Journal of Control- Special Issue on COVID-19
ISSN (print) 2008-8345
ISSN (online) 2538-3752
Vol. 14, No. 5, Special Issue on COVID-19, page 131-140

Rapid COVID-19 Screening Based on the Blood Test using Artificial Intelligence Methods

S. Mehralian[1], E. Jalaeian Zaferani[1], Sh. Shashaani[1], F. Kashefi[1], M. Teshnehlab[1], H. Sokhandan[2], Z. Sadat Dibaji Forooshani[2], B. Montazer[2], Z. Joneidi[3], M. Vafapeyvand[2]

[1]Intelligent Systems Lab., Electrical & Computer Eng. Faculty of K. N. Toosi University of Technology, Tehran, Iran
[2]BMI hospital, Tehran, Iran
[3]Department of genetics and molecular medicine, Zanjan University of Medical Sciences, Zanjan, Iran

Received: 2021/01/29, Accepted: 2021/03/07

Abstract— Coronavirus Disease 2019 (COVID-19) caused by the SARS-CoV-2 virus is spreading rapidly worldwide and has led to widespread deaths globally. As a result, the early diagnosis of patients with COVID-19 is vital to control this dangerous virus's release. There are two common diagnosing methods, chest computed tomography scan (CT-scan) and Reverse Transcription Polymerase Chain Reaction (RT-PCR) test. The most significant disadvantages of RT-PCR molecular tests are the high cost and the long waiting time for test results. The common weaknesses of chest CT-scan are the need for a radiologist to analyze, a misdiagnosis of flu disease due to its similarity, and risky for pregnancy and infants. This article presents a low-cost, highly available method for early detection of COVID-19 based on Artificial Intelligence (AI) systems and blood tests. In this study, 6635 patient's blood tests are used. Experiments conducted using three machine learning algorithms. The results show that the proposed method can detect COVID-19 with an accuracy of %84 and an F1-score of %83. The trained model is being used in a real-world product through an online website called CODAS.

Keywords: Artificial intelligence, Blood test, Fuzzy system, Neural network, Support vector machine, COVID-19, Screen

غربالگری سریع کووید-۱۹ با استفاده از آزمایش خون و روش‌های هوش‌مصنوعی

سهیل مهرعلیان، عفت جلائیان زعفرانی،شهرزاد شعشعانی، فرناز کاشفی نیشابوری، محمد تشنه‌لب، حسین علی سخندان، زهرا سادات دیباجی فروشانی، بینا منتظر ، زینب جنیدی، مریم وفاپیوند

چکیده: بیماری COVID-19که به دلیل ورود ویروس SARS-CoV-2به بدن ایجاد می‌شود با سرعت چشمگیری در حال شیوع یافتن است. این ویروس به قدری خطرناک است که از زمان ظهور تاکنون جان بسیاری از افراد جهان را گرفته است. همین امر نشان می‌دهد که تشخیص زودهنگام افراد مبتلا به بیماری COVID-19برای کنترل شیوع این ویروس خطرناک بسیار حیاتی است. در حال حاضر دو روش تشخیصی رایج برای این بیماری، تصویربرداری سی تی اسکن از ریه و آزمایش مولکولی RT-PCR[1]است. روش تصویربرداری سی تی اسکن از ریه در کنار نقاط قوتی نظیر سرعت بالای اسکن، هزینه انجام اسکن پایین دارای نقاط ضعفی از جمله نیاز به رادیولوژیست جهت تحلیل تصاویر، معطلات ناشی از در معرض اشعه ایکس قرار گرفتن و خطرناک بودن این روش برای زنان باردار و نوزادان می‌باشد. در خصوص نقاط ضعف روش تشخیصی آزمایش مولکولی می‌توان به هزینه بالای انجام آزمایش، وابستگی به کیت‌های وارداتی و مدت زمان طولانی دریافت نتایج آزمایش اشاره کرد هرچند که این روش دارای صحت تشخیصی بالاتری نسبت به سی تی اسکن ریه می‌باشد، در این مقاله، ما روشی کم‌هزینه، سریع و در دسترس برای تشخیص زود هنگام بیماری COVID-19بر اساس مدل‌های هوش‌مصنوعی[2] و آزمایش خون روتین ارائه دادیم که علاوه بر تشخیص زودهنگام به دلیل قابلیت تکرارپذیری خوبی که دارد می‌توان از آن برای غربالگری افراد و

[1] Reverse Transcription Polymerase Chain Reaction (RT-PCR)
[2] Artificial Intelligence (AI)

Corresponding author: M. Teshnehlab

[DOR: 20.1001.1.20088345.1399.14.5.16.2]
[DOI: 10.52547/joc.14.5.131]

جوامع مختلف استفاده کرد. در این مطالعه، از آزمایش خون ۶۶۳۵ بیمار مراجعه کننده به بیمارستان بانک ملی ایران استفاده شده است. سه مدل یادگیری ماشین از قبیل شبکه عصبی فازی تطبیقی، ماشین بردار پشتیبان و شبکه‌های عصبی مورد ارزیابی قرار گرفت. نتایج نشان می‌دهد که روش پیشنهادی ما می‌توانند افراد مبتلا به COVID-19را با صحت ٪۸۴ و F1-score %83تشخیص دهد. همچنین گروه تحقیقاتی ما وب سایت آنلاین به نام CODAS طراحی کرده است تا مخاطبین بتوانند روش پیشنهادی ما را به راحتی و در دنیای واقعی ارزیابی نمایند.

کلمات کلیدی: هوش‌مصنوعی، آزمایش خون، سیستم فازی، شبکه‌عصبی، ماشین بردار پشتیبان، کووید-۱۹، غربالگری

I. INTRODUCTION

COVID-19 infectious disease or Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is caused by a new coronavirus [1]. Due to this disease's characteristics, such as high transmission rate, rapid transmission, no fixed clinical symptoms, global prevalence, and high mortality rate, the World Health Organization (WHO) has identified the new coronavirus as a pandemic [2]. This virus mainly causes lung damage, such as pneumonia and Acute Respiratory Distress Syndrome (ARDS) [3]. These complications may damage other organs and Disseminate Intravascular Coagulation (DIC) in a particular group of patients [4]. Besides, various organs, including blood vessels and kidneys, may be damaged by Covid-19.

The consequences of late diagnosis of the virus can impose a lot of material and medical costs. In this regard, early diagnosis of COVID-19 disease is vital.

There are several COVID -19 screening methods, which are described below:

1. **Reverse Transcription Polymerase Chain Reaction (RT-PCR) test**: It is a laboratory technique that examines the virus's genetic structure. This test shows an active virus in the upper and lower respiratory specimens [5]. Although the RT-PCR test is the most accurately known method for detecting the COVID-19, its drawbacks, such as 1) its results are dependent on expensive kits that are limited to a few countries. 2) Only molecular laboratories are allowed to do so. 3) Due to this test's high cost, not all sections of society can use it. 4) Announcing the results is time-consuming (24 to 72 hours) [6, 7].
2. **Computed Tomography scan (CT-scan) of the lungs:** CT-scan is one of the X-ray medical imaging methods that can be used to view the tissues inside the body and examine their shape [8]. This imaging technique has been instrumental in diagnosing the coronavirus and its destructive effects on the lungs, but the results are not reliable in some cases. 1) The shared border between coronavirus and the flu, which has similar effects on the lungs [9], 2) In some people, the virus does not infect the lungs, and a normal CT-scan does not mean the absence of the virus. 3) If the virus affects the lungs, this effect can be detected on CT images three days after the virus enters the body. Imaging earlier than three days shows normal results when, in fact, it is not [10]. 4) Due to the higher dose of exposure, the usage is limited in a year (prohibition of use for pregnant women and infants) [11].
3. **C - Reactive Protein (CRP) test**: CRP test is a blood test marker for inflammation in the body [12]. Experience has shown that although this parameter increases in the blood when the patient's body is exposed to the coronavirus, this increase indicates any other infectious virus type. In other words, increasing this parameter can indicate exposure to the virus, but this is not necessarily the case [13].
4. **Antibody tests**: The antibody test does not look for the virus itself but to see if the immune system has responded to the infection. Therefore, it is a method to show a past infection with the virus that causes COVID-19 and not suitable for screening [14].
5. **Physical symptoms**: The most common severe COVID-19 symptoms upon the initial presentation are fever, dry cough, dyspnea, and tiredness [15]. Although the physical symptoms do not occur in the same way in all people, the doctor can decide on the treatment process according to the known symptoms. The problem is that the patient may not pay attention to the symptoms or may be confused with other diseases such as the common cold or flu due to the similarity of their symptoms [16]. Therefore, the virus progresses in the infected person's body and will have severe consequences for them and other people.

Although each of the methods mentioned above has its advantages and disadvantages, the main problem is that all methods show the person is infected currently. Whereas population screening to determine COVID-19 infection is a longer-term need, it is necessary to perform one of the tests periodically, and the cost of

performing the periodic test and its long-term disadvantages should be considered.

This paper proposes a new screening method that identifies the probability of getting a coronavirus infection based on blood test data with artificial intelligence algorithms. This method is easy, fast, low cost, and available to everyone with the ability to do the test frequently suitable for screening. Besides, this paper shows how blood test data can be acquired and used to detect data set.

In the next section, the summary of related works is expressed. The data section, including data science, data collection, miss value, and data limitation, are explained in section III. The proposed methodology is described in section IV, including a brief overview of ANFIS, SVM, and neural network techniques. Eventually, results are presented in V.

II. RELATED WORKS

Since the coronavirus outbreak, researchers worldwide have conducted sufficient research methods to help governments fight the dangerous virus.

In the article [17] regarding the RT-PCR test, the sensitivity and effect of CT imaging to detect the presence of COVID-19 virus in the patient's body were investigated. Article [18] is a case study conducted from February 2020 to January 2020 on coronavirus. It is an example of research using artificial intelligence to diagnose coronavirus. However, research does not end there, and several studies, including [19] and [20], have been conducted to increase coronavirus diagnosis accuracy based on CT images.

In the article [21], people were screened using deep neural network algorithms on CT-scan images. In this paper, relying on screening and quarantine and appropriate treatment to control the prevalence of coronavirus disease, it has been hypothesized that in-depth machine learning methods can extract specific features of COVID-19 and make a clinical diagnosis before pathogen testing. It was thus saving critical time to control the disease. This study examined 453 CT images and finally obtained the results with an overall accuracy of 82.9%.

In [22], first, the advantages and disadvantages of conventional diagnostic methods were discussed, then two models of machine learning classification examine to identify coronavirus using common blood parameters. The number of samples was 279 patients (177 positive results, 102 negative responses) collected in San Riley Emergency Hospital (Milan, Italy). Samples Targets were chosen by using the RT-PCR test. The results showed an accuracy between 82% to 86% and sensitivity between 92% and 95%.

Although RT-PCR molecular test is known as the golden test for coronavirus detection, it has disadvantages such as incomplete sensitivity (about 70%). Spending much time, especially in countries with limited energy sources, is not cost-effective was discussed in [23]. Therefore, it has proposed the automatic and accurate diagnosis of COVID-19 using medical imaging techniques, which are more widely available and can reap the benefits. In paper [23], a hierarchical neural network model is designed to classify chest CT-scan with COVID-19. The data includes 5,801 images with 97.8% cross-validation, 99.3% sensitivity and 99.6% positive predictive value.

Research [24] and [25] with a variety of radiographic images including CT, Positron Emission Tomography (PET), lung ultrasound, and Magnetic Resonance Imaging (MRI) for the diagnosis, treatment, and follow-up of people with coronavirus. These articles use artificial intelligence to model the relationship between image properties and coronavirus.

Due to the advantages of artificial intelligence algorithms in controlling the COVID-19 epidemic, the number of artificial intelligence techniques in this field has increased. Although these techniques are a good starting point for COVID-19 pandemic control, they differ in purpose, artificial intelligence synthesis methods, data sets, and validation approaches. This increase and variation in the number of proposed artificial intelligence techniques can confuse decision-makers and lead them to the dilemma of using the appropriate algorithm. However, there are limited studies that evaluate, analyze, and summarize the unresolved problems and shortcomings of current artificial intelligence techniques for COVID-19.

A routine blood test has useful information about the infection of the body. Although blood tests provide valuable information about the disease to a physician, the nature of the coronavirus is such that specialists cannot easily and only rely on the results of blood tests to decide whether or not a person gets COVID-19. Evidence of this statement is that blood tests are not used for COVID-19 screening confidently to the best of our knowledge.

However, efforts are still underway to find a practical relation between blood parameters and the coronavirus. In Article [26], 169 patients' blood tests (a total of 256 samples) containing flue, COVID-19, tuberculosis, and lung cancer were used to train a machine learning model. The random forest was applied to the 11 parameters chosen by statistical methods. The online test is provided on their website[3]. Although the results report high accuracy, to our knowledge, the complexity of the coronavirus in the human body is so great that a small number of samples (256 samples out of 169 patients) cannot show this

[3] http://lishuyan.lzu.edu.cn/COVID2019_2/

complexity well. On the other hand, data were collected from several hospitals. Because each medical device has a different accepted range according to the measurement method, it must be considered in the preprocessing stage.

Table 1 Diagram of average value of features for male and female induvituals grouped by test result (Covid+ and -)

Gender	Covid-19 Result	age	HbA1C	CRP	Lymphocyte	Platelet	W.B.C
Female	Covid +	62.27	7.08	28.94	24.71	247.45	8.08
	Covid -	48.13	6.26	6.18	30.91	263.05	7.76
Male	Covid +	59.67	7.47	26.57	24.14	217.1	7.92
	Covid -	47.46	6.53	5.47	32.06	232.87	7.82

III. DATA

A. *Data collection*

The data collected in this article are the laboratory data of Bank Melli Iran (BMI) Hospital[4], Tehran, Iran, from March 2020 to June 2020. All these data are real and collected from the hospital laboratory software. It contains routine blood tests such as White Blood Cell (WBC), platelet, CRP, Lactate Dehydrogenase (LDH), SGOT, SGPT, lymphocyte, Fast Blood Sugar (FBS), HbA1c, blood group, Rh factor, zinc, and procalcitonin. For some patients, however, some of these blood parameters are not available. Therefore, some occasional miss values appear in the dataset, inevitably. We deal with the miss values in the next section.

Table 1 shows a summary of the characteristics of the features used in the study. The table shows the average value of different features grouped by Gender and label (Covid+ or Covid-). As you can see in the table, there is a significant distance between values for patients affected by Covid. For instance, the average value for CRP in male Covid+ patients is 26.56, while it is just ۵/۴۷ for Covid- ones. Another fact that can be statistically derived from the table is about the age of patients. The average age for Covid+ patients for both male and female groups is noticeably higher.

The data were labeled using a CT-scan report. Individuals who reported positive or suspected coronations were assigned the label "*one*," and the remainder was labeled "*zero*."

In general, the number of 10328 blood tests of 8939 unique people, including 3989 women and 4950 men, were collected, in which 623 test reports were positive, and 9705 ones were negative. Patients include outpatients and inpatients and aged from 3 to 102 years.

Figure 1 shows daily and cumulative Covid+ patients in BMI hospital from February 20 to late January. The green line also shows the overall trend of daily patients peaked in late May and decreased until the end of January.

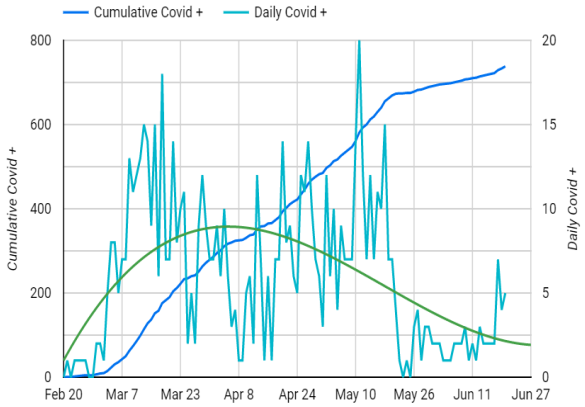

Figure 1 Daily and cumulative Covid+ test results in BMI hospital

One can consider the blood parameters related to coronavirus by drawing related figures. Figure 2 indicates the distribution of blood groups related to positive COVID-19.

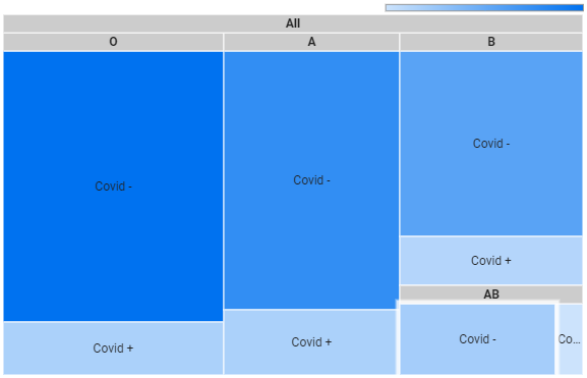

Figure 2: Distribution of blood groups with different test results in the BMI hospital

Figure 3 demonstrates the histograms of coronavirus, related to the sex of infected and non-infected people.

[4] http://bmihis.com/

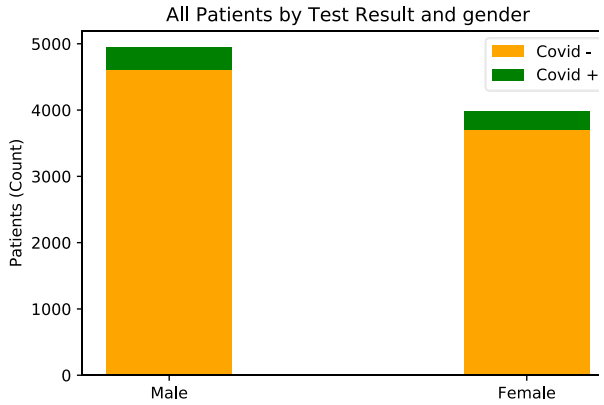

Figure 3: Histogram of patients' gender grouped by the test result

B. Miss Values

Working with real-world data, there are many challenges before making data suitable for learning. One of these challenges is to missing values in data. The empty holes in data tables can make difficulties for data scientists, especially in healthcare. There are several ways to handle this difficulty: There are some typical reasons for missing values that range from programming errors to faults of the users responsible for data entry.

There are some strategies to deal with miss values. Choosing which strategy to use depends on factors such as the feature itself, the expected type of the feature (Numerical, Boolean, and Categorical), and the ratio of missed values. For example, suppose the ratio of the missed values is enormous. In that case, we can omit the feature because the present data cannot represent the real distribution from which the feature came and can be misleading for machine learning algorithms. In this study, we omitted some features because of the high ratio of miss values. Discard of the data can be applied to either a row or a column of the data.

If the ratio of miss values is not very high to omit, we can impute the missed values using different well-known methods, i.e., infer the values from a known part of the data. Imputing missing values can be Univariate or Multivariate: In univariate methods, if we have missed values in i^{th} feature, only known data of the i^{th} feature will be used to fill the missed values. In contrast, in multivariate methods.

C. Data limitations

AI systems need accurate training samples to achieve precise results. Each sample of data gets 0 or 1 label related to positive or negative for COVID-19, respectively. Labeling is obtained form of CT-scan imaging reports which preparing by a radiologist. COVID-19 labeling based on CT images' report enters some error in the simulation results, which is inevitable.

These probability errors are as follows:

1. Human error in reporting
2. In some case, coronavirus does not affect on lung, and the report shows Normal, but it is not correct
3. CT Imaging less than three days after the virus enters the body
4. No clear boundary between coronavirus and the flu.
5. Suspected patient: some people suspected of being infected with coronavirus or not. They reported as positive for COVID-19, which can also affect the accuracy of the AI method.

As can be deduced, there is a kind of uncertainty in dealing with this problem. A fuzzy system is used to model this uncertainty.

IV. METHODOLOGY

In this research, we have used different intelligent methods for COVID-19 detection. The dataset contains real clinical data. Clinical technicians acquire the data in BMI hospital. Therefore, in this research, we are two teams, including the *clinical* and *artificial intelligent* teams.

- The *clinical team* was responsible for acquiring data from all infected and non-infected people and their analysis.
- The *artificial intelligent* team is responsible for preparing the scientific data analysis and designing the appropriate algorithm with high performance.

The framework of the study is indicated in Figure 4. It has several steps, as follows:

- **Data acquisition**: the real laboratory data is collected from the hospital, which is described in section II in detail.
- **Preprocessing**: In this stage, data will be prepared for subsequent analyzing such as normalizing and solving miss value.
- **Data analysis**: In this phase, the data is visually described to medical experts and designers of intelligent models to give an overview of the data.
- **Decision making with AI**: designing the classification method, training, and testing of the model are done in this section.
- **Website design and development**: The output model of this project is deployed on a website in the form of a web app and can be accessed online.

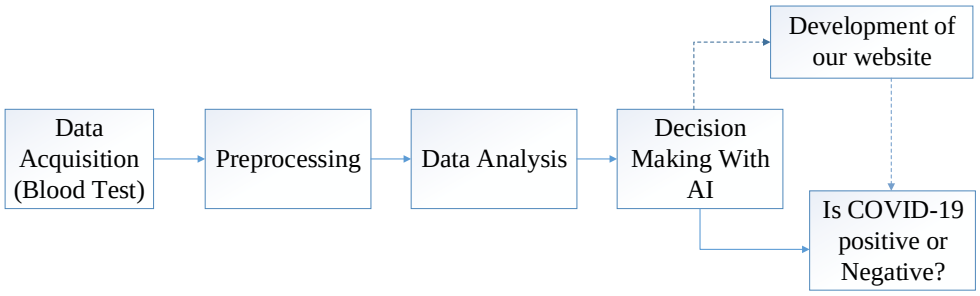

Figure 4: Framework of our study

In the following sections, the three AI methods are introduced.

A. ANFIS

Adaptive Neural Fuzzy Inference System (ANFIS) is a type of neural network structure based on the Takagi-Sugeno fuzzy system. This structure inherits both the strength of fuzzy systems and neural networks. In other words, it enjoys interpretability and representation of the prior knowledge from fuzzy systems nature and can tune parameters using back-propagation from neural networks.

The main structure includes two parts of the *antecedent* and *conclusion*. The typical rule is as follows:

R^{l}: If x_1 is M_{i1} and x_2 is M_{i2}...and x_n is M_{in} Then $y=f^{l}(x_1,x_2,...,x_n)$

Where IF and THEN are *antecedent* and the *conclusion* parts, respectively. x_n is the input vector and M_{ij} is a membership function. $f^{l}(.)$ is a mathematical function which is not fuzzy.

The output is presented as:

$$y(x)=\frac{\sum w_i f_i}{\sum_{i=1}^{m} w_i} \quad (1)$$

Where *m* is rule number and W_i is firing rule base product inference which is obtained as:

$$w^{i}(x)=\mu_{1i}(x_1)*....*\mu_{ni}(x_n) \quad (2)$$

Moreover, rule normalization is defined:

$$\bar{W}_j=\frac{w^{j}(x)}{\sum_{j}^{m} w^{j}(x)} \quad (3)$$

Finally, the simplification of fuzzy output is defined as:

$$y=\sum_{j}^{m} \bar{W}_j f_j \quad (4)$$

For generating the Gaussian membership, the function is depicted as Figure 5:

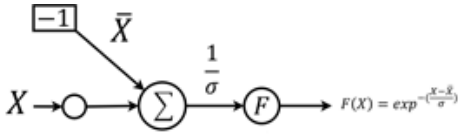

Figure 5: Generating of Gaussian membership function.

Using this Gaussian membership function (MFs), it can be possible to generate the antecedent part with the following steps, as shown in Figure 6. These steps have two inputs, and for each input, three MFs.

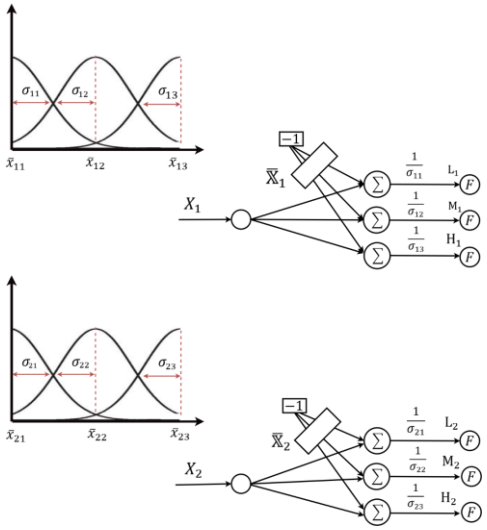

Figure 6: Design of an antecedent part for two inputs.

Where $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ are inputs, and $\bar{\boldsymbol{X}}_{\boldsymbol{ij}}$ and $\boldsymbol{\sigma_{ij}}$ are mean center and standard deviation of j^{th} MF, respectively. The next step generates a rule base, which has nine rules and is showing in Figure 7.

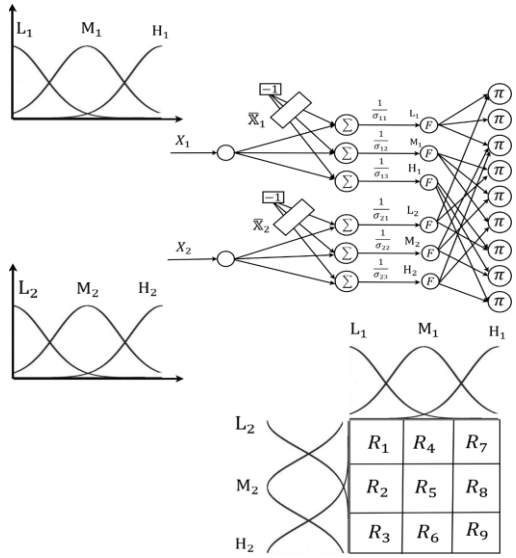

Figure 7: General shape of the rule base for two inputs, and each input includes three MFs.

This structure is for *n* inputs and one output with three MFs for each input. This structure is called the first-order Sugeno type of ANFIS structure and has tunable parameters in antecedent and conclusion parts that lead to more flexibility and precision. This structure is shown in Figure 8.

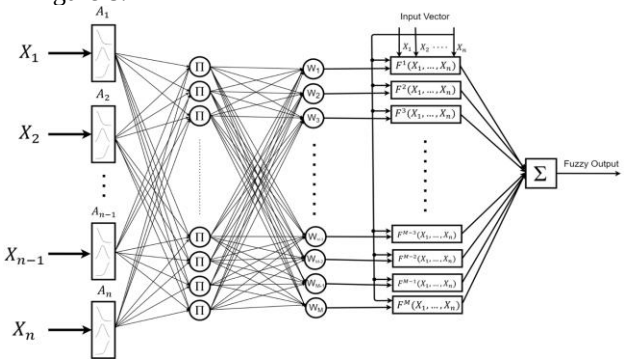

Figure 8: Typical structure of ANFIS with *n* inputs and one output.

B. Support Vector Machine

Support Vector Machines (SVMs) are among the most popular machine learning tools that can be used for both classification and regression problems. It is based on a statistical learning framework and offered one of the most robust prediction models among shallow learning methods. It deals with a subset of data, called Support Vector, for training the algorithm. This study used three implementations of SVM algorithms to classify data: Support Vector Classification (SVC), Linear-SVC, and Nu-SVC. They are different in the number of parameters they use to work and kernel function.

C. Neural Network

Inspired by the human brain, neural networks are currently among the most efficient machine learning algorithms in various facets and applications. In our research work, we also tried to use this ideology for our research application too. To have a fair comparison between algorithms, we use one of the simplest and popular forms of neural networks called Multi-Layer Perceptron (MLP). In its vanilla form, this network consists of three layers: Input, Hidden, and Output layers. It also utilizes a supervised technique for optimization parameters of neural networks, which are called Back-propagation. Neural networks are powerful tools that have a high capability to learn according to a data set.

V. SIMULATION AND RESULTS

To show the proposed method's stability and performance and selected features, we have implemented different versions of the algorithms described above, namely ANFIS, SVM, and NN. The results of these methods are summarized in Table 1. All metrics in this table are average, and standard deviation of five runs, low variance in each metric shows the algorithms' stability.
The data used in the study split into train and test set, %80 for train and the rest for testing purpose, the results in Table 1 are reports the performance of the algorithms on test data and train data just used in the training phase of the models.
As shown in the table, the ANFIS method has the best recall, about 0.81, compared with other methods. A better recall is vital because each false-negative prediction cost is far higher than the false-positive one in this application. The other methods, such as MLP neural network, also have an acceptable performance with a recall of 0.78. Regarding execution time, the SVC-base method, thanks to dealing with a subset of data, namely support vectors, outperforms neural network-based ones, however. The SVC, Linear SVC, and Nu-SVC are significantly faster than the ANFIS and MLP neural networks. It makes them an appropriate approach for the situations in which the final model's speed and execution time are crucial.
Another metric that describes the performance of algorithms used in the study is the F1-score; it is the harmonic mean of recall and precision. In applications in which both recall and precision must have a higher value, F1-score can be used. Regarding this metric ANFIS have better performance among all method implemented.

Table 2: The prediction of different methods of simulation results.

Method	Metric	Precision	Recall	F1-score	Time (sec.)
		Avg. ± STD	Avg. ± STD	Avg. ± STD	Avg. ± STD
ANFIS	Covid-	0.814 ± 0.01	0.868 ± 0.03	0.84 ± 0.01	80.65 ± 0.98
	Covid+	0.866 ± 0.03	0.81 ± 0.01	0.84 ± 0.01	
	Marco Avg.	0.838 ± 0.01	0.838 ± 0.01	0.836 ± 0.01	
	Weighted Avg.	0.838 ± 0.02	0.836 ± 0.01	0.836 ± 0.01	
MLP	Covid-	0.796 ± 0.03	0.888 ± 0.3	0.838 ± 0.02	84.9 ± 1.52
	Covid+	0.878 ± 0.03	0.78 ± 0.05	0.824 ± 0.04	
	Marco Avg.	0.84 ± 0.02	0.83±0.03	0.83±0.03	
	Weighted Avg.	0.838 ± 0.02	0.832 ± 0.03	0.832 ± 0.03	
SVC	Covid-	0.78 ± 0.02	0.888 ± 0.03	0.83 ± 0.01	0.14 ± 0.012
	Covid+	0.868 ± 0.03	0.744 ± 0.04	0.8 ± 0.02	
	Marco Average	0.822 ± 0.02	0.812 ± 0.02	0.812 ± 0.02	
	Weighted Average	0.824 ± 0.02	0.814 ± 0.01	0.812 ± 0.02	
Linear SVC	Covid-	0.774 ± 0.01	0.854 ± 0.01	0.81 ± 0.008	0.005 ± 0.0006
	Covid+	0.826 ± 0.02	0.734 ± 0.01	0.778 ± 0.01	
	Marco Average	0.8 ± 0.01	0.794 ± 0.01	0.796 ± 0.01	
	Weighted Average	0.8 ± 0.01	0.796 ± 0.01	0.796 ± 0.01	
Nu-SVC	Covid-	0.752 ± 0.02	0.902 ± 0.03	0.82 ± 0.02	0.116 ± 0.01
	Covid+	0.886 ± 0.03	0.716 ± 0.02	0.79 ± 0.02	
	Marco Average	0.822 ± 0.02	0.808 ± 0.02	0.806 ± 0.02	
	Weighted Average	0.824 ± 0.02	0.806 ± 0.02	0.806 ± 0.02	

VI. OUR WEBSITE INFORMATION

As mentioned before, the trained ML model is used as a backend and decision-making engine of an application as a real-world website named CoDAS. This website takes the blood test metrics of the patients and outputs the probability of COVID-19 infection.
CoDAS has three main modules:

- A Website that is a simple Flask application written in Python and serves as an interface layer.
- Database that holds track of patients tests and user information
- ML model is the central part of the application and takes inputs from the interface layer and delivers the results to it. This module is the main contribution of this study.

An overview of our website is shown in the following figures.

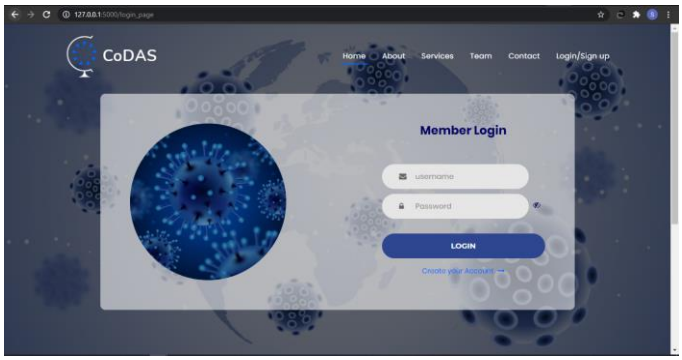

Figure 9: Login page of the website

After login into the website, the new test page appears and asks for data entry. On this page, four blood parameters and three profile information must be entered—Figure 10.

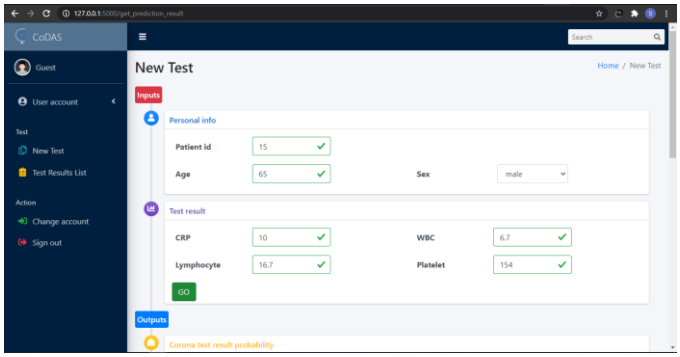

Figure 10: Data entry of the website

The normal ranges of four metrics from laboratory tests and the current test's values are visualized on the website. A test result is an integer number between 0-1, representing the probability of COVID-19 infection (Figure 11).

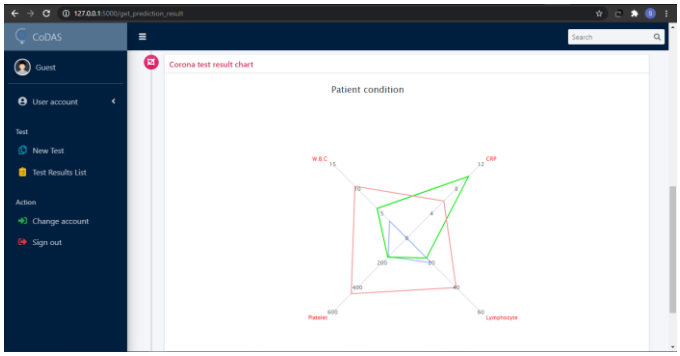

Figure 11: Result chart and probability of getting coronavirus

VII. CONCLUSION

In this paper, a method based on artificial intelligence methods is proposed to screen people infected with the coronavirus. For this purpose, the routine blood tests of 6635 patients of BMI hospital were collected. After preprocessing, due to existing uncertainty in the dataset, the fuzzy classification was used. Besides, two artificial intelligent methods were used as well as a better comparison.

The results show that our models can effectively distinguish COVID+ test results from COVID- one with accuracy and F1-score of more than %84. It is worth mentioning, all methods in this study are shallow machine learning models because using a deep model is a kind of over-engineering in designing a model for an order of few thousand with less than ten dimensions. The deep models show their superiority when there are plenty of high dimensional data to train a model with millions, or even billions, parameters.

The properties of our method that distinguish it from other screening methods are as follows:

1. Focus on practical solutions
2. Effectiveness and efficiency everywhere
3. The simplicity of analyzing the results
4. Repeatability of the results
5. ease of use
6. availability for everyone

Using CT scans of the patients' chest beside blood test results can improve the proposed method's accuracy and reliability; we are planning for this augmentation; however, it can be regarded as a future work recommendation.

VIII. REFERENCES

[1] T. Acter, N. Uddin, J. Das, A. Akhter, T. R. Choudhury, and S. Kim, "Evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as coronavirus disease 2019 (COVID-19) pandemic: A global health emergency," *Science of the Total Environment*, p. 138996, 2020.

[2] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges," *International journal of antimicrobial agents*, p. 105924, 2020.

[3] P. MARKOWICZ *et al.*, "Multicenter prospective study of ventilator-associated pneumonia during acute respiratory distress syndrome: incidence, prognosis, and risk factors," *American journal of respiratory and critical care medicine*, vol. 161, no. 6, pp. 1942-1948, 2000.

[4] A. D. Makatsariya *et al.*, "Coronavirus disease (COVID-19) and disseminated intravascular coagulation syndrome," *Obstetrics, gynecology and reproduction*, vol. 14, no. 2, pp. 123-131, 2020.

[5] V. Corman *et al.*, "Detection of a novel human coronavirus by real-time reverse-transcription polymerase chain reaction," *Eurosurveillance*, vol. 17, no. 39, p. 20285, 2012.

[6] A. T. Xiao, Y. X. Tong, and S. Zhang, "False-negative of RT-PCR and prolonged nucleic acid conversion in COVID-19: rather than recurrence," *Journal of medical virology*, 2020.

[7] F. Cabitza *et al.*, "Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests," *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 1, no. ahead-of-print, 2020.

[8] A. Bhandary *et al.*, "Deep-learning framework to detect lung abnormality–A study with chest X-Ray and lung CT scan images," *Pattern Recognition Letters*, vol. 129, pp. 271-278, 2020.

[9] H. Khorramdelazad, M. H. Kazemi, A. Najafi, M. Keykhaee, R. Z. Emameh, and R. Falak, "Immunopathological similarities between COVID-19 and influenza: Investigating the consequences of Co-infection," *Microbial pathogenesis*, p. 104554, 2020.

[10] B. S. Bleier and K. C. Welch, "Preprocedural COVID-19 screening: Do rhinologic patients carry a unique risk burden for false-negative results?," in *International forum of allergy & rhinology*, 2020, vol. 10, no. 10, pp. 1186-1188: Wiley Online Library.

[11] Y. Li and L. Xia, "Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management," *American Journal of Roentgenology*, vol. 214, no. 6, pp. 1280-1286, 2020.

[12] J. Vernarelli and J. Lambert, "Flavonoid intake is inversely associated with obesity and C-reactive protein, a marker for inflammation, in US adults," *Nutrition & diabetes*, vol. 7, no. 5, pp. e276-e276, 2017.

[13] W. Ling, "C-reactive protein levels in the early stage of COVID-19," *Medecine et maladies infectieuses*, 2020.

[14] J. J. Deeks *et al.*, "Antibody tests for identification of current and past infection with SARS-CoV-2," *Cochrane Database of Systematic Reviews*, no. 6, 2020.

[15] "World Health Organization."

[16] T. Struyf *et al.*, "Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19 disease," *Cochrane Database of Systematic Reviews*, no. 7, 2020.

[17] X. Mei *et al.*, "Artificial intelligence–enabled rapid diagnosis of patients with COVID-19," *Nature Medicine*, pp. 1-5, 2020.

[18] Y. Xing, P. Mo, Y. Xiao, O. Zhao, Y. Zhang, and F. Wang, "Post-discharge surveillance and positive virus detection in two medical staff recovered from coronavirus disease 2019 (COVID-19), China, January to February 2020," *Eurosurveillance*, vol. 25, no. 10, p. 2000191, 2020.

[19] Y. Fang *et al.*, "Sensitivity of chest CT for COVID-19: comparison to RT-PCR," *Radiology*, p. 200432, 2020.

[20] E. Neri, V. Miele, F. Coppola, and R. Grassi, "Use of CT and artificial intelligence in suspected or COVID-19 positive patients: statement of the Italian Society of Medical and Interventional Radiology," *La radiologia medica*, p. 1, 2020.

[21] S. Wang *et al.*, "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)," *MedRxiv*, 2020.

[22] D. Brinati, A. Campagner, D. Ferrari, M. Locatelli, G. Banfi, and F. Cabitza, "Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: a Feasibility Study," *medRxiv*, 2020.

[23] M. Zokaeinikoo, P. Kazemian, P. Mitra, and S. Kumara, "AIDCOV: An Interpretable Artificial Intelligence Model for Detection of COVID-19 from Chest Radiography Images," *medRxiv*, 2020.

[24] D. Dong *et al.*, "The role of imaging in the detection and management of COVID-19: a review," *IEEE reviews in biomedical engineering*, 2020.

[25] L. Li *et al.*, "Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT," *Radiology*, 2020.

[26] J. Wu *et al.*, "Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results," *medRxiv*, 2020.