Southeast Europe Journal of Soft Computing ISSN 2233 – 1859
Available online: http://scjournal.ius.edu.ba

Vol 11, No. 2, September 2022, pp. 06-17

Machine Learning Algorithms in Analysis, Diagnosing and Predicting COVID-19: A Systematic Literature Review

[1] Jasim Mohammed Dahr, [2*]Alaa Khalaf Hamoud, [2] Nahla Hamad Abdul-Samad , [2] Fatima Jasim Muhammed , [2] Aliaa Saad Aljubair

[1] Educational Planning Department, Directorate of Education in Basrah, Iraq,
[2] Department of Computer Information Systems, Colleg of Computer Science and Information Technology, University of Basrah, Iraq
**Corresponding Author*: alaa.hamoud@uobasrah.edu.iq

Article Info

Article history:

Article received on 10 October 2022

Received in revised form 01 November 2022

Keywords:

Supervised Machine Learning; Unsupervised Machine Learning; COVID-19; SARS-CoV-2; Predicting; Diagnosing; Analysis.

ABSTRACT: Since the COVID-19 corona virus first appeared at the end of 2019, in Wuhan province, China, the analysis, diagnosis, and prognosis of COVID-19 (SARS-CoV-2) has attracted the greatest attention. Since then, every part of the world needs some sort of system or instrument to assist judgments for prompt quarantine and medical treatment. For a variety of uses, including prediction, classification, and analysis, machine learning (MLR) have demonstrated their accuracy and efficiency in the fields of education, health, and security. In this paper, three main questions will be answered related with COVID-19 analysis, predicting, and diagnosing. The performance evaluation, fast process and identification, quick learning, and accurate results of MLR algorithms make them a base for all models in analyzing, diagnosing, and predicting COVID-19 infection. The impact of using supervised and unsupervised MLR can be used for estimating the spread level of COVID-19 to make the proper strategic decisions. The researchers next compared the effects of various datatypes on diagnosing, forecasting, and assessing the severity of COVID-19 infection in order to examine the effects of MLRs. Three fields are associated to COVID-19, according to the analysis of the chosen study (analysis, diagnosing, and predicting). The majority of researches focuses on the subject of COVID-19 diagnosis, where they use their models to identify the infection. In the selected studies, several algorithms were employed; however a study revealed that the neural network is the most used method when compared to other algorithms. The most used method for identifying, forecasting, and evaluating COVID-19 infection is supervised MLR.

1. INTRODUCTION

The globe was caught off guard by the COVID-19 epidemic. More than (622,775,232) instances have been documented, and more than (6,548,639) people have died as a result [1]. Despite how concerning these figures are, it is possible that there are more incidents than this. This virus has had an influence on many facets of life, including politics, education, the economy, social issues, the environment, and climate change. It has also raised concerns about how well-equipped governments, civil society, and healthcare systems are to handle an unknown illness. The unexpected coronavirus disease epidemic of 2019 (COVID-19) highlighted the necessity for quick and trustworthy

automation solutions to support medical teams. A precise and trustworthy diagnosis is essential for giving suspected or infected persons prompt medical attention. It also aids government organizations in stopping the disease's spread and saving lives [2]. But the lack of medical resources to make early diagnosis may prevent early prevention or treatment.

Supervised machine learning (MLR) algorithms classify the current instances and produce hypotheses and patterns according to the supplied instances. These algorithms can also categorize and predict future cases based on prior data. Supervised learning is frequently used in different sectors and has proved its accuracy and effieciency in both of predicting, and classifying. Using previous information, algorithms for supervised MLR classification are created to categorize data. Classification is commonly used in data science issues. [3], [4]. Different algorithms in the field of supervised learning such as (support vector machine (SVM), decision tree (DT), artificial neural network (ANN), naïve bayes, association rule mining, gradient boosting, and regression) are used in different fields such as education [5]–[11], financial management [12], [13], healthcare [14], [15], wireless sensor network (WSN) [16]–[18], sentiment analysis [19], [20], image processing [21], [22], intrusion detection [23], [24], and disaster management [25], [26]. Unsupervised MLR on the other side, discover the hidden patterns for unlabeled data [3][27]. Different unsupervised MLR algorithms are utilized such as (clustering, hierarchical clustering, ANN, and principle component analysis) are utilized in different sectors and proved their accuracy. The fields are education [28][29][30], financial management [31][32], healthcare, WSN [18][33], disaster management [26], [34], image processing [35], [36], and sentiment analysis [37], [38].

For this, a variety of supervised and unsupervised MLR techniques have been utilized to analyze, diagnose and predict COVID-19. The medical decision may be supported by the application of AI-based techniques to help with the screening of individuals who may be contaminated with COVID-19. In this study, a systematic study was conducted on the most prominent researches that used supervised and unsupervised MLR algorithms that were applied to predict COVID-19. The total number of papers in this study is (147) papers in which different sectors are handled, such as (analysis, predicting, and diagnosing) of COVID-19 based on supervised and supervised MLR algorithm. The other sections of the study were structured as follows: previous studies linked to the application of MLR algorithms to predict COVID-19 were included, methodology for extracting studies, as well as discussion, findings, and conclusion.

2. LITERATURE REVIEW

In [39], the purposed study developed and evaluated MLR-based algorithm models for the predicting and diagnosing of the COVID-19 severity using data from urine, hematological, and biochemical tests from big datasets. Four methods in all were used: the k-nearest neighbor approach (KNN), decision trees (DT), ANN, and partial least squares discriminant analysis (PLS-DA). All of the MLR-based models (PLS-DA, DT, ANN, and KNN) were proved their ability to accurately predict the diagnosis of COVID-19 and the seriousness of the illness with an accuracies over 84%, which is comparable to the outcomes of RT-PCR and the minimal threshold advised for diagnostic testing. The ANN model had the highest performance (94%, and 98%) and might be utilized by healthcare practitioners as a supplemental decision-making tool in the real world. Hypoxemia, hyperferritinaemia, hypocalcemia, metabolic acidosis, respiratory acidosis, pulmonary hypoxia, high lactate dehydrogenase levels and low urine pH were all associated with the diagnosis and severity of COVID-19. Further clinical studies should more thoroughly examine these indicators since they are potential treatment targets.

This study [40] introduces the COVID-19 Patients Detection Approach, a novel COVID-19 diagnostic strategy (CPDS). Two contributions account for much of CPDS' uniqueness. First up is a novel-new hybrid feature selection methodology (HFSM), which selects the most helpful characteristics from those produced from non-COVID-19 people and chest computed tomography (CT) images for COVID-19 patients. Due to the fact that it combines information from both filter (FS) and wrapper methods, HFSM is a hybrid methodology. Fast Selection Stage (FS2) and Accurate Selection Stage are the two steps that make up this process (AS2). While AS2 employs Genetic Algorithm as a wrapper approach, FS2 depends on filters. HFSM selects the important characteristics for the subsequent detection phase as a hybrid technique. The second benefit is an improved KNN classifier, which by adding sound heuristics while selecting the neighbors of the tested item, overcomes the trapping issue of standard KNN. EKNN selects just the qualifying neighbors for categorization after evaluating the degree of each neighbor's closeness and strength to the tested item. As a result, using the key traits chosen by the HFSM approach, EKNN can precisely and quickly detect

infected individuals. On chest CT scans, several tests have been conducted taking into account the suggested detection approach as well as current rival techniques. According to experimental findings, the suggested detection strategy works better than more current methods since it introduces the highest accuracy rate.

In relation to the development of the COVID-19 epidemic [41], In order to give inputs to assist decision-making for public policies, this work presents a research that examines linear regression processes in conjunction with a sliding and cumulative time window method, including whether they are easing isolation or hardening. To forecast the behavior of the curve for cases and fatalities of COVID-19, data from five Brazilian states were gathered and a Ridge regression was used. In light of instances and fatalities, respectively, a reported Explained Variance Status (EVS) between 0.998 and 0.999. Since public policy may change within a short period of time, It was shown that compared to cumulative time windows, sliding time windows offer better insight into the illness.

In [42], the current study's objective was to build a clinical decision support tool to aid in the diagnosis of COVID-19 utilizing MLR algorithms and findings from standard laboratory tests. We created MLR algorithms using lab data (n = 1,391) that included findings from 6 clinical chemistry (CC) tests, the gold standard test results, and 14 CBC parameter tests for the severe acute respiratory syndrome coronavirus 2. Eight MLR algorithms were created using a mix of CC and CBC parameters and CBC using four MLR algorithms: gradient boosting (XGBoost), random forest (RF), logistic regression, and SVM. On the test set of data and the external Brazilian set of validation data, performance evaluation was done. All models' accuracy percentages varied from 74% to 91%. On the data set used for this study, the RF model trained from CC and CBC analytes performed the best (accuracy, 79.6%; sensitivity; specificity; 91.2%). With 82.8% accuracy, the RF model, which was trained using only CBC parameters, identified COVID-19 instances. The SVM model trained using CBC and CC parameters had the greatest results on the external validation data set (91.18% accuracy, 100% sensitivity, and 84.21% specificity).

3. RESEARCH METHOD

In this study, the authors aim to conduct a systematic literature review (SLR) to focus on studies related to the use of MLR algorithms in predicting COVID-19, where accurate results are extracted based on the studies extracted according to a specific methodology. SLR focuses on a particular topic and is used as a means of summarizing research evidence rather than expert commentary and anecdotal reviews [43]. There is a significant differences among systematic reviews, meta-analyses, and narrative reviews. It follows a precise and rigorous method of extracting the literature to obtain accurate and confirmed information [2, 3]. Whereas, the narrative review is a more broad and less comprehensive method of analysis, as it arranges a group of previous studies by topic and summarizes some of the comprehensive results from the impressions of the trends. Meta-analysis is a procedure band of statistical method for integrating outcome-form quantitative studies for an interpretation of general trends. Often this type of analysis methods focuses on one variable relationship in the analysis. The protocol is one of the features that must be provided in any systematic review [46]. As a result of this study, the protocol used by researchers is dependent on a set of steps put in by Kitchenham and Charters. [47]. In figure (1), SLR stages are depicted.

Figure 1: Systematic Literature Review Steps

3.1 Formulating the systematic review questions

The framing of the research questions that the study seeks to address is the first stage in a systematic literature review (SLR). Therefore, the researchers in this study picked the following as the first question in order to discover the most crucial algorithms connected to unsupervised MLR and employed in prediction in particular:

What is the impact of using MLR algorithms in predicting COVID-19?

Instead of providing general summaries of the area in which you are interested, the primary goal of a systematic literature review is to address a specific research issue [48]. To undertake research on all prediction-related algorithms and to provide a summary of the literature, the first research question was developed. Undoubtedly, past research in the field of prediction employed a variety of methods, which inspired academics to come up with new research questions:

What is most supervised/unsupervised MLR algorithm used for predicting COVID-19?

Numerous sectors will be provided based on studying the COVID-19 DW field. This motivated us to look for and investigate the many COVID-19 DW regions that peaked researchers' curiosity. Consequently, the following question will be put forth:

What is the sector that faced a lot of attention in the field of predicting COVID-19?

3.2 Constructing the search

It is vital to identify a set of pertinent keywords that will be used in the study, as well as the digital libraries that are accessible to researchers, in order to conduct a search for prior studies and find the answers to the research questions.

3.2.1 Search process

The study concentrated on the following set of research-related digital libraries: IEEE electronic library, Science Direct, and PubMod. The researchers' available libraries were taken into consideration while selecting the digital libraries.

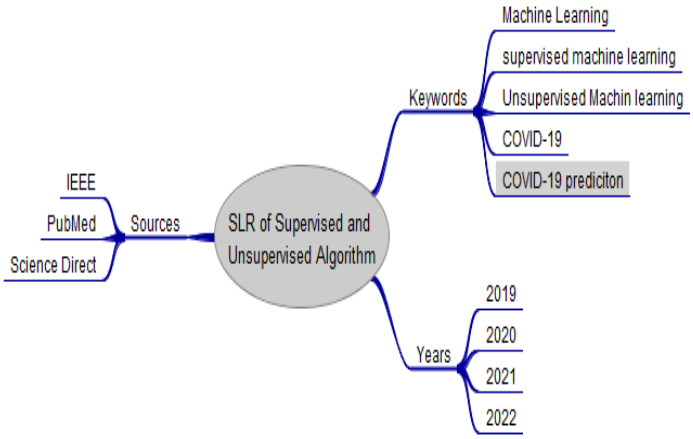

Figure 2: Search strategy of SLR

3.2.2 Terminology

It is a term or group of keywords that will be used to find papers that are relevant to a particular topic. In order to address their research questions, researchers created a set of research words to describe earlier experiments. The following keywords are among the standard search phrases: “Machine learning ”, “COVID-19”, “Supervised Machine Learning”, “Unsupervised Machine Learning” and “COVID-19 prediction”.

3.3 Study selection

A set of criteria will be used to eliminate some prior research in order to decide which studies should be included and which should be excluded. If a research is duplicated in more than one digital library, only one will be picked, concentrate solely on research produced in English, disregard all other research, and studies that do not include COVID-19 DW will also be excluded. Additionally, according to these criteria, all studies outside of the study period (from 2019 to 2022) will be rejected, as will all studies that are irrelevant to the research issue. The search method for a systematic literature review is shown in Figure 2.

3.4 Data Extraction

Each research that was included in accordance with the standards of the systematic review is listed in detail in the next step [49]. This table's objective is to correctly and openly record the data gathered from preliminary investigations by researchers. For data extraction, this research will depend on the method described by Salvado, Nakasone, and Pow-Sang (2014) [50]. The following information is included in the strategy form: (1) Study title; (2) Author(s); (3) Publication type; (4) Date of extraction; (5) The digital library name where the study was located; and (6) Additional details on algorithmic precision and research scope. The numerical information that will be generated from this form is crucial for summarizing the findings of the earlier research.

3.5 Synthesis of the extracted data

The results from the included studies are gathered and summarized in this last stage of the systematic review process. There are three rounds in it: In the initial round, any studies that are pertinent to the topic of study will be considered. Restrict or choose research that, in general, uses supervised algorithms and educational prediction systems in the second round. Determine studies that satisfy all inclusion requirements and have the capacity to appropriately synthesize their findings to address the research question. In the methodological phases, 147 studies were retrieved that satisfied all requirements. Details about the chosen studies that were discovered during the search process are shown in Figure (3), and the whole extract information is described in the following link: https://drive.google.com/file/d/1Ln_NKfJYpI75F49BI3RrC3ytBGZlRnwJ/view?usp=sharing

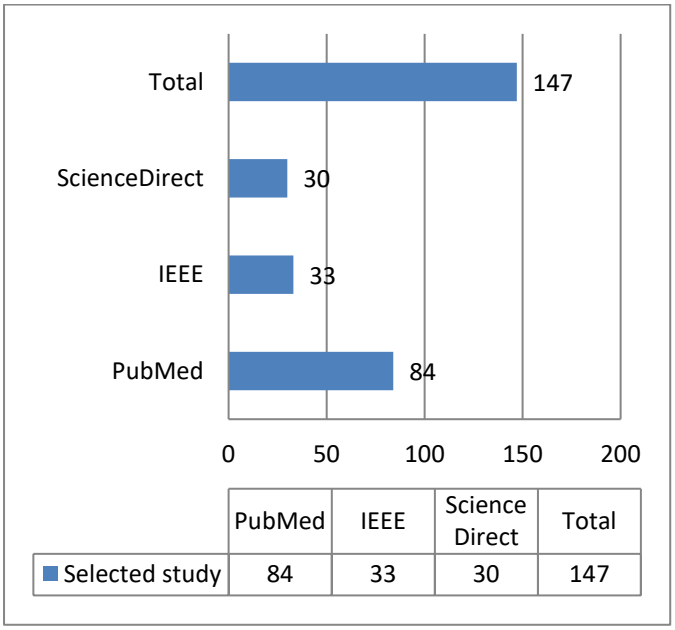

Figure 3: The total number of studies selected.

4. RESULTS AND DISCUSSION

The general analysis of the selected studied found the papers published in the journals (118) papers, while the papers published in the conferences (29) papers as shown in figure (4)-A. The year (2021) faced the most published papers with (80) papers, while the years (2020, and 2022) with (29,38) papers alternatively as shown in figure (4)-B. The supervised MLR algorithms take the most count in the published papers with (70) papers, while the unsupervised MLR with (45) papers. The papers that utilized (supervised and unsupervised) MLR are (15) while other papers with statistical and other MLR algorithms are (17) papers as shown in figure (4)-C.

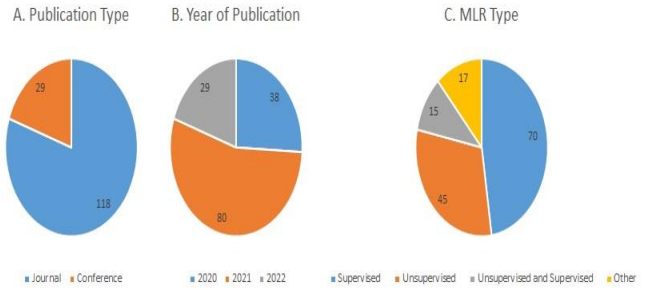

Figure 4: A. Publication Types, B. Year of Publication, C. MLR Type

Figure (5) shows the tree of supervised and unsupervised MLR algorithms in the selected studies. There are three decision tree algorithms (RF, C4.5, and CART) as supervised algorithms, while the Gradient Boosting with three algorithms (XGBoost (XGB), Gradient Boosting Decision Tree (GBDT), and Gradient Boosting Machine (GBM)). The algorithms and techniques in regression are (Logistic Regression (LR), Polynomial Regression (PR), Lasso Regression, Ridge Regression, Stepwise Regression, Multidimensional Regression with Interaction, Multivariate Regression Analysis, Multilinear, and Linear Regression). Naïve Bayes and Support Vector Machine (Support Vector Regression (SVR) were utilized in the selected studies. The clustering algorithms (KNN, and K-means clustering and correlation) are found in the study sample papers. Many types of Neural Networks (NN) algorithms were found such as (ANN, Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Multilayer Perceptron (MLP), Long-short Term Memory (LSTM), ResNEt-50, Mask-CNN, R-CNN, and Conventional MLR). Other proposed MLR were utilized in the selected study as shown in figure (5).

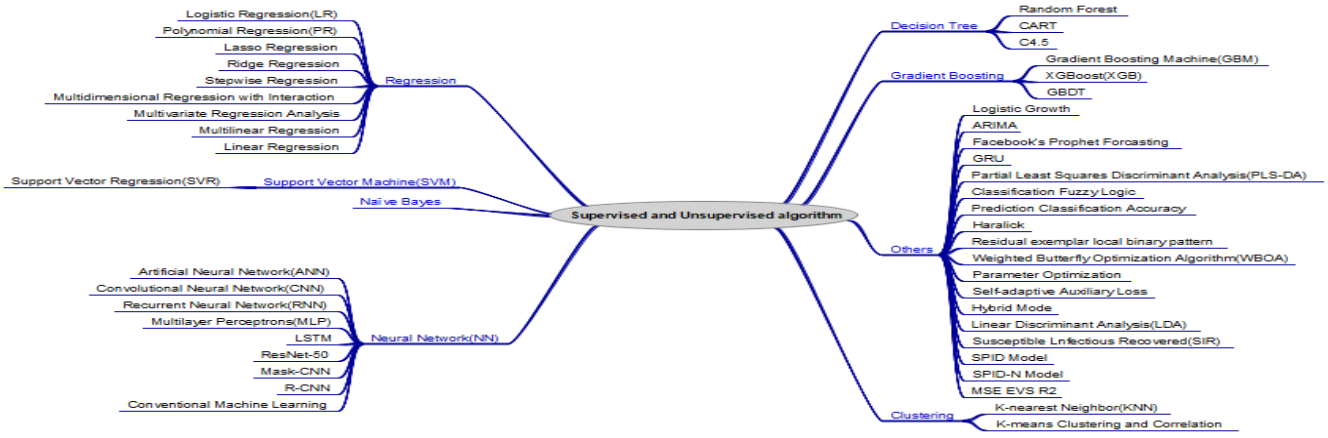

Figure 5: Supervised and Unsupervised MLR Algorithms

Clustering is the process of dividing large data into groups based on similarity so that data in that groups are similar to each other and dissimilar to the data in the other groups [51]. It is an important step in the analysis of a data set to understand its characteristics and prepare it to another process of analysis [52]. There are many methods used for grouping data that are different based on the values of the attribute such as Probabilistic and Generative Models, Feature Selection (FS) Methods, Leveraging Dimensionality Reduction Methods, Distance-Based Algorithms, Density- and Grid-Based Methods. The High Dimensional Scenario, Scalable Techniques for Cluster Analysis [53].

The structure of NN is an information processing structure that is distributed in parallel in the form of a directed graph, where the directed graph is defined as node objects (which consists of a set of points called nodes with links between them) [54]. Similar to the human's neural network, in a neural network the neuron is a central processing unit, which performs a mathematical operation that classifies information based on a specific architecture to generate the output [55]. Each neuron in a neural network perform a simple mathematical operation and then the entire neural network does the computation of the outputs of all the neurons. The simple neural network contains three layers: input layer , hidden layer, and output layer weights bias activation functions [55].

The supervised MLR method known as the SVM was proposed by Cortes and Vapnik. Finding the ideal decision boundary with a maximum margin hyperplane between the various classes of samples is the goal [56]. The algorithm iteratively generates the hyperplanes and provides the best class division and selection of the best plane among them. A point for a line that is remote from the characteristics of the data classes is chosen. The best selected hyperplane is used to determine the separation between each data type. SVM is divided into three parts: (a) model creation for classification or regression problems, (b) application of optimization techniques to new models and paradigms, (c) formulation and solution of optimization problems [57].

Decision Tree (DT) is one sort of supervised learning algorithm, is a type of data flow diagram where each leaf node represents a decision, a branch reflects a decision rule, and an inner node represents a function or attribute. The summit of the tree is where the root is located. Typically, it divides according to the attribute's value [58]. From the perspective of the variable to be predicted, DT aims to divide individuals into groups of people which are as similar as feasible [59]. The hierarchical connections between variables in a tree result are exposed. An iterative process is used where in each iteration a sub-population of individuals is obtained by selecting the explanatory variable that allows a better separation of individuals. The algorithm terminates when no further splits are feasible.

Regression is one of a data mining techniques (supervised MLRs) which is used to predict the numeric value of the final class. The statistical method is utilized to predict the numerical class of the dataset. The process of distribution trends identification in the training set is included. The linear regression (LR) is the simplest form of regression in which the relationship between two variables is predicted. The classification process in regression is formed by dividing the objects and assigning these objects into exhaustive and exclusive classes. The regression is utilized to predict the continuous values of classes [60][61][62][63].

Naive Bayes (NB) is one of the supervised MLRs that is used for classification and proves its performance in multiple domains. The assumption behind NB comes by assuming that the variables are conditional independent. Bayesian theorem is used in NB as a theory of mathematical probability thatis utilized to find the classification [64]. The following equation (1) states the Bayesian theorem:

$$P(A \mid B)=(P(B|A)*P(A))/(P(B))$$

Where P(A) represents the probability of (A) event, and event (B) is represented by P(B). P(A|B) for events (A, and B) represents the conditional probability of P(A) to probability of P(B). The classification for large datasets with NB is performed by combining the prior probability with the conditional probabilities to calculate the possible probabilities. NB produces an acceptable results despite assuming that the value of one attribute on a given probability is independent of the other features values probabilities [62][65].

Gradient Boosting Machines (GBMs) are considered as powerful techniques and have proved their success in wide range of applications. They can be customized by different applications needs that are learned with respect to loss functions. They were originally introduced and proposed to solve the classification problems by the ML community (MLC). The approach behind implementing GBM is combining several (weak learners) models to obtain (strong learners) with high prediction accuracy. The loss function is connected with the boosting algorithm by a statistical method in [66][67]. The GBM is an extended boosting machine with regression which was introduced by Friedman. The GBM can be

considered as an optimization numerical algorithm that find a way to minimize the loss function. GBM adds at each step a new (weak learner) to reduce the loss function. In regression, the DT is utilized as a first guess to initialize the model where the DT reduces the mean squared error (MSE). Then at each step, a new DT is added to the last model in order to update the residuals after fitting the current residuals. This process will continue until reaching the maximum iteration number [20][68], [69].

The systematic review questions conducted before were established to find the most used algorithms in the field of predicting, analyzing, and diagnosing COVID-19 infection. The literature of 147 papers utilized supervised and unsupervised ML algorithms as bases for their models for the same purpose. The impact of MLRs in predicting COVID-19 is measured by finding the accuracy of algorithms in predicting the COVID-19 infection and hence to reduce the efforts, logistics, and deaths.

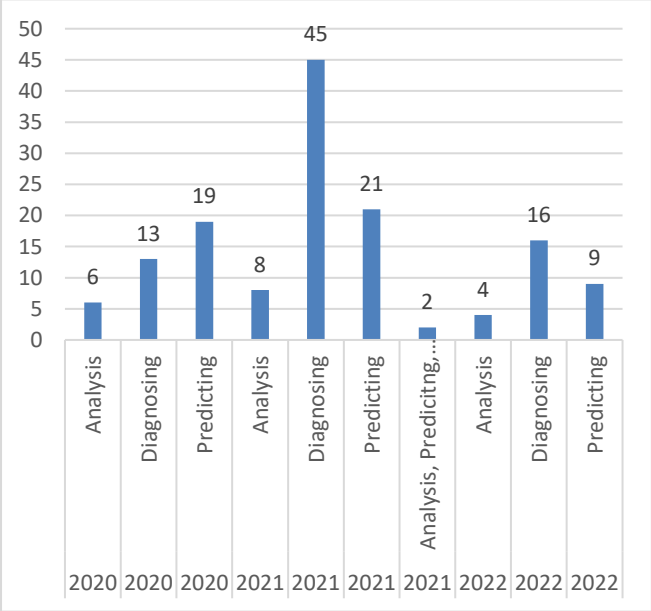

Figure 6: MLR sectors according to years

To answer the first question of the review related to the impact of using ML algorithms in predicting COVID-19, see Figure (6). According to figure (6), the field of diagnosing COVID-19 takes the intention of most researchers with 74 papers (13 papers in 2020; 45 papers in 2021; and 16 papers in 2022), while the analysis sector takes 18 papers (6 papers in 2020; 8 papers in 2021; and 4 papers in 2022). The total number of papers in the field of predicting COVID-19 is (49) with (19 papers in 2020; 21 papers in 2021; and 9 papers in 2022). The data types of the datasets that are utilized in this sector are: (Images for 19 papers, textual for 12 papers, while 19 papers with mixture of text and images datasets). The minimum rows of dataset utilized in this field is (229) while the maximum number are (110533973) rows. The performance evaluation, fast process and identification, quick learning, and accurate results of MLR algorithms make them a base for all models in analyzing, diagnosing, and predicting COVID-19 infection. The impact of using supervised and unsupervised MLR can be summarized by:

1. Estimating the spread level of COVID-19 to make the proper strategic decisions.
2. Real time prediction system of COVID-19 infection in different regions and countries according to previous data for the future directions.
3. Assisting ICUs in their work to handle different cases and take the pre-planned strategies.
4. Helping in making the strategies for overcoming the pandemic.
5. Analyzing the effect of the pandemic on future status.
6. Planning to control the pandemic by finding the accurate infected cases.
7. Implementing automatic detection, and diagnosing systems to reduce the workload in the medical centers.
8. Implementing early warning systems to give alerts for incoming infections.
9. Identifying the home quarantine impact on the psychological stability of individuals.
10. Discovering the factors that increase deaths.
11. Analyzing different aspects of the pandemic in the future.
12. Assessing the decisions for public policies.
13. Identification of the factors that reduce the infections.
14. Early diagnosing systems based on self-estimated symptoms.
15. Determining the required doses of COVID-19 vaccines.
16. Differentiating the COVID-19 infection from other lung infections.
17. Comparing the impact of datatypes on the diagnosing, predicting and analyzing COVID-19 infection severity.

18. Classifying COVID-19 infection types, time of infection, and time for recovery.

19. Helping doctors and hospitals to make a priority strategy for triaging patients and deaths in the cases when the hospitals are overrun.

20. Predicting the moralities in the ICUs.

The next question which handles the sector that faced a lot of attention in the field of predicting COVID-19, the diagnosing COVID-19 field takes the attention of the most researchers.

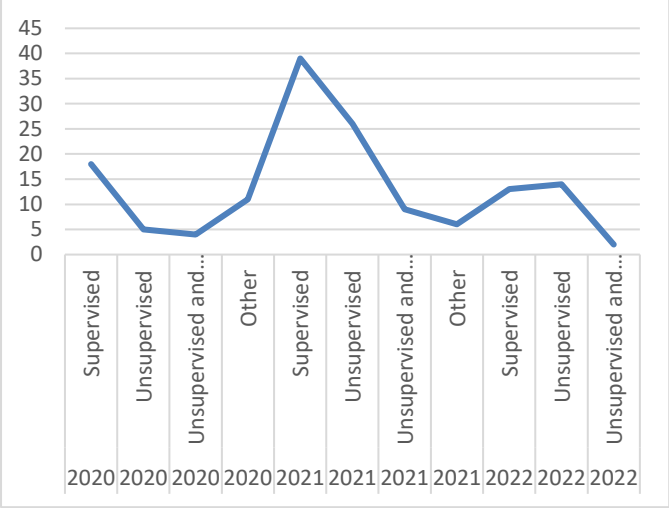

Figure 7: Count of MLR types according to year for all papers

The last question that finds the most supervises/unsupervised ML algorithm that used for predicting COVID-19, according to figure (7), the supervised ML algorithms are the most algorithms that utilized for diagnosing, predicting, and analyzing COVID-19 infection. The total number of papers is (70) with 18 papers using supervised MLR algorithms in 2020; 39 papers in 2021; while 13 papers in 2022. The unsupervised MLRs are utilized in (45) papers with (5 papers in 2020; 26 in 2021; and 14 in 2022). The supervised MLR is the most used sector in the diagnosing, predicting and analyzing the COVID-19 infection.

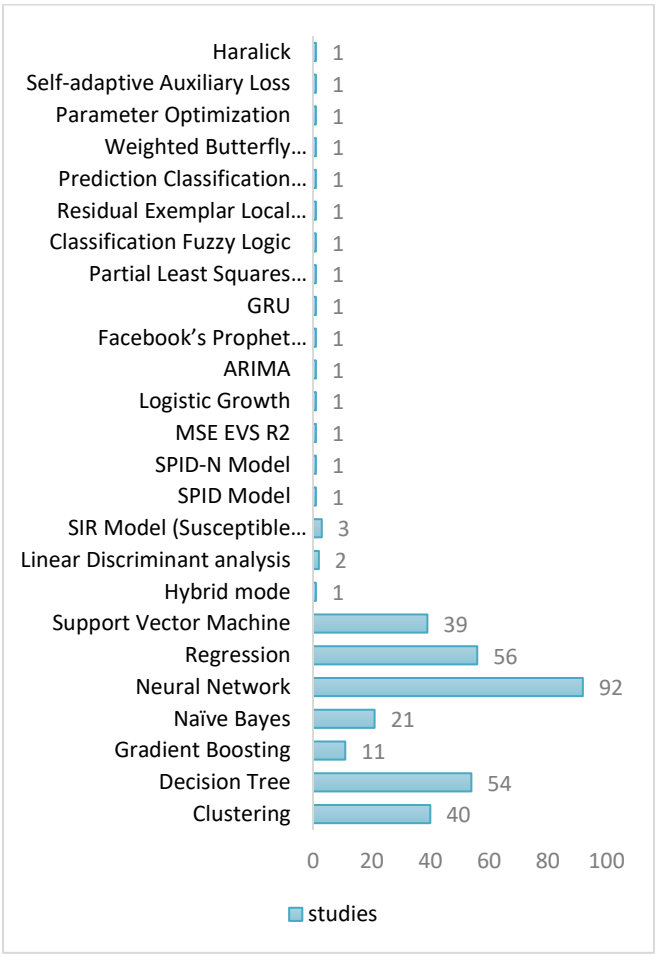

Figure 8: The use of algorithms in the selected studies

According to figure (8), NN is the most used algorithm in 92 papers, while Regression comes next in 56. The DTs are utilized in 54 papers, while Clustering and SVM in (40, and 39) papers alternatively. Naïve Bayes and Gradient Boosting are found in (21, and 11) alternatively, while other proposed or supervised/unsupervised algorithms in (21) papers. Many researchers implemented their models using different supervised and unsupervised algorithms. The analysis of the selected studies found that the NN is the most used algorithm compared with other algorithms.

5. CONCLUSION AND FUTURE WORKS

The analysis, diagnosing and predicting of COVID-19 (SARS-CoV-2) became the most active topic since the first appearance of the COVID-19 corona virus in Wuhan, China, at the end of 2019. Since that, the whole regions of the world need any tool or system to help in supporting decisions related to medical treatment, and timely quarantine. MLRs as a base and supporting algorithms proved their accuracy and efficiency in

different sectors (education, health, and security) for different purposes such as prediction, classification, and analyzing. In this paper, three main questions will be answered related to COVID-19 analysis, predicting, and diagnosing. The performance evaluation, fast process and identification, quick learning, and accurate results of MLR algorithms make them a base for all models in analyzing, diagnosing, and predicting COVID-19 infection. The impact of using supervised and unsupervised MLR can be used for estimating the spread level of COVID-19 to make the proper strategic decisions. Real time prediction systems of COVID-19 infections are implemented in different regions and countries according to previous data for the future directions. The proposed systems based on MLRs are utilized for assisting ICUs in their work to handle different cases and take the pre-planned strategies. They can be used for helping in making the strategies for overcoming the pandemic and analyzing the effect of the pandemic on the future status. Planning to control the pandemic by finding the accurate infected cases can also be a need that accomplished by such systems. MLRs are also utilized for implementing automatic detection, and diagnosing systems to reduce the workload in the medical centers and implementing early warning systems to give alerts for the incoming infections. The pandemic makes a need for systems for identifying the home quarantine impact on the psychological stability of individuals, and discovering the factors that increase deaths. The intensive care units (ICUs) in the pandemic also found the importance of MLRs by the proposed systems that help doctors and hospitals to make a priority strategy for triaging patients and deaths in the case when the hospitals are overrun. Predicting the moralities in the ICUs. Classifying COVID-19 infection types, time of infection, and time for recovery are other sectors that help ICUs. Governments and decision-makers found the impact of MLRs by analyzing different aspects for the pandemic in the future, assessing the decisions for the public policies, and identification the factors that reduce the infections in order to make the right strategies. The medical institutions also found the impact of MLRs by systems that early diagnosis systems based on self-estimated symptoms, determining the required doses of COVID-19 vaccines, and differentiating the COVID-19 infection from other lung infections. The researchers then investigated the impact of MLRs by comparing the impact of datatypes on the diagnosing, predicting, and analyzing of COVID-19 infection severity. The analysis of the selected study found that there are three fields related to COVID-19 (analysis, diagnosing, and predicting). The diagnosing of COVID-19 field takes the attention of most researchers where they implemented their models to diagnose the infection. Different algorithms were utilized in the selected studies, while the analysis found that the NN is the most used algorithm compared with other algorithms. The supervised MLR is the most used sector in the diagnosing, predicting and analyzing of COVID-19 infection. The reinforcement and semi-supervised MLRs will be investigated to find the impact of such algorithms on COVID-19 diagnosing, analyzing and prediction. A comparison among all MLRs can be implemented to find the most accurate with high performance sector and algorithm in diagnosing and predicting COVID-19 infection.

REFERENCES

[1] "COVID Live - Coronavirus Statistics - Worldometer." https://www.worldometers.info/coronavirus/ (accessed Oct. 05, 2022).

[2] M. A. Alves et al., "Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs," Comput. Biol. Med., vol. 132, p. 104335, 2021.

[3] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A. J. Aljaaf, "A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science," 2020.

[4] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in Proceedings of the 10th INDIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016, 2016.

[5] A. K. Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis," Int. J. Interact. Multimed. Artif. Intell., 2018, doi: 10.9781/ijimai.2018.02.004.

[6] S. Karlos, G. Kostopoulos, and S. Kotsiantis, "Predicting and interpreting students' grades in distance higher education through a semi-regression method," Appl. Sci., vol. 10, no. 23, 2020, doi: 10.3390/app10238413.

[7] A. Hamoud, "Applying association rules and decision tree algorithms with tumor diagnosis data," Int. Res. J. Eng. Technol., vol. 3, no. 8, pp. 27–31, 2017.

[8] A. Khalaf, A. Majeed, W. Akeel, and A. Salah, "Students' Success Prediction based on Bayes Algorithms," Int. J. Comput. Appl., vol. 178, no. 7, pp. 6–12, Nov. 2017, doi: 10.5120/ijca2017915506.

[9] A. Hamoud and A. Humadi, "Student's success prediction model based on artificial neural networks (ANN) and a combination of feature selection methods," J. Southwest Jiaotong Univ., vol. 54, no. 3, 2019.

[10] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, "Student Performance Prediction Model based on Supervised Machine Learning Algorithms," in IOP Conference Series: Materials Science and Engineering, 2020, vol. 928, no. 3, p. 32019.

[11] H. Hassan, N. B. Ahmad, and S. Anuar, "Improved students' performance prediction for multi-class imbalanced problems using hybrid and ensemble approach in educational data mining," in Journal of Physics: Conference Series, 2020, vol. 1529, no. 5, p. 52041.

[12] B. Lev, "The deteriorating usefulness of financial report information and how to reverse it," Account. Bus. Res., vol. 48, no. 5, pp. 465–493, 2018.

[13] A. Alshehhi, H. Nobanee, and N. Khare, "The impact of sustainability practices on corporate financial performance: Literature trends and future research potential," Sustainability, vol. 10, no. 2, p. 494, 2018.

[14] A. K. Hamoud, "Applying Association Rules and Decision Tree Algorithms with Tumor Diagnosis Data," Int. Res. J. Eng. Technol., vol. 3, no. 8, pp. 27–31, 2016.

[15] G. Jia, H. K. Lam, and Y. Xu, "Classification of COVID-19 chest X-Ray and CT images using a type of dynamic CNN modification method," Comput. Biol. Med., vol. 134, 2021, doi: 10.1016/j.compbiomed.2021.104425.

[16] I. A. Najm, A. K. Hamoud, J. Lloret, and I. Bosch, "Machine learning prediction approach to enhance congestion control in 5G IoT environment," Electronics, vol. 8, no. 6, p. 607, 2019.

[17] T. Abar, A. Ben Letaifa, and S. El Asmi, "Machine learning based QoE prediction in SDN networks," in 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), 2017, pp. 1395–1400.

[18] A. Darabseh and P. Khethavath, "Supervised and Unsupervised Outlier Detection Machine Learning Algorithms in Wireless Sensor Networks," Int. J. Bus. Inf. Syst., vol. 1, no. 1, 2021, doi: 10.1504/ijbis.2021.10042585.

[19] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, and G. S. Choi, "A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis," PLoS One, vol. 16, no. 2, 2021, doi: 10.1371/journal.pone.0245909.

[20] K. F. Hew, X. Hu, C. Qiao, and Y. Tang, "What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach," Comput. Educ., vol. 145, 2020, doi: 10.1016/j.compedu.2019.103724.

[21] J. Fang, F. Yang, R. Tong, Q. Yu, and X. Dai, "Fault diagnosis of electric transformers based on infrared image processing and semi-supervised learning," Glob. Energy Interconnect., vol. 4, no. 6, 2021, doi: 10.1016/j.gloei.2022.01.008.

[22] U. Delli and S. Chang, "Automated Process Monitoring in 3D Printing Using Supervised Machine Learning," 2018, vol. 26, doi: 10.1016/j.promfg.2018.07.111.

[23] S. Khatoon et al., "Development of social media analytics system for emergency event detection and crisismanagement," Comput. Mater. Contin., vol. 68, no. 3, 2021, doi: 10.32604/cmc.2021.017371.

[24] G. K. Palshikar, M. Apte, and D. Pandita, "Weakly Supervised and Online Learning of Word Models for Classification to Detect Disaster Reporting Tweets," Inf. Syst. Front., vol. 20, no. 5, 2018, doi: 10.1007/s10796-018-9830-2.

[25] M. A. Kumar and A. J. Laxmi, "Machine Learning Based Intentional Islanding Algorithm for DERs in Disaster Management," IEEE Access, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3087914.

[26] X. Jiang et al., "Rapid and large-scale mapping of flood inundation via integrating spaceborne synthetic aperture radar imagery with unsupervised deep learning," ISPRS J. Photogramm. Remote Sens., vol. 178, 2021, doi: 10.1016/j.isprsjprs.2021.05.019.

[27] H. U. Dike, Y. Zhou, K. K. Deveerasetty, and Q. Wu, "Unsupervised Learning Based On Artificial Neural Network: A Review," in 2018 IEEE International Conference on Cyborg and Bionic Systems, CBS 2018, 2019, doi: 10.1109/CBS.2018.8612259.

[28] A. K. HAMOUD, "CLASSIFYING STUDENTS'ANSWERS USING CLUSTERING ALGORITHMS BASED ON PRINCIPLE COMPONENT ANALYSIS.," J. Theor. Appl. Inf. Technol., vol. 96, no. 7, 2018.

[29] O. Iatrellis, I. K. Savvas, P. Fitsilis, and V. C. Gerogiannis, "A two-phase machine learning approach for predicting student outcomes," Educ. Inf. Technol., vol. 26, no. 1, pp. 69–88, 2021.

[30] Y. Auh and H. R. Sim, "Uses of social network topology and network-integrated multimedia for designing a large-scale open learning system: case studies of unsupervised featured learning platform Design in South Korea," Multimed. Tools Appl., vol. 78, no. 5, 2019, doi: 10.1007/s11042-018-6658-1.

[31] J. Kebede, A. Naranpanawa, and S. Selvanathan, "Financial inclusion: Measures and applications to Africa," Econ. Anal. Policy, vol. 70, 2021, doi: 10.1016/j.eap.2021.03.008.

[32] V. Babenko, A. Panchyshyn, L. Zomchak, M. Nehrey, Z. Artym-Drohomyretska, and T. Lahotskyi, "Classical machine learning methods in economics research: Macro and micro level examples," WSEAS Trans. Bus. Econ., vol. 18, 2021, doi: 10.37394/23207.2021.18.22.

[33] Y. Chang, X. Yuan, B. Li, D. Niyato, and N. Al-Dhahir, "A joint unsupervised learning and genetic algorithm approach for topology control in energy-efficient ultra-dense wireless sensor networks," IEEE Commun. Lett., vol. 22, no. 11, 2018, doi: 10.1109/LCOMM.2018.2870886.

[34] H. Li, D. Caragea, C. Caragea, and N. Herndon, "Disaster response aided by tweet classification with a domain adaptation approach," J. Contingencies Cris. Manag., vol. 26, no. 1, 2018, doi: 10.1111/1468-5973.12194.

[35] B. Ma, Y. Zhu, X. Yin, X. Ban, H. Huang, and M. Mukeshimana, "SESF-Fuse: an unsupervised deep model for multi-focus image fusion," Neural Comput. Appl., vol. 33, no. 11, 2021, doi: 10.1007/s00521-020-05358-9.

[36] W. Wöber, P. Tibihika, C. Olaverri-Monreal, L. Mehnen, P. Sykacek, and H. Meimberg, "Comparison of Unsupervised Learning Methods for Natural Image Processing," Biodivers. Inf. Sci. Stand., vol. 3, 2019, doi: 10.3897/biss.3.37886.

[37] H. Jigneshkumar Patel, J. Prakash Verma, and A. Patel, "Unsupervised Learning-Based Sentiment Analysis with Reviewer's Emotion," in Lecture Notes in Electrical Engineering, 2021, vol. 694, doi: 10.1007/978-981-15-7804-5_6.

[38] Z. Li, Y. Fan, F. Wang, and W. Liu, "Unsupervised feature learning assisted visual sentiment analysis," Int. J. Multimed. Ubiquitous Eng., vol. 11, no. 10, 2016, doi: 10.14257/ijmue.2016.11.10.11.

[39] A. de Fátima Cobre et al., "Diagnosis and prediction of COVID-19 severity: can biochemical tests and machine learning be used as prognostic indicators?," Comput. Biol. Med., vol. 134, p. 104531, 2021.

[40] W. M. Shaban, A. H. Rabie, A. I. Saleh, and M. A. Abo-Elsoud, "A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier," Knowledge-Based Syst., vol. 205, p. 106270, 2020.

[41] P. R. S. Dos Santos et al., "Prediction of COVID-19 using time-sliding window: the case of Piauí state-Brazil," in 2020 IEEE International Conference on E-Health Networking, Application & Services (HEALTHCOM), 2021, pp. 1–6.

[42] H. C. Çubukçu, D. İ. Topcu, N. Bayraktar, M. Gülşen, N. Sarı, and A. H. Arslan, "Detection of COVID-19 by machine learning using routine laboratory tests," Am. J. Clin. Pathol., vol. 157, no. 5, pp. 758–766, 2022.

[43] P. Hemingway and N. Brereton, "What is a systematic review', What is Series," Bandolier, April, 2009.

[44] J. H. Littell, J. Corcoran, and V. Pillai, Systematic reviews and meta-analysis. Oxford University Press, 2008.

[45] H. Cooper, A. B. Allen, E. A. Patall, and A. L. Dent, "Effects of full-day kindergarten on academic achievement and social development," Rev. Educ. Res., vol. 80, no. 1, pp. 34–70, 2010.

[46] A. M. O'Connor, K. M. Anderson, C. K. Goodell, and J. M. Sargeant, "Conducting systematic reviews of intervention questions I: writing the review protocol, formulating the question and searching the literature," Zoonoses Public Health, vol. 61, pp. 28–38, 2014.

[47] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," 2007.

[48] D. Gough, S. Oliver, and J. Thomas, "An introduction to systematic reviews. 2017." Sage.

[49] M. Petticrew and H. Roberts, Systematic reviews in the social sciences: A practical guide. John Wiley & Sons, 2008.

[50] C. Salvador, A. Nakasone, and J. A. Pow-Sang, "A systematic review of usability techniques in agile methodologies," in Proceedings of the 7th Euro American Conference on Telematics and Information Systems, 2014, pp. 1–6.

[51] H. Jiawei, K. Micheline, and P. Jian, "Data mining concepts and techniques." 2016.

[52] J. Sander, X. Qin, Z. Lu, N. Niu, and A. Kovarsky, "Automatic extraction of clusters from hierarchical clustering representations," in Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2003, pp. 75–87.

[53] C. C. Aggarwal and C. K. Reddy, "Data clustering," Algorithms Appl. Chapman&Hall/CRC Data Min. Knowl. Discov. Ser. Londra, 2014.

[54] N. Yadav, A. Yadav, and M. Kumar, An introduction to neural network methods for differential equations, vol. 1. Springer, 2015.

[55] G. Ciaburro and B. Venkateswaran, Neural Networks with R: Smart models using CNN, RNN, deep learning, and artificial intelligence principles. Packt

Publishing Ltd, 2017.

[56] S. A.-F. Sayed, A. M. Elkorany, and S. S. Mohammad, "Applying different machine learning techniques for prediction of COVID-19 severity," Ieee Access, vol. 9, pp. 135697–135707, 2021.

[57] L. W. Mary and S. A. A. Raj, "Machine Learning Algorithms for Predicting SARS-CoV-2 (COVID-19)–A Comparative Analysis," in 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), 2021, pp. 1607–1611.

[58] F. Ullah, J. Moon, H. Naeem, and S. Jabbar, "Explainable artificial intelligence approach in combating real-time surveillance of COVID19 pandemic from CT scan and X-ray images using ensemble model," J. Supercomput., pp. 1–26, 2022.

[59] 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," SN Comput. Sci., vol. 3, no. 4, pp. 1–35, 2022.

[60] E. Frank, M. A. Hall, and I. H. Witten, "The WEKA Workbench Data Mining: Practical Machine Learning Tools and Techniques," Morgan Kaufmann, Fourth Ed., 2016.

[61] I. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques Third Edition. Elsevier," 2012.

[62] D. J. Hand, "Principles of data mining," in Drug Safety, 2007, vol. 30, no. 7, doi: 10.2165/00002018-200730070-00010.

[63] C. C. Aggarwal, Data mining: the textbook. Springer, 2015.

[64] A. Hamoud, A. Humadi, W. A. Awadh, and A. S. Hashim, "Students' success prediction based on Bayes algorithms," Int. J. Comput. Appl., vol. 178, no. 7, pp. 6–12, 2017.

[65] C. Y. Ko and F. Y. Leu, "Examining Successful Attributes for Undergraduate Students by Applying Machine Learning Techniques," IEEE Trans. Educ., vol. 64, no. 1, 2021, doi: 10.1109/TE.2020.3004596.

[66] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Ann. Stat., vol. 29, no. 5, 2001, doi: 10.1214/aos/1013203451.

[67] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," Annals of Statistics, vol. 28, no. 2. 2000, doi: 10.1214/aos/1016218223.

[68] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," Front. Neurorobot., vol. 7, no. DEC, 2013, doi: 10.3389/fnbot.2013.00021.

[69] S. Touzani, J. Granderson, and S. Fernandes, "Gradient boosting machine for modeling the energy consumption of commercial buildings," Energy Build., vol. 158, 2018, doi: 10.1016/j.enbuild.2017.11.039.