63 % Tools: MATLAB FIS Accuracy – 71.51% ANFIS Accuracy – 78.79% Note: J48graft is the best algorithm in WEKA, Naïve Bayes is the best algorithm in TANAGRA and ANFIS is the best algorithm in MATLAB. Based on the average accuracy results, TANAGRA	768 instances 9 attributes Tool – WEKA, MATLAB, TANAGRA

			Machine Learning tools are the best compared to WEKA and MATLAB	
[46]	Diabetes	SVM Naïve Bayes Decision Trees ANN MLP Logistic Regression	Tool: WEKA SVM Accuracy -97.21 % Naïve Bayes Accuracy – 96.28% Decision Trees Accuracy – 93.5% ANN Accuracy -95.12 % MLP Accuracy – 98.83% Logistic Regression Accuracy -98.60 % Tool: Rapid Miner SVM Accuracy -97.98 % Naïve Bayes Accuracy – 95.30% Decision Trees Accuracy – 91.90% ANN Accuracy – 97.76% MLP Accuracy – 99.10% Logistic Regression Accuracy – 98.65% Note: Both the Weka and Rapid Miner tools show very high success rates for all algorithms. MLP algorithm has been the best algorithm with the highest success percentage in both of the tools. The Decision Trees	768 instances 9 attributes Tool – WEKA, Rapid Miner

			algorithm has been the algorithm whose success percentage is the lowest in both tools.	
[47]	Diabetes	Naïve Bayes Logistic MLP SMO KStar AdaBoostM1 Bagging ClassificationVia Clustering ClassificationVia Regression Multi Class Classifier VFI OneR ZeroR BFTree FT Random Tree Decision Table J48 IBk JRip	Naïve Bayes Accuracy – 78.2471% Logistic Accuracy – 79.1702% MLP Accuracy – 79.5245% SMO Accuracy – 78.4662% KStar Accuracy – 79.9487% AdaBoostM1 Accuracy – 78.5594% Bagging Accuracy – 80.8205% Classification Via Clustering Accuracy – 68.5595% Classification Via Regression Accuracy – 78.0559% Multi Class Classifier Accuracy – 79.1702% VFI Accuracy – 78.0559% OneR Accuracy – 78.4662% ZeroR Accuracy – 53.7763% BFTree Accuracy – 77.1375% FT Accuracy – 79.0816% Random Tree Accuracy – 80.5641%	400 instances 13 attributes Tool - WEKA

			DecisionTable Accuracy – 77.5478% J48 Accuracy – 79.8135% IBk Accuracy – 80.0606% JRip Accuracy – 77.7529% Note: The most top ranking classification is bagging, followed by Multiclass classifier and third is Random Tree	
[48]	Diabetes	Random Forest SVM	Random Forest Accuracy – 75.7813% Error rate – 24.2188% SVM Accuracy – 65.1042% Error rate- 34.8958% Note: Random forest has the maximum accuracy, minimum error rate, and takes less time to build the model than other classifiers.	768 instances 9 attributes Tool- WEKA
[23]	Kidney Disease	SVM Naïve Bayes	SVM Accuracy – 76.32% Execution Times – 3.22 Naïve Bayes Accuracy – 70.96% Execution Times – 1.29 Note: Naïve Bayes has the best execution time but SVM has the maximum value of accuracy.	584 instances 6 attributes Tool - MATLAB
[49]	Diabetes	Deep Neural Network (DNN)	DNN Five-fold cross validation Accuracy – 98.35%	7768 instances 8 attributes Tool – WEKA

			Ten-fold cross validation Accuracy – 97.11% Note: Accuracy on the Five-fold cross validation shows the best result compared to ten-fold cross validation	
[50]	Diabetes (Early stages)	Naïve Bayes Kstar ZeroR OneR J48 Random Forest	Naïve Bayes Accuracy – 76.30% Kstar Accuracy 69.14- % ZeroR Accuracy - 65.10% OneR Accuracy – 71.48% J48 Accuracy – 73.83% Random Forest Accuracy – 75.78% Note: Naïve Bayes has better result than other classifiers in terms of accuracy value and time taken to build the model	768 instances 9 attributes Tool - WEKA
[9]	Chronic Disease	Naïve Bayes Decision Tree Logistic Regression KNN and CNN	Naïve Bayes Accuracy – 52% Decision Tree Accuracy – 62% Logistic Regression Accuracy – 86% KNN and CNN Accuracy – 96% Note: The Proposed system (CNN and KNN) has the best accuracy result.	630 instances 52 attributes Tool – WEKA and MATLAB
[51]	Diabetes	Random Forest SVM Naïve Bayes Decision Tree	Random Forest Accuracy – 98.9% SVM Accuracy – 78.3%	2768 instances 8 attributes Tool – Not stated

			Naïve Bayes Accuracy – 77.6% Decision Tree Accuracy – 97.1% Note: Random Forest outperformed with the highest accuracy value compared to SVM, Naïve Bayes and Decision Tree.	
[25]	Diabetes	Bayesian Naïve Bayes J48 Random Forest Random Tree REP Tree FT Tree Cart SMO	Bayesian Accuracy – 78.25% Naïve Bayes Accuracy – 76.30% J48 Accuracy – 84.11% Random Forest Accuracy - 100% Random Tree Accuracy - 100% REP Tree Accuracy -83.07 % FT Tree Accuracy – 78.38% Cart Accuracy – 77.21% SMO Accuracy – 77.47% Note: Both Random Tree and Random Forest has the best accuracy result	768 instances 9 attributes Tool - WEKA
[52]	Abdomen Disease (liver disease kidney disease)	SVM Ripper Random Forest Hybrid Weighted Random Forest	Liver Diseases SVM Accuracy – 74.8% Ripper	**Liver Diseases** 583 instances **Kidney Diseases**

		Support Vector Machine (WRFSVM)	Accuracy – 84.73% Random Forest Accuracy – 88.93% WRFSVM Accuracy – 91.22% Kidney Diseases SVM Accuracy – 75.19% Ripper Accuracy – 83.21% Random Forest Accuracy – 88.85% WRFSVM Accuracy – 93.49% Note: It shows that WRFSSVM has the best accuracy results for both liver disease and kidney disease.	584 instances 5 attributes Tools - MATLAB
[53]	Covid-19	Logistic Regression SVM Decision Tree Naïve Bayes Random Forest KNN	Logistic Regression Accuracy – 97.49% SVM Accuracy – 98.85% Decision Tree Accuracy – 99.85% Naïve Bayes Accuracy – 97.52% Random Forest Accuracy – 99.60% KNN Accuracy – 98.06% Note: Decision Tree has the highest accuracy among other algorithms because Decision Tree is more efficient in predicting the recovery	3254 instances 8 attributes Tool - Python Programming

			possibility of the Covid-19 infected patients.	
[54]	MERS-CoV	J48 Naïve Bayes	Class: Recovery and Death J48 Accuracy - 68% Naïve Bayes Accuracy – 71.58% Class: Stable and Critical J48 Accuracy – 55.69% Naïve Bayes Accuracy – 53.63% Note: The overall results show that J48 has better accuracy.	1082 records Tool - WEKA
[55]	COVID-19	1. Random Forest	Random Forest Accuracy – 99.72% Note: These results have been compared with other algorithms as well such as KNN, Logistic Regression, SVM and Decision Tree.	1080 records Tool – Not stated
[56]	COVID-19	1. KNN 2. Logistic Regression 3. Decision Tree 4. SVM 5. MLP	KNN Accuracy – 80.37% Logistic Regression Accuracy – 78.54% Decision Tree Accuracy – 75.34% SVM Accuracy – 79.00% MLP Accuracy – 77.17% Note: The KNN has the highest accuracy compared to Logistic Regression, SVM, MLP and Decision Tree	730 records Tool – Not stated
[57]	COVID-19	1. Feature Correlated	Feature Correlated Naïve Bayes Accuracy – 99.00%	140 records of Covid-19 patients and

		Naïve Bayes (FCNB)	Note:	non-Covid-19 people Tool – Genetic Algorithm
[27]	COVID-19	1. J48 2. Hoeffding Tree	J48 (2 Fold) Accuracy – 83.60% Precision – 85.90% Recall – 83.60% Hoeffding Tree (2Fold) Accuracy – 82.65% Precision – 84.60% Recall – 82.70% Note: J48 shows the best result in terms of accuracy compared to hoeffding Tree	31,740 records 13 attributes Tool – WEKA
[58]	COVID-19	1. SVM and Deep Features	SVM and Deep Features Accuracy – 95.38% Note: The Resnet50 plus SVM has the highest accuracy compared to others classification	25 numbers of Covid-19 patients 25 numbers of x-tray images Tool - MATLAB
[59]	COVID-19	1. SVM 2. KNN 3. Naive-Bayes 4. Random forest 5. Gradient Boosting Machine (GBM) 6. Logistic 7. AdaBoost 8. XGBoost 9. Ensemble 10. Multilayer Perceptron – Nueral Network (MLP-NN)	SVM Accuracy –73.74% KNN Accuracy – 81.01% Naive-Bayes Accuracy –72.80% Random forest Accuracy – 87.02% Gradient Boosting Machine (GBM) Accuracy – 82.16% Logistic Accuracy – 73.44%	B-cell Tool – Not stated

			AdaBoost Accuracy – 83.33% XGBoost Accuracy – 81.29% Ensemble Accuracy – 87.80% MLP-NN Accuracy – 79.14% Note: The most accurate result was obtained using the ensemble classification which scored 87.80% validation accuracy.	
[60]	COVID-19	1. ANN 2. SVM 3. Logistic Regression 4. Stacking	ANN Accuracy – 96.2% SVM Accuracy – 90.7% Logistic Regression Accuracy – 96.7% Stacking Accuracy – 96.9% Note: Stacking model has the best accuracy rate compared to other models.	3486 instances 3 classes Tool – Not stated
[61]	COVID-19	1. Random Forest 2. Decision Tree 3. Bagging	Random Forest Accuracy – 98.3% Decision Tree Accuracy – 97.6% Bagging Accuracy – 97.8% Note: Random Forest achieves the best performance amongst the three classification algorithms employed	29315 instances 15 attributes Tools Python (Notebook)
[62]	COVID-19	1. Naïve Bayes 2. KNN 3. SVM	Naïve Bayes Accuracy – 98.99%	95839 instances 19 attributes

		4. J48 5. BayesNet 6. Random Forest	KNN Accuracy – 99.79% SVM Accuracy – 100% J48 Accuracy – 99.99% BayesNet Accuracy – 99.77% RandomForest Accuracy – 99.98% Note: SVM provides the best classification accuracy results with 100%	Tools WEKA
[63]	COVID-19	1. Logistic Regression 2. CNN	Logistic Regression Accuracy – 78.82% CNN Accuracy – 97.41% Note:	133MB COVID-19 3 sttributes Tools – Not Stated
[64]	COVID-19	1. SVM 2. FS-SVM 3. HPO-FS-SVM	SVM Accuracy – 80.42% FS-SVM Accuracy – 85.67% HPO-FS-SVM Accuracy – 90.73 Note: HPO-FS-SVM achieved better performance than FS-SVM and SVM	1000 images Tools – Not Stated
[65]	COVID-19	1. Naïve Bayes 2. BayesNet 3. Decision Tree 4. Random Forest 5. Logistic Regression	Naïve Bayes Accuracy – 68.53% BayesNet Accuracy – 60.39% Decision Tree Accuracy – 79.36 % Random Forest Accuracy – 87.28 % Logistic Regression	22852 instances 10 attributes

			Accuracy – 69.52% Note: Random forest has the highest accuracy value compares to others	

Table 3. Analysis on frequency of use and usage percentage Classification Algorithms for various diseases

Num.	Machine Learning Algorithm	Frequency of Use	Usage %
1	Naïve Bayes	26	15.03%
2	SVM	17	9.83%
3	Random Forest	14	8.09%
6	Decision Tree	11	6.36%
4	J48	11	6.36%
5	MLP	9	5.20%
7	KNN	9	5.20%
8	Logistic Regression	9	5.20%
9	JRip	4	2.31%
10	Decision Table	3	1.73%
11	C4.5	3	1.73%
12	AdaBoost	3	1.73%
13	SMO	3	1.73%
14	FT	3	1.73%
15	ANN	3	1.73%
16	RBF Network	2	1.16%
17	FLR	2	1.16%
18	Kstar	2	1.16%
19	OneR	2	1.16%
20	ZeroR	2	1.16%
21	Random Tree	2	1.16%
27	Bagging	2	1.16%
52	BayesNet	2	1.16%
22	Neural Network	1	0.58%
23	Stochastic Gradient Decent	1	0.58%
24	PNN	1	0.58%
25	FIS	1	0.58%

26	ANFIS	1	0.58%
28	ClassificationViaClustering	1	0.58%
29	ClassificationViaRegression	1	0.58%
30	MulticlassClassfier	1	0.58%
31	VFI	1	0.58%
32	BFTree	1	0.58%
33	FCNB	1	0.58%
34	IBk	1	0.58%
35	Hoeffding	1	0.58%
36	GBM	1	0.58%
37	Logistic	1	0.58%
38	XGBoost	1	0.58%
39	Ensemble	1	0.58%
40	DNN	1	0.58%
41	KNN and CNN	1	0.58%
42	Bayesian	1	0.58%
43	REP Tree	1	0.58%
44	Cart	1	0.58%
45	Ripper	1	0.58%
46	WRFSSVM	1	0.58%
47	Stacking	1	0.58%
48	MLP-NN	1	0.58%
49	CNN	1	0.58%
50	FS-SVM	1	0.58%
51	HPO-FS-SVM	1	0.58%

Table 4. Analysis on frequency of use and usage percentage Classification Algorithms for COVID-19 and Mers-Cov disease

Num.	Machine Learning Algorithm	Frequency of Use	Usage %
1	Naïve Bayes	6	12.00%
2	Support Vector Machine (SVM)	6	12.00%
3	Random Forest	6	12.00%
4	Logistic Regression	5	10.00%
8	Decision Tree	4	8.00%
6	J48	3	6.00%

7	K-Nearest Neighbor (KNN)	3	6.00%
5	Multilayer Perceptron (MLP)	2	4.00%
19	BayesNet	2	4.00%
9	AdaBoost	1	2.00%
10	Hoeffding	1	2.00%
11	GBM	1	2.00%
12	Logistic	1	2.00%
13	XGBoost	1	2.00%
14	Ensemble	1	2.00%
15	Stacking	1	2.00%
16	MLP-NN	1	2.00%
17	FCNB	1	2.00%
18	Bagging	1	2.00%
20	CNN	1	2.00%
21	FS-SVM	1	2.00%
22	HPO-FS-SVM	1	2.00%

Table 5. Summary of Classification Algorithms for various diseases and COVID-19, Mers-Cov

Num.	Machine Learning Algorithm	Usage Percentage (%)	
		Various Diseases	Covid-19 and Mers-Cov
1.	Naïve Bayes	15.03	12.00
2.	Support Vector Machine (SVM)	9.83	12.00
3.	Random Forest	8.09	12.00
4.	J48	6.36	6.00
5.	Multilayer Perceptron (MLP)	5.20	4.00
6.	Decision Tree	6.36	8.00
7.	K-Nearest Neighbour (KNN)	5.20	6.00
8.	Logistic Regression	5.20	10.00

*Author for correspondence

5. Conclusion

This research has reviewed and analyzed the classification algorithms in various diseases such as COVID-19, Mers-Cov, heart diseases, diabetes, coronary artery diseases, breast cancer, hepatitis, thalassemia, liver and kidney diseases. The results show that the classification algorithms such as Naïve Bayes, support vector machine (SVM), multilayer perceptron

(MLP), J48, K-nearest neighbor (k-NN) decision tree, random forest and logistic regression were listed as the most applied algorithms in the previous research as demonstrated in Table 4. However, this reviewed research identified Naïve Bayes, SVM, random forest, MLP and J48 algorithms as the top five most utilized algorithms.

The Naïve Bayes classification was the number one choice of many researchers because it is a simple, easy, and powerful model. It returns not only the prediction but also the level of certainty, which can be very useful [29]. According to [39], the SVM functions effectively in high-dimensional spaces and is relatively memory efficient. Apart from that, MLP is one of the crucial classes in the neural network, containing three layers, an input layer, one or more hidden layers, and an output layer [8]. This model aims to minimize the difference between the intended results of the network and the achieved result [45]. Alternatively, the random forest was identified as one of the efficient algorithms as it gives a good accuracy value [60]. Meanwhile, J48 has also been the choice of many researchers in data mining as it is one of the fastest and easiest models because it does not require any domain information [61-69]. The coronavirus (COVID-19) is extremely unpredictable with the emergence of COVID-19 variants is becoming a barrier because each variant shows different symptoms.

In conclusion, classification algorithms will help researchers work easily and improve prediction accuracy. Furthermore, there are many suggestions and solutions provided in this paper for future work. Therefore, the researcher will test the results in Table 5 as guidance for the next research.

Acknowledgment
This research was supported by the Research Management and Innovation Centre, Universiti Sultan Zainal Abidin (UniSZA).

Conflicts Of Interest

The authors have no conflicts of interest to declare.

References

1. R. Rosly, M. Makhtar, M. K. Awang, M. I. Awang, and M. N. A. Rahman, “Analyzing performance of classifiers for medical datasets,” Int. J. Eng. Technol., vol. 7, no. 2, pp. 136–138, 2018, doi: 10.14419/ijet.v7i2.15.11370.
2. World Health Organization, “The top 10 causes of death,” World Health Organization, 2020. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.
3. S. Basij-Rasikh, M. Khalil, and N. Safi, “Early responses to covid-19 in afghanistan,” Eastern Mediterranean Health Journal, vol. 26, no. 12. World Health Organization, pp. 1442–1445, 2020, doi: 10.26719/emhj.20.137.
4. World Health Organization, “COVID-19 Weekly Epidemiological Edition 53,” World Health Organization, 2021. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20210817_weekly_epi_update_53.pdf.
5. A. Elengoe, “COVID-19 outbreak in Malaysia,” Osong Public Health and Research Perspectives, vol. 11, no. 3. Korean Disease Control and Prevention Agency, pp. 93–100, 2020, doi: 10.24171/j.phrp.2020.11.3.08.
6. M. Z. Alom et al., “A state-of-the-art survey on deep learning theory and architectures,” Electron., vol. 8, no. 3, 2019, doi: 10.3390/electronics8030292.

7. L. Wynants et al., "Prediction models for diagnosis and prognosis of covid-19 : systematic review and critical appraisal," BMC Public Health, 2020, doi: 10.1136/bmj.m1328.
8. E.-H. A.Rady and A. A. S., "Prediction of kidney disease stages using data mining algorithms," Informatics Med. Unlocked, vol. 15, no. December 2018, p. 100178, 2019, doi: 10.1016/j.imu.2019.100178.
9. R. Alanazi, "Identification and Prediction of Chronic Diseases Using Machine Learning Approach," J. Healthc. Eng., vol. 2022, 2022, doi: 10.1155/2022/2826127.
10. B. Khalid and N. Abdelwahab, "A Model for Predicting Ischemic Stroke Using Data Mining Algorithms," IJISET - Int. J. Innov. Sci. Eng. Technol. Vol. 2 Issue 11, Novemb. 2015, vol. 2, no. 11, pp. 18–23, 2015, [Online]. Available: http://ijiset.com/vol2/v2s11/IJISET_V2_I11_04.pdf.
11. Ahmed M. Salah EL-Bohy, Atallah I. Hashad, and Hussien Saad Taha, "Performance Evaluation of Hepatitis Diagnosis using Single and Multi-Classifiers Fusion," Int. J. Eng. Res., vol. V4, no. 04, pp. 293–298, 2015, doi: 10.17577/ijertv4is040503.
12. S. R. Joseph, H. Hlomani, and K. Letsholo, "Data Mining Algorithms: An Overview," Int. J. Comput. Technol., vol. 15, no. 6, pp. 6806–6813, 2016, doi: 10.24297/ijct.v15i6.1615.
13. Pan American Health Organization, "Why Predictive Modeling is Critical in the Fight against COVID-19," Pan American Health Organization, pp. 1–4, 2020.
14. ROHIT GARG, "7 Types of Classification Algorithms," Developers Corner, p. https://analyticsindiamag.com/7-types-classificati, Jan. 2018.
15. R. Rosly, M. Makhtar, M. K. Awang, M. I. Awang, M. N. A. Rahman, and H. Mahdin, "Comprehensive study on ensemble classification for medical applications," Int. J. Eng. Technol., vol. 7, no. 2.14 Special Issue 14, pp. 186–190, 2018, doi: 10.14419/ijet.v7i2.14.12822.
16. Gauraz Sharma, "5 Classification Algorithms you should know – introductory guide!," Data Science Blogathon, 2021. https://www.analyticsvidhya.com/blog/2021/05/5-classification-algorithms-you-should-know-introductory-guide/#:~:text=Introduction,%2C multiclass classification%2C multilabel classification.
17. S. Bashir, U. Qamar, and F. H. Khan, "IntelliHealth: A medical decision support application using a novel weighted multi-layer classifier ensemble framework," J. Biomed. Inform., vol. 59, pp. 185–200, 2016, doi: 10.1016/j.jbi.2015.12.001.
18. J. Brownlee, "What is Deep Learning?," Machine Learning Mastery, 2019. https://machinelearningmastery.com/what-is-deep-learning/.
19. I. Goodfellow, Yoshua Bengio, and Aaron Courville, "Deep Learning," Int. J. Semant. Comput., vol. 10, no. 3, pp. 417–439, 2016, doi: 10.1142/S1793351X16500045.
20. M. Makhtar, R. Rosly, M. K. Awang, M. Mohamad, and A. H. Zakaria, "A multi-classifier method based deep learning approach for breast cancer," Int. J. Eng. Trends Technol., no. 1, pp. 102–107, 2020, doi: 10.14445/22315381/CATI3P217.
21. L. P. Jin and J. Dong, "Ensemble Deep Learning for Biomedical Time Series Classification," Comput. Intell. Neurosci., vol. 2016, 2016, doi: 10.1155/2016/6212684.
22. B. Bahrami and M. H. Shirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques," 2015. [Online]. Available: www.jmest.org.
23. V. S and D. S, "Data Mining Classification Algorithms for Kidney Disease Prediction," Int. J. Cybern. Informatics, vol. 4, no. 4, pp. 13–25, 2015, doi: 10.5121/ijci.2015.4402.
24. R. Ghorbani and R. Ghousi, "Predictive data mining approaches in medical diagnosis: A review of some diseases prediction," International Journal of Data and Network Science, vol. 3, no. 2. Growing Science, pp. 47–70, 2019, doi: 10.5267/j.ijdns.2019.1.003.
25. L. AlThunayan, N. AlSahdi, and L. Syed, "Comparative analysis of different classification algorithms for prediction of diabetes disease," ACM Int. Conf. Proceeding Ser., pp. 1–6, 2017, doi: 10.1145/3018896.3036387.

26. Edpresso Team, “What is a multi-layered perceptron?,” Edpresso Team, 2022. .
27. N. Rochmawati et al., “Covid Symptom Severity Using Decision Tree,” Proceeding - 2020 3rd Int. Conf. Vocat. Educ. Electr. Eng. Strength. Framew. Soc. 5.0 through Innov. Educ. Electr. Eng. Informatics Eng. ICVEE 2020, 2020, doi: 10.1109/ICVEE50212.2020.9243246.
28. R. Kirkby, E. Frank, and P. Reutemann, “WEKA Explorer User Guide,” vol. 1, p. 22, 2008.
29. V. Chaurasia, S. Pal, and B. B. Tiwari, “Prediction of benign and malignant breast cancer using data mining techniques,” J. Algorithms Comput. Technol., vol. 12, no. 2, pp. 119–126, Jun. 2018, doi: 10.1177/1748301818756225.
30. N. Review, B. Saleh, A. Saeidi, A. Al-Aqbi, L. Salman, and B. S. Al, “MEDICAL REVIEWS Analysis of Weka Data Mining Techniques for Heart Disease Prediction System,” Int J Med Rev, vol. 7, no. 1, pp. 15–24, 2020, doi: 10.30491/ijmr.2020.221474.1078.
31. S. Nikhar and A. M. Karandikar, “Prediction of Heart Disease Using Machine Learning Algorithms,” Int. J. Adv. Eng. Manag. Sci., vol. 2, no. 6, 2016, [Online]. Available: www.ijaems.com.
32. H. Ayatollahi, L. Gholamhosseini, and M. Salehi, “Predicting coronary artery disease: A comparison between two data mining algorithms,” BMC Public Health, vol. 19, no. 1, Apr. 2019, doi: 10.1186/s12889-019-6721-5.
33. K. M. Almustafa, “Prediction of heart disease and classifiers’ sensitivity analysis,” BMC Bioinformatics, vol. 21, no. 1, Jul. 2020, doi: 10.1186/s12859-020-03626-y.
34. J. Soni, U. Ansari, D. Sharma, S. Soni, and S. Associate, “Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction,” 2011.
35. H. Benjamin, F. David, and S. A. Belcy, “HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES,” ICTACT J. SOFT Comput., p. 1, 2018, doi: 10.21917/ijsc.2018.0253.
36. B. Sarangam Kodati, R. Vivekanandam, S. Kodati α, and R. Vivekanandam σ, “Analysis of Heart Disease using in Data Mining Tools Orange and Weka Sri Satya Sai University Analysis of Heart Disease using in Data Mining Tools Orange and Weka,” 2018.
37. P. Mahani and N. Ruhil, “APPLICATION OF DATA MINING IN HEALTH CARE,” 2016. [Online]. Available: http://www.ijates.com/images/short_pdf/1455803777_593Y.pdf.
38. N. Chidozie Egejuru, S. Olayinka Olusanya, A. Onyenonachi Asinobi, O. Joseph Adeyemi, V. Oluwatimilehin Adebayo, and P. Adebayo Idowu, “Using Data Mining Algorithms for Thalassemia Risk Prediction,” Int. J. Biomed. Sci. Eng., vol. 7, no. 2, p. 33, 2019, doi: 10.11648/j.ijbse.20190702.12.
39. C. Fiarni, E. M. Sipayung, and S. Maemunah, “Analysis and prediction of diabetes complication disease using data mining algorithm,” Procedia Comput. Sci., vol. 161, pp. 449–457, 2019, doi: 10.1016/j.procs.2019.11.144.
40. S. Joshi and Mydhili K. Nair, “Prediction of Heart Disease Using Classification Based Data Mining Techniques,” Smart Innov. Syst. Technol., vol. 32, no. December 2015, 2015, doi: 10.1007/978-81-322-2208-8.
41. S. Omokanye and Taye Aro, “Homogenous Ensembles Of Data Mining Algorithms In Predicting Liver Disease,” vol. XVI, 2018.
42. N. Munsch et al., “A benchmark of online COVID-19 symptom checkers,” medRxiv, p. 2020.05.22.20109777, 2020, [Online]. Available: http://medrxiv.org/content/early/2020/05/26/2020.05.22.20109777.abstract.
43. R. Manimaran and Dr. M.Vanitha, “Prediction of Diabetes Disease Using Classification Data Mining Techniques,” Int. J. Eng. Technol., vol. 9, no. 5, pp. 3610–3614, 2017, doi: 10.21817/ijet/2017/v9i5/170905319.
44. R. M. Rahman and F. Afroz, “Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis,” J. Softw. Eng. Appl., vol. 06, no. 03, pp. 85–97, 2013, doi: 10.4236/jsea.2013.63013.

45. T. Pala and A. Y. Camurcu, “EVALUATION OF DATA MINING CLASSIFICATION AND CLUSTERING TECHNIQUES FOR DIABETES,” Malaysian J. Comput., vol. 2, 2014.
46. K. Ahmed and J. Tasnuba, “Comparative Analysis of Data Mining Classification Algorithms in Type-2 Diabetes Prediction Data Using WEKA Approach,” Int. J. Sci. Eng., vol. 7, no. 2, Oct. 2014, doi: 10.12777/ijse.7.2.150-154.
47. H. Rashid Abdulqadir, A. Mohsin Abdulazeez, and D. Assad Zebari, “Data Mining Classification Techniques for Diabetes Prediction,” Qubahan Acad. J., vol. 1, no. 2, pp. 125–133, 2021, doi: 10.48161/qaj.v1n2a55.
48. S. Islam Ayon and M. Milon Islam, “Diabetes Prediction: A Deep Learning Approach,” Int. J. Inf. Eng. Electron. Bus., vol. 11, no. 2, pp. 21–27, 2019, doi: 10.5815/ijieeb.2019.02.03.
49. S. Sivakumar, S. Venkataraman, and A. Bwatiramba, “Classification Algorithm in Predicting the Diabetes in Early Stages,” J. Comput. Sci., vol. 16, no. 10, pp. 1417–1422, 2020, doi: 10.3844/jcssp.2020.1417.1422.
50. M. O. Edeh et al., “A Classification Algorithm-Based Hybrid Diabetes Prediction Model,” Front. Public Heal., vol. 10, no. March, pp. 1–7, 2022, doi: 10.3389/fpubh.2022.829519.
51. S. Vijayarani, C. Sivamathi, and P. Tamilarasi, “ASEAN Journal of Science and Engineering A Hybrid Classification Algorithm for Abdomen Disease Prediction,” vol. 3, no. 3, 2022.
52. L. J. Muhammad, M. M. Islam, S. S. Usman, and S. I. Ayon, “Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients’ Recovery,” SN Comput. Sci., vol. 1, no. 4, pp. 1–7, 2020, doi: 10.1007/s42979-020-00216-w.
53. I. Al-Turaiki, M. Alshahrani, and T. Almutairi, “Building predictive models for MERS-CoV infections using data mining techniques,” J. Infect. Public Health, no. January, pp. 19–21, 2016.
54. S. Mallick, B. Rajak, A. K. Verma, and D. S. Kushwaha, “A Best Effort Classification Model For Sars-Cov-2 Carriers Using Random Forest,” Int. J. Comput. Sci. Netw. Secur., vol. 21, no. 1, pp. 27–33, 2021.
55. P. Theerthagiri, I. J. Jacob, A. U. Ruby, and Y. Vamsidhar, “Prediction of COVID-19 Possibilities using KNN Classification Algorithm,” Res. Sq., 2021.
56. N. A. Mansour, A. I. Saleh, M. Badawy, and H. A. Ali, Accurate detection of Covid-19 patients based on Feature Correlated Naïve Bayes (FCNB) classification strategy, no. 0123456789. Springer Berlin Heidelberg, 2021.
57. P. K. Sethy, S. K. Behera, P. K. Ratha, and P. Biswas, “Detection ofCoronavirus Disease (COVID-19) Based on Deep Features,” Int. J. Math. Eng. Manag. Sci., vol. 5, no. 4, pp. 643–651, 2020, doi: 10.20944/preprints202003.0300.v1.
58. N. Jain et al., “Prediction modelling of COVID using machine learning methods from B-cell,” Results Phys., no. January, 2020.
59. Y. S. Taspinar, I. Cinar, and M. Koklu, “Classification by a stacking model using CNN features for COVID-19 infection diagnosis,” J. Xray. Sci. Technol., vol. 30, no. 1, pp. 73–88, 2022, doi: 10.3233/XST-211031.
60. Y. Meraihi, A. B. Gabis, S. Mirjalili, A. Ramdane-Cherif, and F. E. Alsaadi, Machine Learning-Based Research for COVID-19 Detection, Diagnosis, and Prediction: A Survey, vol. 3, no. 4. Springer Nature Singapore, 2022.
61. B. Srinivasan and K. Pavya, “A STUDY ON DATA MINING PREDICTION TECHNIQUES IN HEALTHCARE SECTOR,” 2016. [Online]. Available: www.irjet.net.
62. Myagmarsuren Orosoo, J Chandra Sekhar, Manikandan Rengarajan and Nyamsuren Tsendsuren, Adapa Gopi, Yousef A.Baker El-Ebiary, Prema S, Ahmed I. Taloba “Analysing Code-Mixed Text in Programming Instruction Through Machine Learning for Feature Extraction” International Journal of Advanced Computer Science and Applications(IJACSA),15(7), 2024. http://dx.doi.org/10.14569/IJACSA.2024.0150788.
63. Anna Gustina Zainal, M. Misba, Punit Pathak, Indrajit Patra, Adapa Gopi, Yousef A.Baker El-

Ebiary and Prema S, "Cross-Cultural Language Proficiency Scaling using Transformer and Attention Mechanism Hybrid Model" International Journal of Advanced Computer Science and Applications(IJACSA), 15(6), 2024. http://dx.doi.org/10.14569/IJACSA.2024.01506116.

64. A. Greeni, Yousef A.Baker El-Ebiary, G. Venkata Krishna, G. Vikram, Kuchipudi Prasanth Kumar, Ravikiran K and B Kiran Bala, "BrainLang DL: A Deep Learning Approach to FMRI for Unveiling Neural Correlates of Language across Cultures" International Journal of Advanced Computer Science and Applications(IJACSA), 15(6), 2024. http://dx.doi.org/10.14569/IJACSA.2024.01506114.
65. Taviti Naidu Gongada, Girish Bhagwant Desale, Shamrao Parashram Ghodake, K. Sridharan, Vuda Sreenivasa Rao and Yousef A.Baker El-Ebiary, "Optimizing Resource Allocation in Cloud Environments using Fruit Fly Optimization and Convolutional Neural Networks" International Journal of Advanced Computer Science and Applications(IJACSA),15(5),2024.http://dx.doi.org/10.14569/IJACSA.2024.015051
66. Kambala Vijaya Kumar, Y Dileep Kumar, Sanjiv Rao Godla, Mohammed Saleh Al Ansari, Yousef A.Baker El-Ebiary and Elangovan Muniyandy, "Enhancing Water Quality Forecasting Reliability Through Optimal Parameterization of Neuro-Fuzzy Models via Tunicate Swarm Optimization" International Journal of Advanced Computer Science and Applications(IJACSA), 15(3), 2024. http://dx.doi.org/10.14569/IJACSA.2024.01503110.
67. Belal Alifan, Mokhairi Makhtar, Yousef A. Baker El-Ebiary; A review study of electronic health care systems in Jordan. AIP Conf. Proc. 22 March 2024; 2816 (1): 180002. https://doi.org/10.1063/5.0177566.
68. Franciskus Antonius Alijoyo, Taviti Naidu Gongada, Chamandeep Kaur, N. Mageswari, J.C. Sekhar, Janjhyam Venkata Naga Ramesh, Yousef A.Baker El-Ebiary, Zoirov Ulmas, Advanced hybrid CNN-Bi-LSTM model augmented with GA and FFO for enhanced cyclone intensity forecasting, Alexandria Engineering Journal, Volume 92, 2024, Pages 346-357, ISSN 1110-0168, https://doi.org/10.1016/j.aej.2024.02.062.
69. V Moses Jayakumar, R. Rajakumari, Kuppala Padmini, Sanjiv Rao Godla, Yousef A.Baker El-Ebiary and Vijayalakshmi Ponnuswamy, "Elevating Neuro-Linguistic Decoding: Deepening Neural-Device Interaction with RNN-GRU for Non-Invasive Language Decoding" International Journal of Advanced Computer Science and Applications(IJACSA),15(2), 2024. http://dx.doi.org/10.14569/IJACSA.2024.0150233.

Appendix 1

S. No.	Abbreviations	Description
1.	COVID-19	Coronavirus
2.	SARS-Cov-2	Severe Acute Respiratory Syndrome
3.	SVM	Support Vector Machine
4.	MLP	Multilayer Perceptron
5.	KNN	Knowledge Nearest Neighbor
6.	ANN	Artificial Neural Network
7.	FLR	Fuzzy Inference System
8.	PNN	Probabilistic Neural Network
9.	FIS	Fuzzy Inference system
10.	ANFIS	Adaptive Neuro-fuzzy Inference System
11.	CNN	Convolutional Nueral Network

12.	WRFSSVM	Hybrid Weighted Random Forest Support Vector Machine
13.	FCNB	Feature Correlated Naïve Bayes
14.	MLP-NN	Multilayer Perceptron – Neural Network
15.	DNN	Deep Neural Network
16.	VFI	Voting Feature Interval
17.	IHDPS	Intelligent Heart Disease Prediction System
18.	WEKA	Waikato Environment for Knowledge Analysis