Develop application for prediction COVID-19 using artificial intelligence

Noor abdulmuttaleb jaafar[1], Noor Razzaq Abbas [2], Ammar Kadi [3], Abdelhameed Ibrahim[4], Abdelaziz A. Abdelhamid[5]

[1]Administration & Finance Department University of Diyala, University of Diyala, Baqubah MJJ2+R9G, Iraq
[2] Al-Furat Al-Awsat Technical University, Technical Institute of Najaf, Iraq
[3] Department of Food and Biotechnology, South Ural State University,
454080 Chelyabinsk, Russia
[4]Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, 35516, Mansoura, Egypt
[5]Computer Science Department, Faculty of Computer and Information Science, Ain Shams University, Cairo, 11566, Egypt
Emails noor84abd84@uodiyala.edu.iq; noor.hachame@atu.edu.iq; ammarka89@gmail.com; afai79@mans.edu.eg; abdelaziz@cis.asu.edu.eg

Abstract

The subset of manufactured insights (AI) known as machine learning starts in design acknowledgment, where information can be organized for human comprehension. For a long time, various applications utilizing machine learning have been created in healthcare, fund, military gear, and space investigation; presently, machine learning is a zone that's extending and progressing quickly. It utilizes information to optimize computer execution. AI is vital in combating modern coronaviruses in 2019 (COVID-19) -related matters and is used additionally in computer-assisted blend-making plans. Computer programs' settings are optimized based on preparing information or past encounters. It can moreover make future forecasts utilizing the information. With the assistance of machine learning, we are creating a numerical demonstration based on the data's measurements. Numerous illustrations outline the viability of machine learning and counterfeit insights in this field. Counterfeit insights strategies can improve the consistency of forecasts and choices by making valuable calculations. AI is useful not for foreseeing people with COVID-19 but for assessing general wellbeing. It can screen the COVID-19 episode at different levels; in our paper, we use three machine learning calculations to analyze and predict. The leading precision was in XGP= 99%, but SVM and RF gave great precision at 98%.

Keywords: Covid 19; SVM; RF; XGB; Machine learning; Internet of things; Prediction

1. Introduction

The subset of manufactured insights (AI) known as machine learning started from design acknowledgment, where information may be orchestrated for human comprehension. Later, a few applications utilizing machine learning in businesses such as healthcare, back, military equipment, circle investigation, and Machine have improved for a long time. Learning could. Be a quick developing both presently and continuously making strides region at the show time. It programs Computers to maximize their execution utilizing data. [3,4,7]. The settings of computer programs are optimized based on their planning to analyze information and previous encounters. Using the information can moreover make future expectations. Utilizing the insights of the information, we are building a numerical demonstration with the help of machine learning. The central point of the magnificence of machine learning is that it automatically learns from information inputs. It

Doi: https://doi.org/10.54216/JAIM.060103
Received: February 10, 2023 Revised: May 11, 2023 Accepted: November 17, 2023

consequently learns from given information (involvement) and gives the specified yield by looking at the data for patterns and designs.

The coronaviruses (Cove) are a tremendous family of infections that can cause various conditions in creatures and people. These ailments run seriously from the conventional cold to the more unsafe extreme intense respiratory syndrome (SARS). These diseases can cause side effects such as fever, hacking, dyspnoea, and trouble breathing. Passing can happen if somebody has an exceptionally awful pneumonia or kidney disappointment. On December 31, 2019, China declared the presence of a cluster of pneumonia cases with an obscure root. They recognized the flare-up as SARS-C0v-2, a modern strain of coronavirus, was distinguished on January 7. At the minute, this disease may be widespread (assigned by the WHO on Walk 11, 2020), influencing a tremendous parcel of the world's populace. Fever, a dry hack, and weariness are the foremost signs of this condition. There are, moreover, less common side effects like nasal blockage and cerebral pains, runs, a misfortune of taste or scent, and skin rashes, but these are still conceivable [1,10,11,14].

Around 80% of patients with this disorder recoup without hospitalization. 15% require inpatient treatment, and 5% require basic treatment (serious care). Individuals over the age of 60 or those who have pre-existing infections such as high blood weight, heart or lung issues, diabetes, or cancer are especially at hazard of creating an extreme condition. In any case, many young individuals with no well-being issues recently got COVID-19, getting exceptionally wiped out after they got it. A few people who have beaten this condition have symptoms for an amplified period. Parece There's inadequate logical proof at the minute to gather that COVID-19 can cause long-term seculars. Still, there's insufficient logical proof that COVID-19 can cause long-term seculars because of the virus's brief nearness in our middle. Covid 19 is mostly spread by contact with a tainted person. The foremost visit mode of transmission is by diminutive beads that can be ousted from a contaminated person's nose or lips when they hack, talk, or wheeze. Moreover, the disease can happen if a contaminated person places their hands over their mouth, nose, or eyes after touching a surface on which beads have landed. This could be anticipated by cleansing hands with warm water and cleanser or using a hand rub alcohol-based hand sanitizer. [2,3,8]

Besides, pressurized canned products can be utilized to scatter COVID-19. Mist concentrates are small solid or fluid particles coasting within the discussion. When an infected individual hacks, sniffles, or talks boisterously, they can make mist concentrates capable of infecting other solid individuals despite their modest measures. This is often particularly genuine inside, where discussion circulation is constrained. One statistic metric recommended is social separation between people to avoid the spread of these beads and pressurized canned products. As a precaution, it is prescribed that the open-wear covers assist in preventing infection transmission as they channel out possibly infectious tiny particles [22,6,7].

Indeed, even though a few patients with COVID-19 have exceptionally mellow side effects, they can still be exchanged by beads made by talking or hacking. There's a scaling issue here. The virus, as numerous people stay undiscovered and may unknowingly spread the infection; according to the World Wellbeing Organization, People who have been in contact with a positive person must self-isolate for 14 days. Quarantine is the method through which unhealthy people are physic [2,12,15,16]. They are entirely separated from solid ones. People who test positive for any infection should be isolated for at least ten days, with three additional days of symptom-free isolation added. If there are no side effects, the individual should be kept absent from other individuals for ten days after the positive test. Indications develop five to six days after infection contact is regular; in any case, this number can run between one and fourteen days [4,8,9,10].

2. Dataset (.Csv)

The.csv record is frequently used to transport and store information in a database. Acing the capacity to peruse, change, and compose reports to and from CSV records utilizing Python is imperative for any information researcher or commerce investigator whose records incorporate comparative but particular data strings. Each column within the table speaks to a database measurement, while each push speaks to a partitioned history when the data communicated in a CSV record is ordinarily alphanumeric [5,6,17,18].

One of the foremost common occurrences of inactive records could be a comma-separated values (CSV) record. A comma isolates each cell's esteem, and each push starts an unused line in a CSV record. Tabs or commas frequently separate columns and bars, as the inactive record database comprises a single table. CSV records are ordinarily produced by applications that oversee colossal amounts of information. They are a helpful strategy to trade and consequence information from spreadsheets and databases into other programs. For occasion, you'll send out the discoveries of an

Doi: https://doi.org/10.54216/JAIM.060103
Received: February 10, 2023 Revised: May 11, 2023 Accepted: November 17, 2023

information mining instrument to a CSV record and input it into a spreadsheet to assess the information, make charts for an introduction, or prepare a report for distribution. CSV records are fundamental to control programmatically. Any programming dialect, such as Python, that permits content record input and string control may work specifically with CSV records [12,14, 20].
The CSV library offers both examined and type-in capabilities for CSV records. Planned to operate out of the box with CSV records made by Exceed expectations, it may be promptly changed to work with several CSV designs. The CSV library incorporates objects and extra code for perusing, composing, and preparing information from and to CSV records. Utilizing CSV Records Utilizing CSV, the peruse question is used to peruse a report from a CSV record. Python's built-in open () strategy opens the CSV record as a content record and returns a record question. Typically, in this way, it is passed on to the peruse, who performs the challenging assignments [13,21,19].
The dataset is multidimensional, combined information. Some information demonstrates whether a persistent has already been analyzed with a particular illness, such as kidney illness or gastrointestinal disorder, whereas others contain exact clinical values already. It includes areas containing content information and a few containing correct values. Content information is encoded with substantial values for the test. The properties considered within the dataset for the machine learning show are displayed in Figure 1.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Breathing Problem	Fever	Dry Cough	Sore throat	Running Nose	Asthma	Chronic Lung Disease	Headache	Heart Disease	Diabetes	Hyper Tension	Fatigue	Gastrointestinal	Abroad travel
2	Yes	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes	No
3	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	No	Yes	No	No
4	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes
5	Yes	Yes	Yes	No	No	Yes	No	No	Yes	Yes	No	No	No	Yes
6	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No
7	Yes	Yes	Yes	No	No	No	No	No	Yes	No	Yes	No	No	No
8	Yes	Yes	Yes	No	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No
9	Yes	Yes	Yes	No	Yes	Yes	No	No	No	Yes	Yes	No	Yes	Yes
10	Yes	Yes	Yes	No	Yes	No	Yes	No	No	Yes	No	Yes	No	Yes
11	Yes	Yes	Yes	No	No	Yes	No	No	No	Yes	Yes	Yes	No	No
12	Yes	Yes	Yes	No	No	No	Yes	No	Yes	No	Yes	No	No	Yes
13	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes
14	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No	No	Yes	No	Yes	Yes
15	Yes	Yes	Yes	Yes	No	Yes	No	No	No	Yes	Yes	No	No	Yes
16	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	No
17	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	No	No	No	No	No
18	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	No	No	No	Yes	No
19	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
20	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	Yes	No	No

Figure 1: Patient dataset

3. Architecture of the application

Python was utilized to code the COVID-19 dataset. The information set was transmitted already. This information collection contains data concerning affirmed cases, side effects, and recouped patients. To figure the guess based on the injury's side effects. Python was utilized to create a forecast show for anticipating affirmed and recouped cases. Based on obtained information and scientific models, expectations are shaped, and a demonstration is built. The reply "YES" is given. "NO for states and side effects as input to decide the connection between both factors." Figure 5. Show outline schematic Set of COVID-19 case records; in expansion, information on contamination indications was included to the members within the data set.

The information is changed and spared as CSV records for ensuing handling, preparing, and testing information: The demonstration was tried and prepared utilizing side effect data and numerical models. The objective of the preparing demonstration is to optimize the show by using the prepared information and organizing the rally with expectation calculations. After the demonstration is ready, the execution of the predicate models is assessed using testing data sets. Within the approval stage, the rightness of the results is considered, and the test information is compared to the forecasts. As seen in Figure 2.

Doi: https://doi.org/10.54216/JAIM.060103
Received: February 10, 2023 Revised: May 11, 2023 Accepted: November 17, 2023

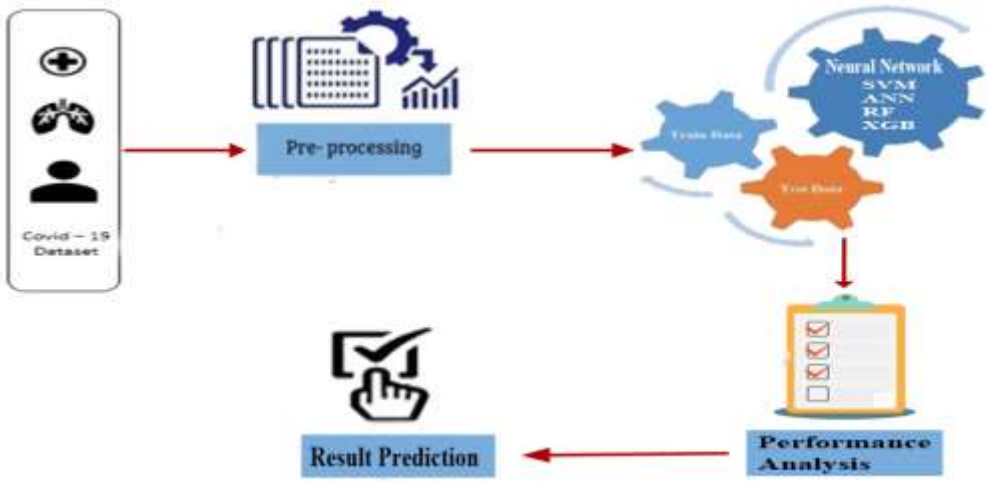

Figure 2: Overview of the model Architecture of the application

4. Performance Metrics

Measuring the execution of a machine learning show may be a vital errand. As our demonstration requires categorization, we chose exactness as the execution metric. In this thesis, exactness is the metric used to assess the calculations. It is the foremost standard measurement utilized to evaluate classification algorithms. This metric makes a difference in determining which demonstrate is best at recognizing designs within the preparing set in order to create more exact expectations within the test set.

$$accuracy = \frac{TP+TN}{TP+TN+FN} \qquad (1)$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False.

Table 1: shows the accuracy of each algorithm

		precision	recall	f1-score	support	accuracy
Accuracy of SVM	0	0.93	0.97	0.95	216	0.98
	1	0.99	0.98	0.99	871	
Accuracy of RF	0	0.93	0.97	0.95	216	0.98
	1	0.99	0.98	0.99	871	
Accuracy of XGB	0	0.95	0.98	0.96	218	0.99
	1	0.99	0.99	0.99	869	

3. Result

A. Support Vector Machine (SVM)

A Bolster Vector Machine (SVM) calculation is prepared with each record set to recognize its exactness at all stages. At all locations, the information was separated into preparing and test information. SVM accomplishes a precision of 98 %. Figure 23 speaks to the exactness of each set of records achieved by the Bolster Vector Machine (SVM) calculation.
Appears the perplexity framework of approval and test information for two-class classification with COVID-19/normal. Fig. 24. shows the preparation, approval precision, and misfortune for the VGG16 show. The demonstration merges well since the crevice between the preparing and approval bends may be the least. They are accomplishing up to 98% exactness, 99% review, 95% exactness, and 99 % F1-score for our show. The execution of the Specific show can be analyzed by implying a network of perplexity; the numbers of cases allude to genuine positives, genuine negatives, wrong positives, and untrue negatives. Figure 3 shows the perplexity frameworks for when information expansion was connected and when it was not connected. The following data can be extracted from the matrix:
True Positives =186, False Positives = 14

Doi: https://doi.org/10.54216/JAIM.060103
Received: February 10, 2023 Revised: May 11, 2023 Accepted: November 17, 2023

True Negatives = 875, False Negatives = 12

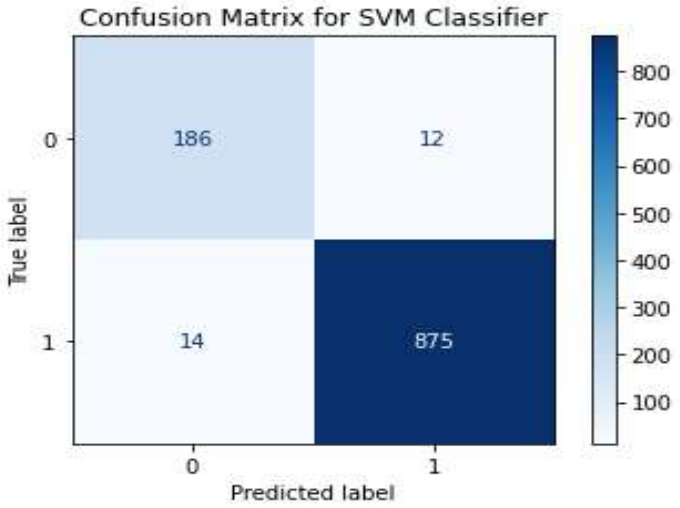

Figure 3: Confusion for SVM Classifier

B. Random Forest (RF)

Arbitrary Timberland (RF) calculation is prepared in a comparable way with each bunch of information to decide its rightness at all stages. At all locations, the news was separated into designing and testing knowledge. RF accomplishes an exactness of 99.00%. The classification exactness of Arbitrary Woodland (RF) calculation for each set of records spoken to in the execution of the Specific show can be analyzed by implying a network of disarray; it appears the numbers of cases alluding to genuine positives, genuine negatives, wrong positives, and untrue negatives. Figure 4 shows the perplexity frameworks for when connected information increases and is not connected individually.
The following data can be extracted from the matrix:
True Positives = 200, False Positives = 22
True Negatives = 859, False Negatives = 6

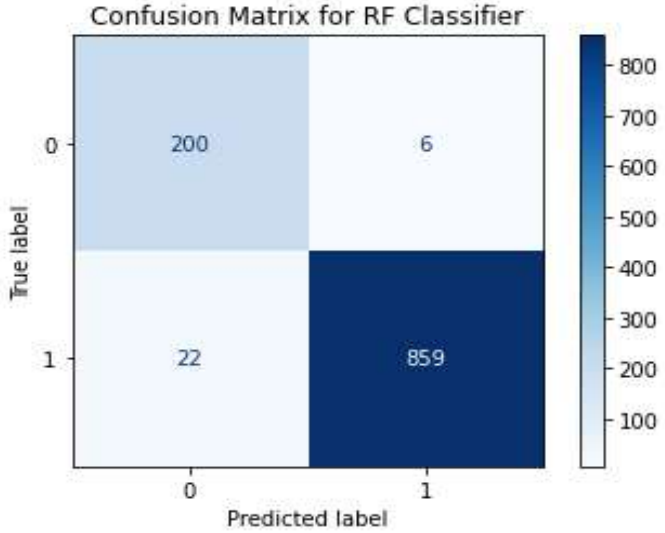

Figure 4: Confusion for RF Classifier

C. XGBoost (XGB)

The XGBoost calculation is prepared with each record set in a comparable design to decide its rightness at all stages with each set in an identical Mold. At all locations, the information was separated into preparing and test information. The precision of XGB is 98.00 percent. The classification exactness of the XGB calculation for each data set. The execution of the specific show

Doi: https://doi.org/10.54216/JAIM.060103
Received: February 10, 2023 Revised: May 11, 2023 Accepted: November 17, 2023

can be analyzed by implying a framework of perplexity. The number of cases allude to genuine positives, genuine negatives, untrue positives, and wrong negatives. Figure 26 shows the disarray frameworks for when connected information enlargement was hooked and when it was not individually.
The following data can be extracted from the matrix:
True Positives = 186 False Positives = 14
True Negatives = 875 False Negatives = 12

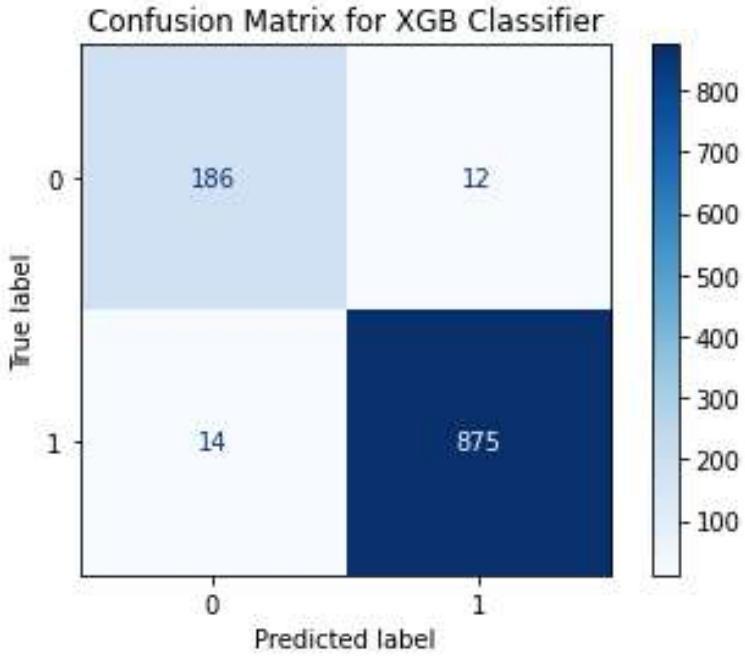

Figure 5: Confusion for XGBoost Classifier

D. Correlation matrix

Figure 6. The colors within the figure speak to a chi-square test of a few interconnected relationships. We note that blue refers to a coherent relationship's nonattendance, whereas white speaks to a tall relationship until zero esteem is reached. These numbers speak to the P-value.

Doi: https://doi.org/10.54216/JAIM.060103
Received: February 10, 2023 Revised: May 11, 2023 Accepted: November 17, 2023

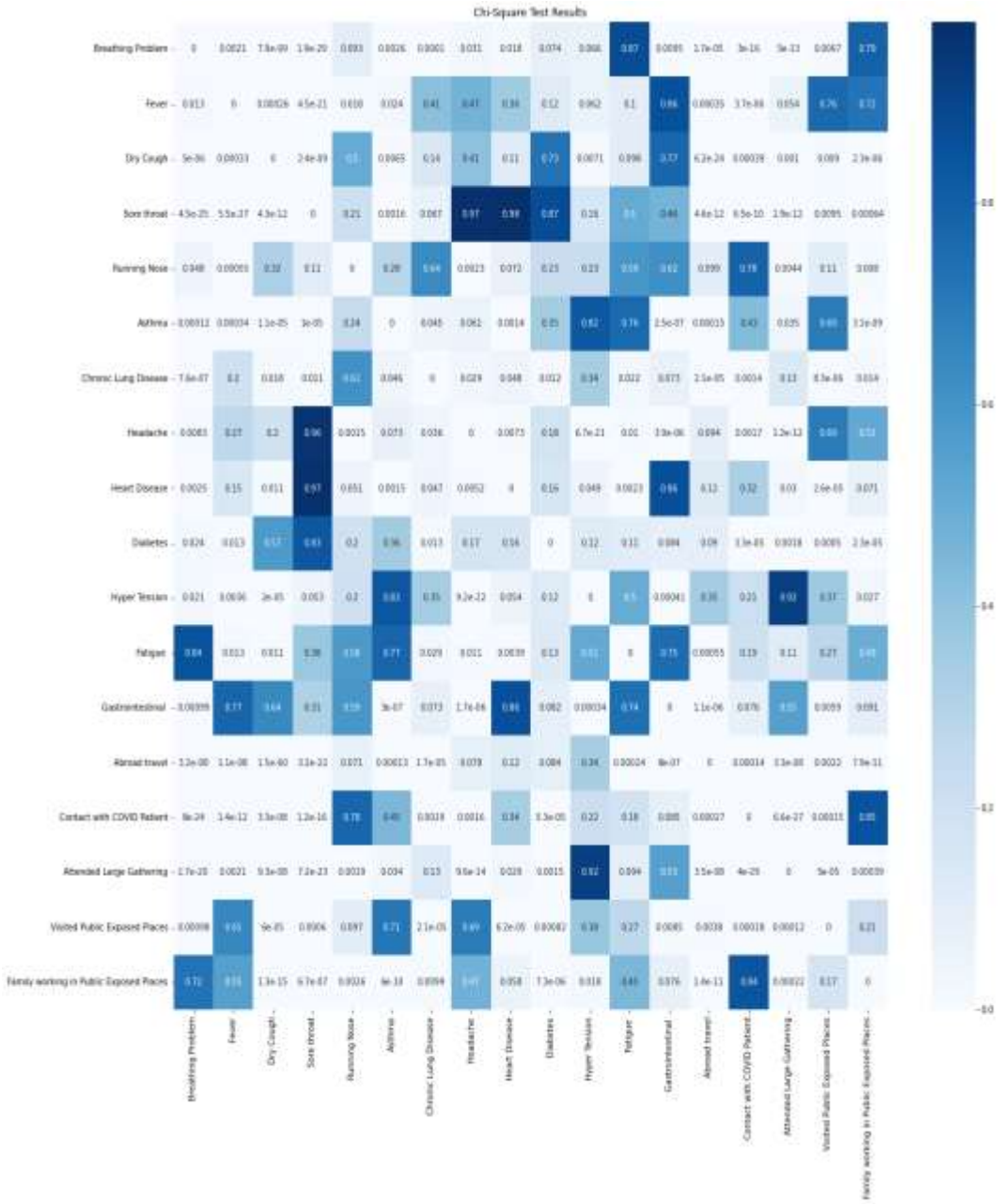

Figure 6: Correlation matrix

6. Conclusion

This thought distinguished the essential writing concerning expectation models for COVID-19. This study's objective was to survey and survey the distributions within the logical scholarly papers concerning COVID-19 expectation models. A prediction show may be a strategy for determining the more extended term utilizing existing information. An examination was conducted to decide which forecast models are right now accessible, with the targets of recognizing different strategies used to create different sorts of prediction models and achieving an effectiveness or quality appraisal of models, which helps in building up their precision.

Concurring to the discoveries of this expectation, we'll be able to require more noteworthy safety measures against the infection since of this figure. Within the most punctual stage of the disease, the symptomatic framework may be critical and offer assistance in deciding the probability of contamination. The examination of this information set is complicated by the reality that it is growing daily and the number of occasions is expanding quickly. A number of future endeavors will include sending the framework with extra machine learning models that give essentially higher exactness.

Doi: https://doi.org/10.54216/JAIM.060103
Received: February 10, 2023 Revised: May 11, 2023 Accepted: November 17, 2023

Funding: "This research received no external funding."

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Coronavirus Disease (COVID-19) - events as they happen. Library Catalog: www.who.int.
[2] Countries where Coronavirus has spread-Worldometer. Library Catalog: www.worldometers.info.
[3] COVID-19 situation reports. Library Catalog: www.who.int.
[4] Diagnosis of covid-19 and its clinical spectrum dataset. url=https://kaggle.com/einsteindata4u/covid19.
[5] WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020. Library Catalog: www.who.int.
[6] WHO EMRO | Questions and answers | COVID-19 | Health topics.
[7] Support Vector Machine Machine learning algorithm with example and code, January 2019. Library Catalog: www.codershood.info Section: Machine learning.
[8] Ali Al-Hazmi. Challenges presented by MERS corona virus, and SARS corona virus to global health. Saudi journal of biological sciences, 23(4):507–511, 2016. Publisher: Elsevier.
[9] Sina F Ardabili, Amir Mosavi, Pedram Ghamisi, Filip Ferdinand, Annamaria R Varkonyi-Koczy, Uwe Reuter, Timon Rabczuk, and Peter M Atkinson. Covid-19 outbreak prediction with machine learning. Available at SSRN 3580188, 2020.
[10] Ibrahim W, Abdullaev S, Alkattan H, Adelaja OA, Subhi AA. Development of a Model Using Data Mining Technique to Test, Predict and Obtain Knowledge from the Academics Results of Information Technology Students. Data. 2022; 7(5):67. https://doi.org/10.3390/data7050067
[11] Taiwo Oladipupo Ayodele. Types of machine learning algorithms. New advances in machine learning, 19–48, 2010.
[12] Doaa S. Khafaga, Hussein Alkattan, Alhumaima A. Subhi. (2022). Evaluating the Effect of Optimized Voting Using Hybrid Particle Swarm and Grey Wolf Algorithm on the Classification of the Zoo Dataset. Journal of Artificial Intelligence and Metaheuristics, 2 (1), 08-15.
[13] Hussein Alkattan, Sanjar Abdullaev, El-Sayed M. El-Kenawy. (2023). The «Climate in Weathers» Approach to Processing of Meteorological Series in Mesopotamia: Assessment of Climate Similarity and Climate Change using Data Mining. Journal of Intelligent Systems and Internet of Things, 10 (1), 48-65.
[14] Hetal Bhavsar and Amit Ganatra. A comparative study of training algorithms for supervised machine learning. International Journal of Soft Computing and Engineering (IJSCE), 2(4):2231–2307, 2012.
[15] Hussein Alkattan, S. K. Towfek, M. Y. Shams. (2023). Tapping into Knowledge: Ontological Data Mining Approach for Detecting Cardiovascular Disease Risk Causes Among Diabetes Patients. Journal of Artificial Intelligence and Metaheuristics, 4 (1), 08-15.
[16] Nanshan Chen, Min Zhou, Xuan Dong, Jieming Qu, Fengyun Gong, Yang Han, Yang Qiu, Jingli Wang, Ying Liu, Yuan Wei, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study. The Lancet, 395(10223):507–513, 2020.
[17] Ahmed T. Alhasanه, Hussein Alkattan, Alhumaima Ali Subhi, El-Sayed M. El-Kenawy, Marwa M. Eid. (2023). A Comparative Analysis of Methods for Detecting and Diagnosing Breast Cancer Based on Data Mining. Journal of Artificial Intelligence and Metaheuristics, 4 (2), 08-17.
[18] Ehsan khodadadi, S. K. Towfek, Hussein Alkattan. (2023). Brain Tumor Classification Using Convolutional Neural Network and Feature Extraction. Fusion: Practice and Applications, 13 (2), 34-41.
[19] Ashok Kumar Dwivedi. Performance evaluation of different machine learning techniques for prediction of heart disease. Neural Computing and Applications, 29(10):685–693, 2018.
[20] Taiwo Oladipupo Ayodele. Types of machine learning algorithms. New advances in machine learning, pages 19–48, 2010.
[21] Ya-Han Hu, Yi-Lien Lee, Ming-Feng Kang, and Pei-Ju Lee. Constructing inpatient pressure injury prediction models using machine learning techniques. Computers, Informatics, Nursing: CIN, 2020.
[22] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. Journal of artificial intelligence research, 4:237–285, 1996.

Doi: https://doi.org/10.54216/JAIM.060103
Received: February 10, 2023 Revised: May 11, 2023 Accepted: November 17, 2023