Archive of SID

http:// ijp.mums.ac.ir
Original Article (Pages: 12723-12737)

COVID-19 Prediction Classifier Model Using Hybrid Algorithms in Data Mining

*Morteza Nikooghadam[1], Adel Ghazikhani[1], Mohammad Saeedi[2]

[1]Assistant Professor, Department of Computer Engineering and Information Technology, Imam Reza International University, Mashhad, Iran.

[2]MS Student, Department of Computer Engineering and Information Technology, Imam Reza International University, Mashhad, Iran.

Abstract

Increase of stored data in medical databases needs allocative tools to get access to data, data mining, discover knowledge and efficient use of data. Medical and treatment fields are two examples of data mining tools to analyze massive data and predictive modelling. In medical sciences, prediction and precise-quick detection of multiple diseases has to reduced exprense and also save people's lives. Group based methods (Ensemble Methods) are approaches that use hybrid models to recover classification. Coronavirus (COVID-19) has killed many people around the world so far, and this could be a good reason to present a new model for diagnosing the disease using data mining algorithms. This research presents a hybrid model of basic data mining and hybrid algorithms according to information in medical and laboratory records of patients suffering Covid-19 in Emam-Reza (AS) hospital in Mashhad, Iran, to diagnose the sickness. The proposed method uses Ensemble base (hybrid) classifiers, where the general model can be used to provide diagnoses with higher precision rather than classifiers. To execute the proposed model, data mining tools including Rapid Miner 9.7 and Python 3.7 were used. This study used stacking classifiers composed of basic algorithms including simple baze, decision tree, K- nearest neighborhood backup vector machine for basic section and uses chaos jungle algorithm in stack section that has gained 86.5% accuracy for diagnosis of Covid-19.

Key Words: Accuracy, Covid-19, Classifier model, Data mining, Hybrid data mining.

*Please cite this article as: Nikooghadam M, Ghazikhani A, Saeedi M. COVID-19 Prediction Classifier Model Using Hybrid Algorithms in Data Mining. Int J Pediatr 2021; 9(1): 12723-737. DOI: **10.22038/ijp.2020.54272.4290**

Corresponding Author:

Dr. Morteza Nikooghadam, Department of Computer Engineering and Information Technology, Imam Reza International University, Mashhad, Iran.

Email: Morteza.nikooghadam@gmail.com

Received date: Apr.20, 2020; Accepted date: Nov. 12, 2020

www.SID.ir

Archive of SID

1- INTRODUCTION

Respiratory and pulmonary disease is already among the 10 top factors causing death and disabilities, manifested as lung or cords infections. The cause of lung infection could be Influenza, pneumonia, tuberculosis, and bronchitis (1). Coronaviruses are a large family of viruses that results in many illnesses with mild and severe symptoms. SARS and MERS are known from coronavirus with severe symptoms. Corona pandemic (COVID-19) is the newly advanced kind of coronavirus that has transferred from animals to humans. One of the factors increasing the virus outbreak that caused illness in our country is the people's lack of knowledge (2). The regional situation of COVID-19 victims and patients worldwide is varied (**Figure.1**). The healthcare industry can be considered as a site with a rich set of data because it generates massive data, including electronic medical history, administrative reports, and other finding branches. Medical data mining is an effective method to uncover the hidden pattern of big raw data in the medical field. Disease prediction and diagnostic systems can reduce the costs of disease, waiting time, and human errors (3).

One of the most significant and essential data mining stages is Knowledge Discovery in Databases (KDD) (4). Data mining and health care sciences create several reliable systems that involve early prediction and related systems for health care from clinical and diagnostics data. Data mining is based on Machine Learning, Artificial Intelligence, statistics, and probability (3). Currently, limited research was conducted to predict and diagnose Coronavirus (COVID-19) with blood experiments of patients (RT-PCR). However, there are numerous studies employing data mining algorithms to diagnose Coronavirus (COVID-19) disease with patients' lung scans (5-18). Eom et al. (2007) proposed a supporting and group-based clinical decision system to predict cardiovascular disease using four machine learning classifiers, including Decision Tree (DT), Bayes Network, Neural Network, and Support Vector Machine (SVM). The resulted accuracy was 94% (5). Chaurasia and Pal (2013) suggested a model for classification of medical data and diagnosing heart diseases employing the Bagging algorithm with 85% accuracy (6). Bashir et al. (2014) presented a group-based method to precisely predict heart disease by employing a compound method with Naïve Bayes, Decision Tree, SVM, and majority voting technique that led to 81% accuracy (7). Elshazly et al. (2015) reached an early and accurate diagnosis of glaucoma (chronic eye disease) integrating Principal Component Analysis (PCA) and Rotational Tree (ROT).

Evaluating performance was carried out with three well-known classifiers including Neural Network, Decision Tree, and Fuzzy logic classifier. The ROT model has achieved 86% accuracy (8). Kurdia et al. (2017) introduce an accurate classifier system to predict MERS-COV disease infection by Naïve Bayes, Decision Tree, K-NN, binary classifiers, and multi-class multi-label classifiers. The Decision Tree algorithm achieved 90% accuracy, the K-NN algorithm in multi-class classification achieved 60% accuracy, and Naïve Bayes achieved 77% accuracy (9).

Arafat et al. (2018) announced a classifier model to predict chronic kidney diseases automatically with clinical data using machine learning algorithms such as Naïve Bayes, Decision Tree, and Random Forest. By comparing all algorithms' results, they understood that the Random Forest method has the best performance with 98% accuracy (10). Kurian et al. (2018) suggested a group-based model for predicting heart diseases combining Naïve Bayes, Decision Tree, and K-NN algorithm. Their model achieved 90% accuracy (11). AlMoammar et al. (2018)

www.SID.ir

Archive of SID

innovated a new model to increase MERS-COV disease accuracy by hiring data mining algorithms K-NN, DT, and SVM. SVM and K-NN algorithms obtain anaccuracy higher than 86% (12). Wang et al. (2019) provided a model to predict COVID-19 disease using deep learning techniques based on radiographic scan changes in COVID patients. The suggested model achieved 82.9% accuracy (13). Narin et al. (2020) employed three neural network-based models, including ResNet50, InceptionV3, and Inception-ResNetV2, to predict COVID-19 automatically. ResNet50 achieved 98% accuracy and showed higher performance compared to two others (14). Barstugan et al. (2020) used CT images of COVID-19 patients to diagnose their disease. Five feature extraction methods have been used to find a feature set that separated contaminated stains with high accuracy. The accuracy of this model classifier is 99% (15). Wiguna and Riana (2020) provided a model that classifies three categories (supervised patients, suspected and asymptomatic individuals) with a C4.5 Decision Tree and achieved 92.8% accuracy (16). Muhammad et al. (2020) predict the recovery of COVID-19 patients by the epidemiologic dataset of South Korean patients and algorithms such as Naïve Bayes, DT, SVM, logistic regression, Random Forest, K-NN in python language. Results showed that this model could predict patients' recovery with 99.8% accuracy (17). Khanday et al. (2020) provided a model for textual clinical reports using basic and hybrid algorithms. These reports have been categorized into four groups [acute respiratory disease, SARS, COVID, and both (acute respiratory disease and COVID)], and various features such as Term Frequency/Inverse Document Frequency and Bags of Words have been extracted from these reports. Finally,Eventually, logistics regression and multivariate Naïve Bayes achieved the highest accuracy, which is 96.2% (18).

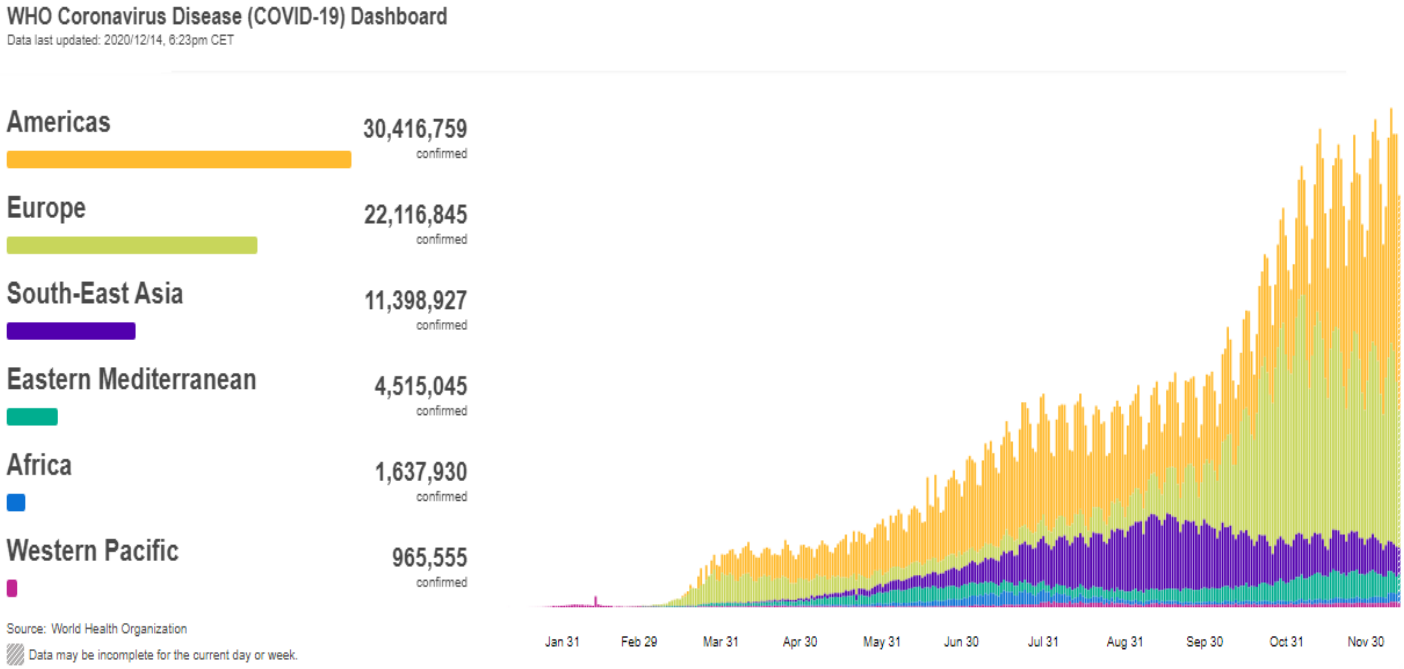


Fig.1: The regional situation of COVID-19 victims and patients worldwide.
(Based on the World Health Organization's (WHO) report, last updated 2020/12/14).

www.SID.ir

Archive of SID

2- MATERIALS AND METHODS

In this cross-sectional study, we investigated and compared basic algorithms and Hybrid algorithms in terms of accuracy, precision, recall, and F-measure benchmarks. We also evaluated the results using the cross-validation method. RapidMiner version 9.7, a data mining tool, was used for preprocessing and experimenting with other algorithms.

2-1. Database

We used the electronic and clinical information of patients with suspected or confirmed COVID-19 in Emam-Reza hospital in Mashhad city, Iran. The employed dataset in this research included 200 cases (124 Male, 76 Female), 29 features, and one diagnostic class. The features comprise clinical information, underlying medical conditions, and blood experiment of suspected and confirmed COVID-19 including gender, age, fever, cough, shortness of breath, feverish cold, body pain, hypertension, diabetes, cardiovascular, pulmonary, other diseases, no medical history, Alanine Aminotransferase (ALT), Aspartate transaminase (AST), Albumin, C-Reactive Protein, Lactate dehydrogenase (CRP), Neutrophil (NEU), Urea, and Diagnosis. Among 29 features, we eliminated those with less significance in the disease diagnosis using Relief feature selection methods, and finally, we experimented and evaluated 20 prominent features and one diagnosis class.

2-2. Preparing and pre-processing procedure

2-2-1. Preparing process

In general, data scientists spend most of their time processing data. This procedure consists of selecting appropriate features, cleaning, and preparing them to become inputs or independent variables for a machine learning model.

a. Data cleaning

The most critical stages in this section are estimating unavailable data in the database, removing noise in data, eliminating outlier and unrelated data, eliminating data anomaly

b. Data integration

In most cases, data have been kept in different files and resources. It is required to integrate data before applying data mining techniques. This phase comprises removing missing values, outliers, and duplicate data.

c. Data reduction

All data are not always demanded in data mining, and only part of the data needs to be processed.

d. Data transformation

Since data are provided through sources that generate or keep data regardless of data mining procedures, it is required to prepare data considering the condition and the given problem for data mining algorithm insertion. To prepare data, we should transform them from their initial form to a suitable one for the algorithm (19).

2-2-2. Feature selection

Big datasets establishment and their requirements for machine learning techniques is a significant challenge. To address this issue, novel approaches are in demand. Feature selection in machine learning refers to selecting the best features in our data to provide for our model (**Figure.2**), (20).

www.SID.ir

Archive of SID

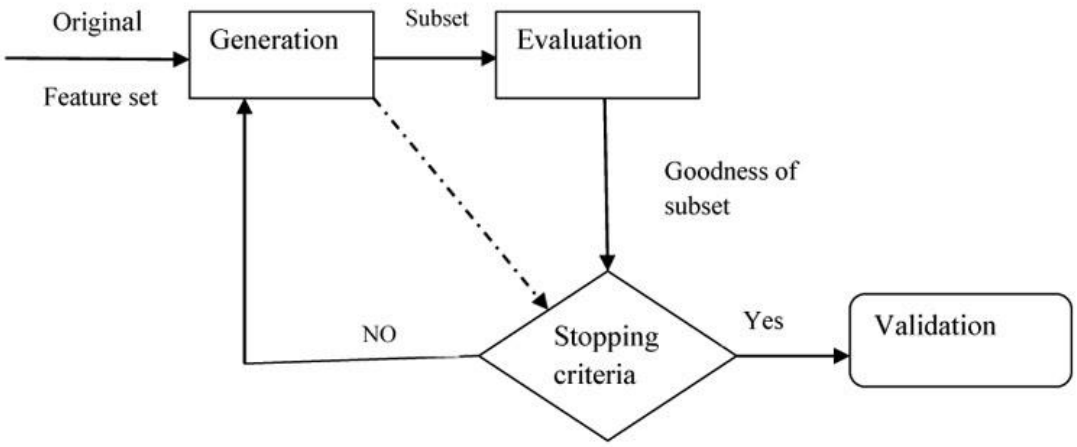


Fig.2: Validation feature selection procedure (21).

2-2-3. Relief feature selection

This method is widely used for filter-based selecting features. In each iteration, this algorithm selects an instance from the available instances in the dataset randomly. It then updates the feature relevance based on the difference between selected instances and two close neighbor instances (22):

$$W_i = W_i - (x_i\text{-nearHit}_i)^2 + (x_i\text{-nearMiss}_i)^2$$

In this paper, we used the Relief feature selection method in Weka software version 3.9. Also, we employed 20 useful features and one diagnosis class in our proposed model.

2-3. Classification in data mining

The classification procedure tries to find a model that discriminates and describes classes and data concepts. It is a data analyzing task.

A. Classifier models

This paper uses binary classifiers and basic and hybrid algorithms to improve the prediction's performance and more accurate diagnosis of COVID-19. In the following, we explain each algorithm briefly.

A1. Basic classification models

This section presents the classification algorithm, such as Naïve Bayes, Decision Tree, K-Nearest-Neighbor, and Support Vector Machine.

A2. Naïve Bayes

This algorithm is generally used for clustering and classification. The fundamental architecture of Naïve Bayes depends on conditional probability. Naïve Bayes classifier predicts the class values considering feature sets (23).

Bayes' theorem:

$$P\left(\frac{Y}{X}\right) = \frac{P(X,Y)}{P(X)} = P\left(\frac{X}{Y}\right) P(Y)/P(X)$$

Bayes' classifier:

$$P(\omega j, x) = \frac{P(\omega j, x)}{P(x)} = P\left(\frac{X}{\omega j}\right) P(\omega j) / \sum_{j=1}^{c} P\left(\frac{X}{\omega j}\right) P(\omega j)$$

A3. Decision Tree

This algorithm is generally used for clustering and classification. The fundamental architecture of Naïve Bayes depends on conditional probability. Naïve Bayes classifier predicts the class values considering feature sets (23).

A4. K-Nearest-Neighbors

www.SID.ir

Archive of SID

K-Nearest-Neighbors algorithm (K-NN) is a simple supervised machine learning algorithm and a conventional non-parametric classifier that can be used for classification and regression problems. The classifier's central part is based on measuring similarity or distances between test and training examples. To detect nearest neighbors, we used various distance measuring techniques in which Euclidean distance, Manhattan distance, and cosine distance are the most traditional methods (23).

A5. Support Vector Machine

This algorithm, which is called SVM, works based on measuring margins. Basically, it considers a margin between classes that maximizes the distance between margins and classes and minimizes the classification error. In the **Figure.3**, data are indicated with blue and red dotted circles. The corresponding support vectors for each category are shown with double border circles, and the solid black line is the SVM. Each support vector has a unique formula that describes the border for each category (23).

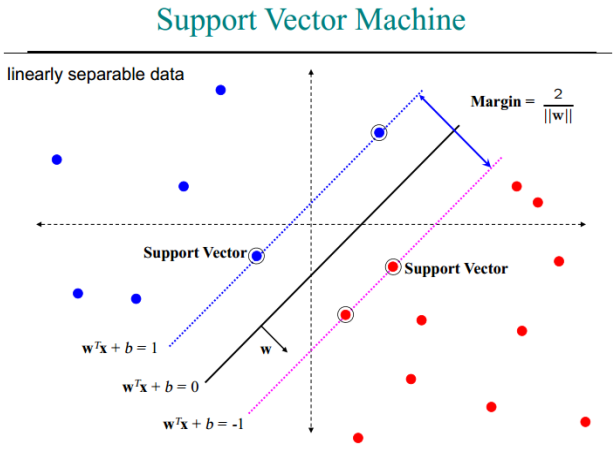


Fig.3. Linear SVM (23).

B. Hybrid classification models

Group-based learning algorithm (hybrid) contributes to machine learning results by combining several methods. This method provides a better prediction performance rather than a single model.

B1. Bootstrap Aggregating

Briefly, this algorithm, which is called "Bagging" is a two-phase approach. First, the subsets of primary data are used to generate mild performance models, and then, their performance is aggregated by combining these models using a particular cost function (Majority Voting). The bagging algorithm also reduces the variance and helps to reduce overfitting (24).

B.2 Random Forest algorithm

Trees learning is a popular basic model of group-based methods. The forest consists of trees that could be shallow or deep. Shallow trees have lower variance and higher deviation. On the other side, deep trees have lower deviation and higher variance. As a result, we understand that in general, Bootstrap aggregating methods focus on reducing variance (25).

B3. Boosting algorithm

This algorithm creates a set of weak learning algorithms and converts them to a

www.SID.ir

Archive of SID

strong learning algorithm. Weak learning is a classifier that cannot be compared to real-world classification. On the contrary, strong learning is a classifier that is meaningful for real-world classification (24).

B4. Adaptive Boosting algorithm

This algorithm, also called AdaBoost, combines weak learning algorithms to form a strong classifier. It is sensitive to noisy and outlier data. AdaBoost improves areas that basic learners cannot handle (26).

B5. Gradient Boosted Trees algorithm

This algorithm abbreviated GBT and is a set of classification and regression models. Both methods are group-based methods that achieve predictive results from improved gradual estimations (26).

B6. Bayesian Boosting algorithm

The Bayesian algorithm is one kind of Boosting model which is based on the Bayes theorem. This operator performs a meta-algorithm that can be used for numerous learning algorithms to improve their performance (26).

B7. Majority Voting

The Bayesian algorithm is one kind of Boosting model which is based on the Bayes theorem. This operator performs a meta-algorithm that can be used for numerous learning algorithms to improve their performance (26).

B8. Stacking algorithm

This algorithm uses a hyper-learning method employing two or more basic machine learning algorithms to learn how to combine predictions. Base level models are trained based on a complete training dataset, and then the meta-model in basic level models' outputs are trained as the features. The base-level includes several learning algorithms, so aggregating groups are heterogeneous (28).

2-4. Review the proposed method

This study proposed a combination of Random Forest and basic algorithms such as Naïve Bayes, Decision Tree, K-Nearest-Neighbors, and SVM in the stacking algorithm. COVID-19 patients' dataset is divided into training and test dataset utilizing 10-Fold Cross-Validation (29). We then evaluated the accuracy, precision, recall, and F-measure using the model output, a 2*2 vector called a sparse matrix. We experimented with different combinations of basic and hybrid algorithms (**Figure.4**).

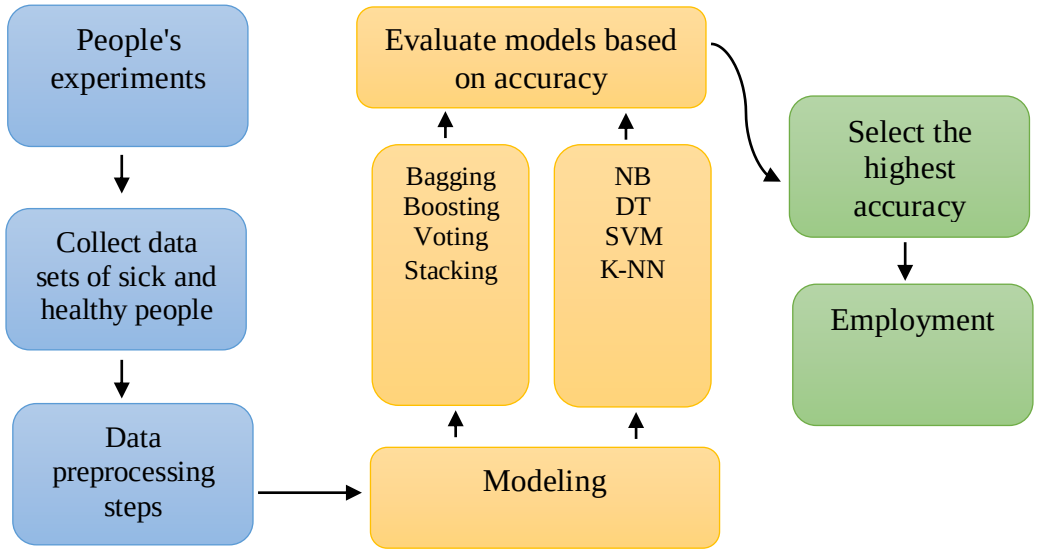


Fig.4: The model proposed by researchers.

www.SID.ir

Archive of SID

2-5. Proposed method

In this model, we combined Random Forest in the stacking part and basic algorithms such as Naïve Bayes, DT, K-NN, and SVM in the stacking algorithm's fundamental part. In the **Figures 5, 6**, we showed the model and the evaluated results.

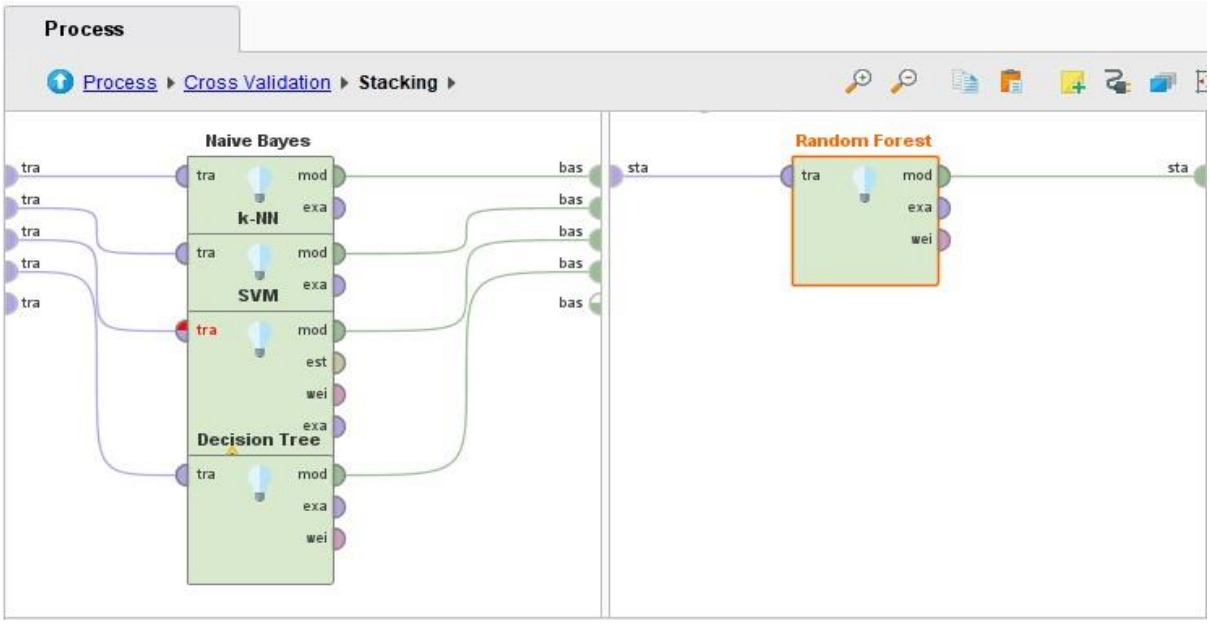


Fig.5: Basic algorithms combination in stacking classifier.

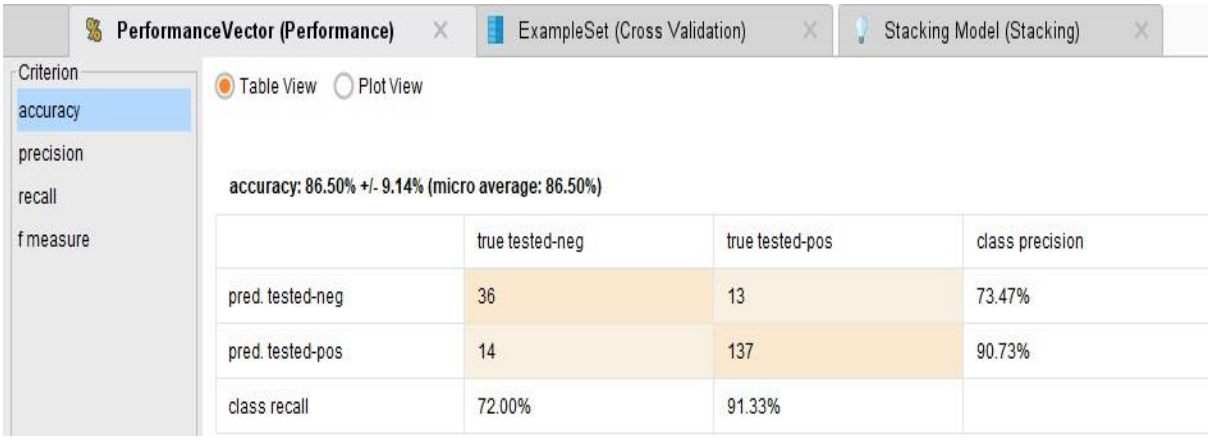


Fig.6. Proposed model's sparse matrix.

Computation and evaluation of the sparse matrix in the combinational stacking classifier model (the fifth method) is shown in **Table.1**.

Table-1: The results and evaluation of the combinational model in stacking classifier.

Accuracy	Precision	Recall	F-measure
86.5%	90.60%	91.31%	90.51%

www.SID.ir

Archive of SID

3- RESULTS

4-1. Evaluate and analyze the results

There is a wide range of classification algorithms, each of which has its strengths and weaknesses. In fact, none of the learning algorithms has the best performance, considering the current supervised learning issues. We evaluated the model to find the optimal solution from various classification models generated through a complicated and repetitive process. Machine learning model evaluation can be intricate. Generally, we divide the dataset into two categories: training dataset and test dataset. The training dataset is used to train the model, and the test dataset is employed to test the model. Then we evaluate the model performance based on error criteria to determine the model accuracy.

If we separate the training dataset to k folds with the same amount of data, in each phase of the cross-validation process, k-1 fold of these k folds will be the training dataset, and one of them will be the validation dataset (in this paper, we considered k = 10). The sparse matrix is hired to measure benchmarks like accuracy, precision, recall, and F-measure. A sparse matrix is a method in machine learning classification to measure the performance. This is kind of table contributes to specifying the model performance in the test dataset and enhances the real values representation. In the combinational classification model for COVID-19 diagnosis, data were divided into confirmed and negative. The relations between real and predicted classes are shown in the following matrix (**Figure.7**).

Where,

TN: the anticipated values are predicted correctly as a negative case.

TP: the anticipated values are predicted correctly as a positive case.

FP: the anticipated values are predicted incorrectly as a positive case.

FN: the anticipated values are predicted incorrectly as a negative case (31).

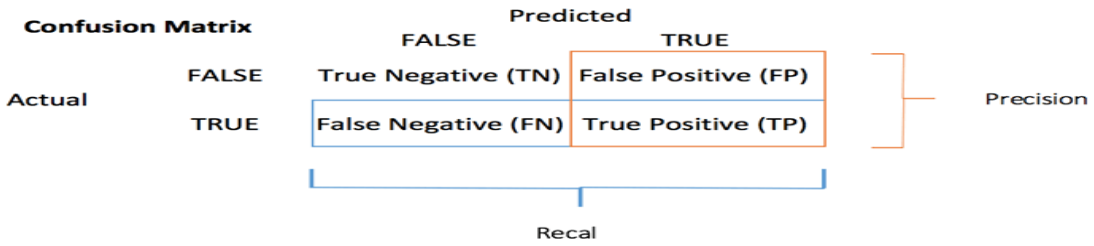


Fig.7: Sparse matrix (30).

In the following, we elaborated on some key benchmarks that can be calculated by the sparse matrix.

4-1-1. Accuracy benchmark

This is the ratio of correct prediction to the total input samples. For example, assume 98% of the training datasets are 'class A' samples, and 2% of data are 'class B' samples. Our model can easily achieve 98% accuracy by predicting all class A examples correctly. We can compute accuracy as follows:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}$$

4-1-2. Precision benchmark

Precision is the ratio of correctly predicted positive cases to all the positive predicted cases:

www.SID.ir

Archive of SID

$$Precision=\frac{TP}{TP+FP}$$

4-1-3. Recall benchmark

Recall is the ratio of correctly predicted positive cases to all the predicted observations:

$$Recall=\frac{TP}{TP+FN}$$

4-1-4. F-measure benchmark

F-measure computes the weighted average between precision and recall. Therefore, this benchmark considers both incorrect positive and negative cases. Intuitively, it is complicated to understand, but this is more useful than accuracy, especially in non-uniform distributions (31).

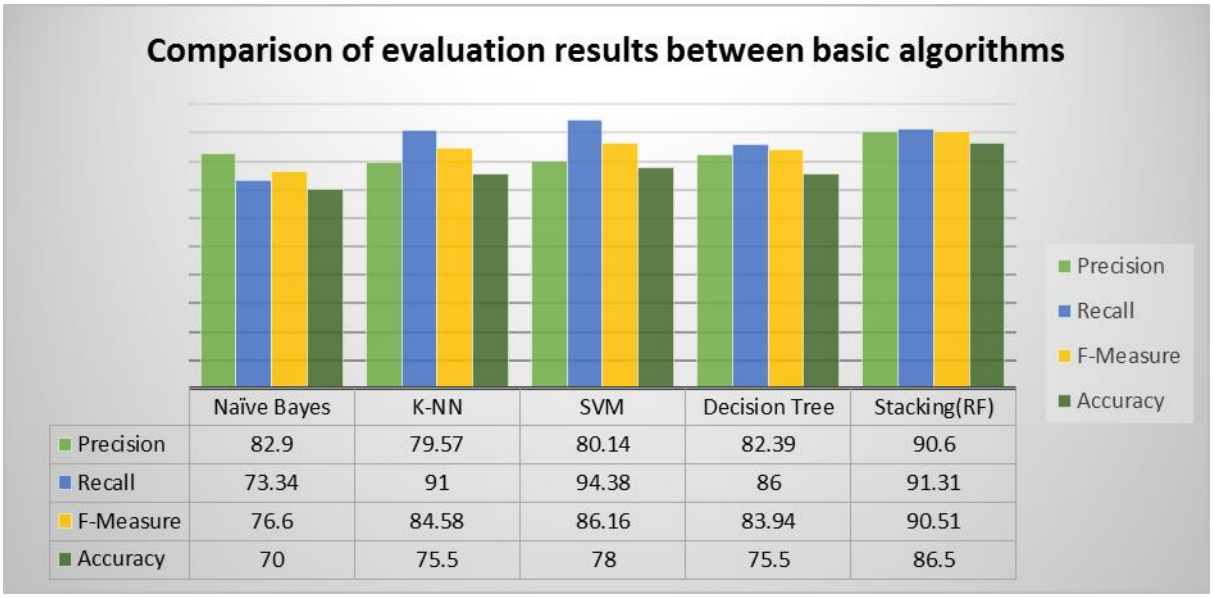
$$F\text{-}measure=2*\frac{Precision*Recall}{Precision+Recall}$$

4-2. Comparing the results of basic models evaluations

Various classification algorithms provide different results for different datasets or problems. A classification algorithm that presents the best solution for a given problem may not work on other problems or dataset. Thus, the results of various classification algorithms should be compared before finding a solution to the problem. The results of other basic algorithms with the proposed algorithm are shown in **Table.2**. The results of integrating basic algorithms with the proposed model are shown in **Figure.8**.

Table-2: Comparing basic classification benchmarks.

Classifier Model	F-Measure	Recall	Precision	Accuracy
Naïve Bayes	76.6%	73.34%	82.9%	**70%**
K-NN	84.58%	91%	79.57%	**75.5%**
SVM	86.16%	94.38%	80.14%	**78%**
Decision Tree	83.94%	86%	82.39%	**75.5%**
Stacking (RF)	90.51%	91.31%	90.60%	**86.5%**



	Naïve Bayes	K-NN	SVM	Decision Tree	Stacking(RF)
Precision	82.9	79.57	80.14	82.39	90.6
Recall	73.34	91	94.38	86	91.31
F-Measure	76.6	84.58	86.16	83.94	90.51
Accuracy	70	75.5	78	75.5	86.5

Fig.8: Evaluating and comparing the results of basic algorithms with the proposed model.

www.SID.ir

Archive of SID

4-3. Comparing the results of hybrid models evaluations

Group-based (Hybrid) methods provide a more accurate solution rather than a single model. Nevertheless, the effectiveness of this method is undeniable, and they benefit relevant applications remarkably. In some areas like healthcare, even a negligible accuracy improvement in machine learning algorithms is valuable. The results of other hybrid algorithms with the proposed algorithm are shown in **Table 3**.

Table-3: Comparing hybrid classification benchmarks.

Combined classifier model	F-Measure	Recall	Precision	Accuracy
Bagging	90.74%	94.90%	87.51%	**86%**
Random Forest	89%	94.55%	84.73%	**83.5%**
AdaBoost	89.51%	94.66%	85.18%	**84%**
GBT	89.26%	92.67%	87.6%	**83.5%**
Bayesian Boosting	87.86%	93.47%	83.72%	**81.5%**
Voting	89.81%	97.63%	83.5%	**84.5%**
Stacking (GBT)	90.12%	90.65%	90.54%	**86%**
Stacking (RF)	90.51%	91.31%	90.60%	**86.5%**

According to the **Figures 8 and 9**, the stacking algorithm, which consists of two parts, obtained the highest accuracy by trial and error method. In the first part, all basic models for predicting the output of test data sets and the second part-contain a meta-classifier, which considers all the predictions of the base models as input and creates a new prediction with a random forest algorithm, which led to more accurate results.

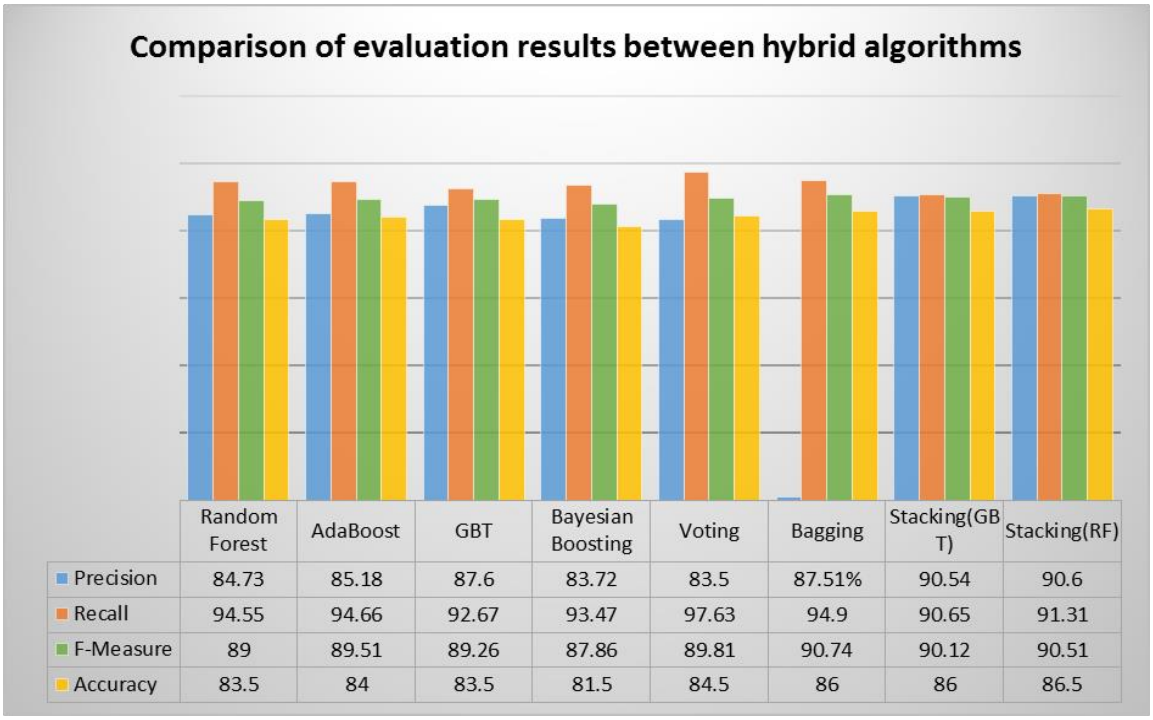


Fig.9: Evaluating and comparing the results of hybrid algorithms.

www.SID.ir

Archive of SID

4- DISCUSSION

In today's world of information, it is required to find an analytical and robust solution to extract valuable information from the data-intensive aggregated and saved information in organization databases or repositories. Researchers currently use models based on data mining algorithms to predict and diagnose different kinds of diseases. We used a group-based learning method, a hybrid learning system that provides multiple analyses to achieve more accurate results than a single model. In December 2019, in Wuhan, China, a new coronavirus named severe acute respiratory syndrome, coronavirus 2 (SARS-CoV-2) produced a new disease (32). Since mid-February, more than 50,000 cases of COVI-19 have been confirmed in China, and more than 1,600 of them have died. Shortly after, COVID-19 spread throughout China to other countries, and it has been reported in more than 25 countries (33, 34). Today, the coronavirus is the main hygiene crisis in the world and has an effect on people at an international level and has shifted to a global epidemic. In recent months, corona virus has been responsible for the considerable increase in death rate (35, 36). In late January 2020, WHO became concerned about the coronavirus and investigated it as hygiene emergency status (37). The first case in Iran was approved in Qom in February 2020 (38). Based on the Iran Ministry of Health and Medical Education, as of December 14, 2020, the number of patients with Covid-19 in Iran was 1,108,269, the number of deaths was 52,196 and the number of patients recovered was 812,270 (**Figure.10**), (39).

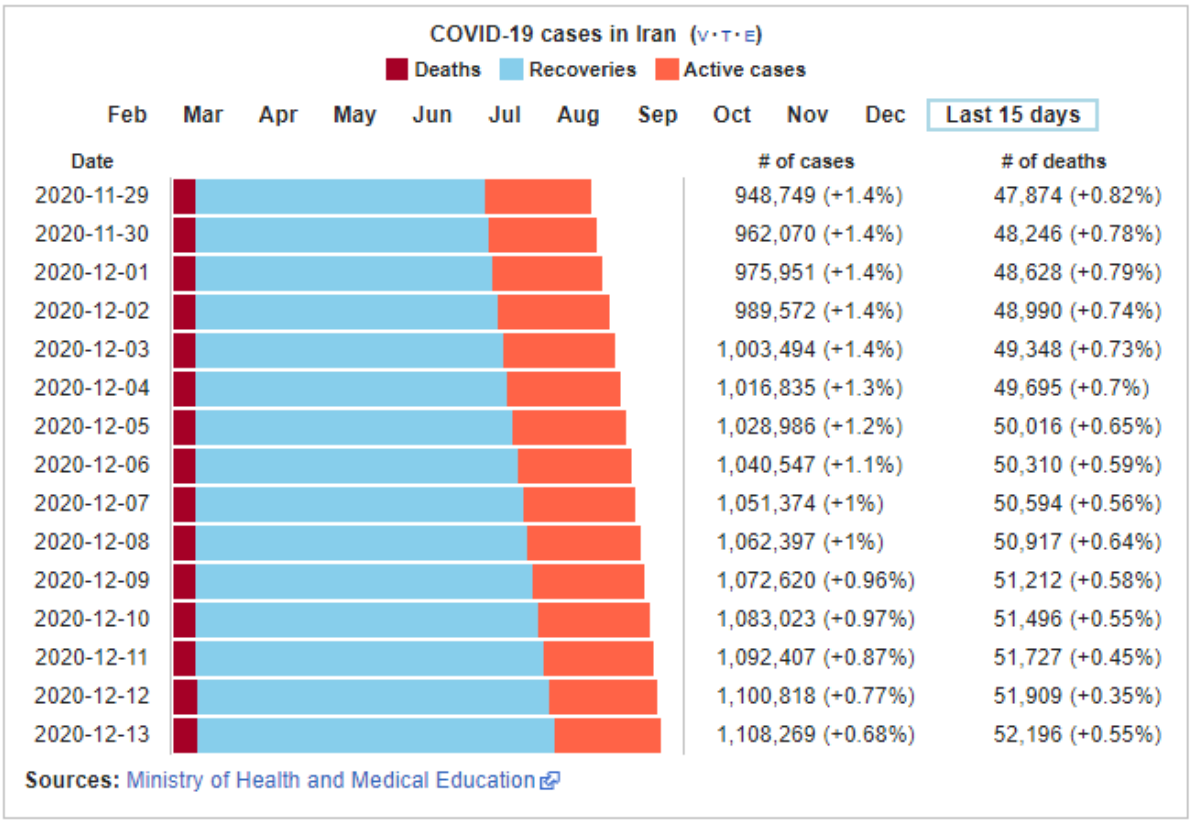


Fig.10: Covid-19 in Iran (39).

www.SID.ir

Archive of SID

COVID-19 is a respiratory disease that belongs to a large family of viruses called Corona. This disease is a subset of respiratory and lung diseases that have killed many people worldwide. The number of deaths contribute to understanding the severity of a disease, distinguishing detecting people at higher risk of getting that disease, and evaluating the quality of healthcare. The development of vaccines and specific drug therapies are under investigation and are undergoing clinical trials (40-42). It is evident that nowadays, researchers worldwide are seeking to find non-clinical methods such as data mining, machine learning, and deep learning approaches. Everybody has felt the need to develop a rapid, accurate, and affordable solution compared to the diagnosis methods in hospitals. This will lead to preventing the spread of the COVID-19 virus.

5- CONCLUSION

In this study, we used Relief feature selection to prepare and preprocess datasets. We combine Random Forest and basic algorithms such as Naïve Bayes, DT, SVM, and K-NN in stacking classifier. The results have shown that the stacking algorithm diagnoses the patients with 85.6% accuracy. We can improve the models' performance by increasing data in the standard format for the COVID-19 dataset, and improve the model accuracy trying other technological techniques like a neural network, Genetic Algorithm, and Logistics Regression.

6- ACNOWLEDGMENT

This paper is derived from the MS dissertation of Mohammad Saeedi. The authors would like to acknowledge the University Research Management, for assisting this project at Imam Reza International University, Mashhad, Iran (ID-number: 552379).

7- CONFLICT OF INTEREST: None.

8- REFERENCES

1. Bousquet J, Dahl R, Khaltaev N. Global alliance against chronic respiratory diseases. European Respiratory Journal. 2007;29(2):233-9.

2. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. Journal of Advanced Research. 2020 Mar 16. https://doi.org/10.1016/j.jare.2020.03.005.

3. Chen M, Hao Y, Hwang K, Wang L, Wang L. Disease prediction by machine learning over big data from healthcare communities. Ieee Access. 2017; 5: 8869-79.

4. Fayyad U, Stolorz P. Data mining and KDD: Promise and challenges. Future generation computer systems. 1997; 13(2-3): 99-115.

5. Eom JH, Kim SC, Zhang BT. AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction. Expert Systems with Applications. 2008; 34(4): 2465-79.

6. Chaurasia V, Pal S. Early prediction of heart diseases using data mining techniques. Caribbean Journal of Science and Technology. 2013; 1: 208-17.

7. S. Bashir, U. Qamar and M. Younus Javed, "An ensemble based decision support framework for intelligent heart disease diagnosis," International Conference on Information Society (i-Society 2014), London, 2014, pp. 259-264, doi: 10.1109/i-Society.2014.7009056.

8. H. I. Elshazly, M. Waly, A. M. Elkorany and A. E. Hassanien, "Chronic eye disease diagnosis using ensemble-based classifier," 2014 International Conference on Engineering and Technology *(ICET)*, Cairo, 2014, pp. 1-6, doi: 10.1109/ICEngTechnol.2014.7016799.

9. Kurdia H, AlMansour N. Identifying accurate classifier models for a text-based MERS-CoV dataset. In 2017 Intelligent

www.SID.ir

Archive of SID

Systems Conference (IntelliSys) 2017 Sep 7 (pp. 430-435). IEEE.

10. Arafat F, Fatema K, Islam S. Classification of chronic kidney disease (ckd) using data mining techniques, 2018.

11. Kurian RA, Lakshmi KS. An ensemble classifier for the prediction of heart disease. International Journal of Scientific Research in Computer Science. 2018; 3(6):25-31.

12. AlMoammar A., AlHenaki L., Kurdi H. (2019) Selecting Accurate Classifier Models for a MERS-CoV Dataset. In: Arai K., Kapoor S., Bhatia R. (eds) Intelligent Systems and Applications. IntelliSys 2018. Advances in Intelligent Systems and Computing, vol 868. Springer, Cham. https://doi.org/10.1007/978-3-030-01054-6_74

13. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, Cai M, Yang J, Li Y, Meng X, Xu B. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). MedRxiv. 2020 Jan 1.

14. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv preprint arXiv:2003.10849. 2020 Mar 24.

15. Barstugan M, Ozkaya U, Ozturk S. Coronavirus (covid-19) classification using ct images by machine learning methods. arXiv preprint arXiv:2003.09424. 2020 Mar 20.

16. Wiguna W, Riana D. Diagnosis of Coronavirus disease 2019 (Covid-19) surveillance using C4. 5 Algorithm. Jurnal Pilar Nusa Mandiri. 2020; 16(1):71-80.

17. Muhammad LJ, Islam MM, Sharif US, Ayon SI. Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients Recovery. SN COMPUT SCI. 2020; 1(4): 206.

18. Khanday, A.M.U.D., Rabani, S.T., Khan, Q.R. et al. Machine learning based approaches for detecting COVID-19 using clinical text data. Int. j. inf. tecnol. 12, 731–739 (2020). https://doi.org/10.1007/s41870-020-00495-9

19. García S, Luengo J, Herrera F. Data preprocessing in data mining. Cham, Switzerland: Springer International Publishing; 2015. ISBN 978-3-319-10247-4.

20. Dash M, Liu H. Feature selection for classification. Intelligent data analysis. 1997; 1(3):131-56.

21. Sahu B, Dehuri S, Jagadev A. A Study on the Relevance of Feature Selection Methods in Microarray Data. The Open Bioinformatics Journal. 2018;11(1):117-39.

22. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH. Relief-based feature selection: Introduction and review. Journal of biomedical informatics. 2018; 85: 189-203.

23. Nikam SS. A comparative study of classification techniques in data mining algorithms. Oriental journal of computer science & technology. 2015; 8(1):13-9.

24. N. C. Oza, "Online bagging and boosting," 2005 IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI, 2005, pp. 2340-2345 Vol. 3, doi: 10.1109/ICSMC.2005.1571498.

25. Alfaro E, Gámez M, García N. Ensemble Classifiers Methods. Ensemble Classification Methods with Applicationsin R. 2018 Oct 29:31-50.

26. Drucker H, Cortes C, Jackel LD, LeCun Y, Vapnik V. Boosting and other ensemble methods. Neural Computation. 1994; 6(6):1289-301.

27. Ruta D, Gabrys B. Classifier selection for majority voting. Information fusion. 2005; 6(1):63-81.

28. Frías-Blanco I, Verdecia-Cabrera A, Ortiz-Díaz A, Carvalho A. Fast adaptive stacking of ensembles. InProceedings of the 31st Annual ACM Symposium on Applied Computing 2016 Apr 4 (pp. 929-934). https://doi.org/10.1145/2851613.2851655.

29. Wong, T.-T. "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation." Pattern Recognition.2015; 48(9): 2839-46.

30. Visa S, Ramsay B, Ralescu AL, Van Der Knaap E. Confusion Matrix-based Feature Selection. MAICS. 2011; 710: 120-7.

31. Lever J, Krzywinski M, Altman N. Classification evaluation. Available at:

www.SID.ir

Archive of SID

https://www.nature.com/articles/nmeth.3945?report=reader.

32. World Health Organization Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV).

33. Zhang, Yi et al. "A Novel Coronavirus (COVID-19) Outbreak. Chest. 2020 Apr; 157(4): e99–e101.

34. Kumar Kar, S., Yasir Arafat, S.M., Kabir, R., Sharma, P., & Saxena, Sh. (2020). Coronavirus Disease 2019 (COVID-19), Coping with Mental Health Challenges dDuring COVID-19, 199-213. doi: 10.1007/978-981-15-4814-7_16.

35. Yu, H., Li, M., Li, Zh., Xiang, We., Yuan, Y., Liu, Y., Li, Z., & Xiong, Zh. Coping style, social support and psychological distress in the general Chinese population in the early stages of the COVID-2019 epidemic. 10.21203/rs.3.rs-20397/v2 (preprint), 2020.

36. Hosseinzadeh Shirayeh B, Shojaie L, Maharatfard Jahromy M, Vaziri MS, Safavi S, Jafari R, et al. Comparison of Clinical Characteristics of Hospitalized Patients with and Without Covid-19 in Mashhad, Iran: A Retrospective, Single-Center Experience. Int J Pediatr 2020; 8(12): 12619-628.

37. Coronavirus. Worldometer. Archived from the original on 9 March 2020. Retrieved 23 March 2020.

38. Ministry of Health and Medical Education. 2020;14,12.

39. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus(2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. Euro Surveill. 2020; 25 doi: 10.2807/1560-7917.ES.2020.25.5.2000062.

40. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. (February 2020). "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study". Lancet. 395 (10223):507-13.

41. Symptoms of Coronavirus. U.S. Centers for Disease Control and Prevention (CDC). 13 May 2020. Archived from the original on 17 June 2020. Retrieved 18 June 2020.

42. Han X, Cao Y, Jiang N, Chen Y, Alwalid O, Zhang X, et al. "Novel Coronavirus Pneumonia (COVID-19) Progression Course in 17 Discharged Patients: Comparison of Clinical and Thin-Section CT Features during Recovery". Clinical Infectious Diseases. 2020; 71 (15):723. doi:10.1093/cid/ciaa271

www.SID.ir