Received: October 14, 2020. Revised: December 14, 2020. 178

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Adaptive Neuro-Fuzzy Inference System (ANFIS) for Rapid Diagnosis of COVID-19 Cases Based on Routine Blood Tests

Mohanad Deif[1] Rania Hammam[1] Ahmed Solyman[2]*

[1]*Department of Bioelectronics, Modern University of Technology and Information (MTI) University, Egypt*
[2]*Department of Electrical and Electronics Engineering, Istanbul Gelişim University, Turkey*
* Corresponding author's Email: aaasahmed@gelisim.edu.tr

Abstract: This article presents an Adaptive Neuro-Fuzzy Inference System (ANFIS) approach to rapidly detect COVID-19 cases using commonly available laboratory blood tests. Current Reverse transcription-polymerase chain reaction (RT-PCR) tests for COVID-19 suffer from several limitations including false-negative results as large as 15-20%, the need for certified laboratories, expensive equipment, and trained personnel; hence the development of an efficient diagnosis system that provides prompt and accurate results is of great importance to control the spread of the virus. Therefore, it was aimed to develop an intelligent system to analyze blood tests and identify significant hematological indicators to support COVID-19 diagnosis. This study interpreted the ANFIS model performance by shapely values to identify the most important and decisive parameters that could assist clinicians in making effective patient management decisions. The findings of this study revealed that WBC (White blood cells) & Platelet counts can act as relevant and significant indicators for the diagnosis of COVID-19 patients. Moreover, the proposed ANFIS model achieved a high prediction accuracy as it was able to discriminate between positive and negative COVID-19 patients with an Accuracy, Sensitivity, and Specificity rates of 95%, 75%, and 97.25% respectively even though 10 % only of the data was positive. Therefore by combining available and low-cost blood test results to analysis based on the ANFIS model, we were able to provide an efficient and robust system to diagnose COVID-19.

Keywords: Adaptive neuro-fuzzy inference system (ANFIS), COVID-19 diagnosis, Routine blood tests, SHAP values, Hematologic parameters.

1. Introduction

The coronavirus epidemic, caused by the SARS-CoV-2 virus and known as COVID-19, started in December 2019 in Wuhan, China, and extent quickly through the world [1]. The incubation period of COVID-19 is as long as 2 weeks or longer, and it is highly infectious [2] in which the lung tissue is the target organ. Some patients rapidly deteriorate into acute respiratory failure, acute respiratory distress syndrome (ARDS) or multiple organ failure, reducing the time required to save them. [3, 4]. Consequently, it is essential to find prognostic tools to classify possible and severe cases of COVID-19, which helps in early treatment. In addition, diagnostics can play a significant part in disease control, permitting the fast employment of controller measures that limit the disease spread through positive patient discovery, separation and contact tracing.

Molecular techniques have been proven accurate diagnoses because they can target and identify specific pathogens by identifying its genetic material or identifying unique markers of the pathogen itself. Several reverse transcription-polymerase chain reaction (RT-PCR) kits have been developed to detect the SARS-CoV-2 genetically [5]. RT-PCR testing relies on its ability to amplify a minimal quantity of viral genetic material in a sample and is considered an essential part for SARS-CoV-2 virus detection. RT-PCR tests for COVID-19 typically use samples obtained using swabs from the upper respiratory system. However, due to the global spread of the virus, the strong demand for RT-PCR tests highlights the drawbacks of this form of diagnosis on

a large scale, such as long time to acquire the results (over 2-3 h to produce results) and the necessity for accredited laboratories, costly tools and skilled staff [6]. In addition, RT-PCR entails general analytical and pre-analytical problems that can risk the diagnostic accuracy of the test [7]. For this type of test, current studies have recorded up to 20 percent false-negative results. [8, 9]. These limits make RT-PCR inadequate for rapid and large-scale screening targeting for fast diagnosis of patients. Such restrictions become even more highlighted in developing countries, which suffer from limited resources. Thus, there is a critical need for substitute tests to rapidly identify infected COVID-19 patients in order to avoid transmission of the virus and ensure a quick treatment for patients.

Other types of tests that are currently in development include serological testing (i.e., blood tests for specific antibodies) which deliver results promptly, however, they are non-generic for COVID-19, and may have quite poor specificity and sensitivity [10-13]. Antibodies tests as Immunoglobulins (IgA, IgG, IgM) tests can't identify the SARS-Cov2 existence directly, in fact, they identify the latest infection serological evidence. Furthermore, Li et al. [6] specified that positive reaction may be from other coronavirus and influenza virus antibodies.

Since laboratory medicine is a vital part in the early detection, diagnosis, and controlling of many diseases, current studies were conducted on the diagnoses of COVID-19 patients via analyzing the blood tests of suspected people [14-16]. The studies were performed in order to illustrate statistically significant variations which may help to detect COVID-19 patients. It has been shown that a few hematological parameters were obviously altered in COVID-19 patients. For instance, Davide Ferrari et al. [14] reported that there was a robust connotation between COVID-19 positively tested patients and a small WBC count in patients who were admitted to the San Raffaele Hospital (Milan, Italy) emergency room. Cheng et al. [17] presented a study on the clinical features and CT manifestations of positively and negatively tested COVID-19 patients in a single-center study in Shanghai, China. The study revealed a significant connotation between low counts of WBC and platelets and patients with COVID-19. Moreover, the studies of Fan et al. [18] and Huang et al. [4] showed cases of leukopenia (low WBC count) and lymphopenia (low lymphocyte counts) of COVID-19 patients upon admission to the hospitals. Guan et al. [19] reported that in the routine blood test of COVID-19 patients upon admission to the hospital, the rate of reduced white blood cells or lymphocyte amount was mutual, and a decrease in the platelets count was observed in the patients. Hu Yun et al. [20] have also proved in their study that there was a decrease in WBC count, Platelet count, Basophils count & Eosinophil count in positively tested COVID-19 patients. Another study was presented by Xiaofang Zhao et al. [21] who investigated the alteration in platelet counts between non-survivors and survivors infected with COVID-19. The authors found that there was a decrease in the platelet count of COVID-19 patients in the primary phases of the illness and that it was reduced in non-survivors more than in survivors in 1 week after admission. They found that the disease progression is related to the risk of reducing the number of platelets.

Recently, Machine learning algorithms have been used to diagnose diseases through the study of hematological parameters [22, 23]. It was proved that a simple blood test with the combination of an effective machine learning method would help diagnose negative and positive COVID-19 patients. Therefore, the present study aims to propose an Adaptive Neuro-Fuzzy Inference System (ANFIS) approach to rapidly detect COVID-19 cases using commonly available laboratory blood tests. The use of blood tests has the advantages of being cheaper and faster than other methods of diagnosis and thus providing a more reachable system. Besides, as pointed out from previous studies that hematological parameters can be indicators for the degree of sternness and the risk factors of COVID-19. Therefore, the identification of these criteria can be necessary for patients to promote clinical treatment. In this context, it was essential to develop intelligent systems to analyze blood tests and identify significant hematological indicators for the presence of COVID-19.

2. Contribution

This study contributed to the proposal of a Neuro-Fuzzy Inference System (ANFIS) model to rapidly detect COVID-19 cases using commonly available laboratory blood tests. Identification of the most relevant Clinical variables that had a significant influence on the classifier decision and therefore supporting the COVID19 diagnosis was also presented. Also developing a model classifier having a high prediction accuracy despite that 10 % only of the data was positive for COVID 19 proved that the ANFIS classifier is a reliable tool for the rapid and efficient detection of COVID-19 patients. Interpretation of the model both globally and locally was easily performed using SHAP values which can further guide in medical decision-making.

3. Materials

The present study was applied on the results taken from the Kaggle dataset Diagnosis of COVID-19 and its clinical spectrum created by the Hospital Israelita Albert Einstein in São Paulo, Brazil [24]. The dataset provided by the Israelita Albert Hospital consisted of 5644 records in which there were 559 cases (which constitutes 10% of the dataset) that were diagnosed with COVID 19 by employing the gold standard method; the reverse transcription-polymerase chain reaction (RT-PCR)[24]. The dataset contained 39 variables (predictors), and one target outcome variable, which is a binary variable that indicates whether the patient is tested positive or negative and it is given by the name COVID-19 exam result.

4. Methodology

4.1 Preliminaries

4.1.1. Adaptive neuro-fuzzy inference system (ANFIS) review

The Neuro-Fuzzy Inference System (NFIS) combines the Fuzzy logic system (FLS) and the artificial neural networks (ANN). Combining this method with neural networks, produces significant results, which can provide rapid and accurate detection of COVID-19 cases by analyzing hematological parameters based on collected blood test samples.

The adaptive network-based fuzzy inference systems (ANFIS) are used to unravel difficulties associated to parameter recognition issues [25]. This parameter recognition is achieved by integrating the back-propagation gradient descent and the least-squares approach via a hybrid learning law.

ANFIS is a representation for graphical network of the Takagi-Sugeno-type fuzzy inference system possessed with the neural learning abilities. The architecture of the ANFIS is shown in Fig. 1. Fixed nodes are represented by the circular nodes, whereas the nodes that have parameters to be learned are represented by the square nodes.

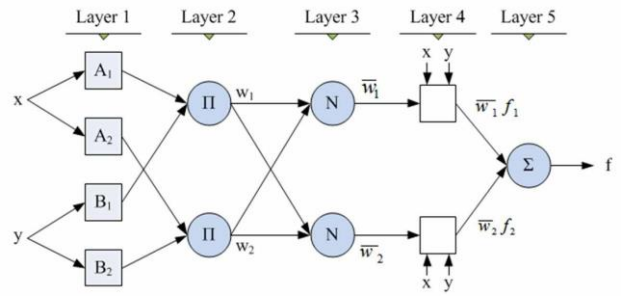


Figure. 1 Structure of the ANFIS network

a. ANFIS Architecture

ANFIS system architecture consist of five layers:

Layer 1: Fuzzification layer is the first layer. Every node i in this layer is a square node and is shown in Eqs. (1) and (2):

$$O_{1,i} = \mu_{A_i}(x) \qquad for\ i = 1,2 \tag{1}$$

$$O_{1,i} = \mu_{B_{i-2}}(y) \qquad for\ i = 1,2 \tag{2}$$

Where $\mu_{A_i}(x)$ and $\mu_{B_{i-2}}(y)$ fuzzy membership function (MF). In this paper, the following Triangular MF are used.

$$Triangular(x;a,b,c) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ \dfrac{c-x}{c-b} & b \le x \le c \\ 0 & c \le x \end{cases}$$

Where $(x;a,b,c)$ is the parameter set that changes the shapes of the MFs. Parameters in this layer are referred to as the premise parameters.

Layer 2: The second layer in the ANFIS network is the rule layer. In this layer, the membership functions are the input values and each node multiplies the input and provides an output that reflects the rule's firing power. This layer's output is given in the Eq. (3).

$$O_{2,i} = w_i = \mu_{A_i}(x)\ \ \mu_{B_i}(y) \qquad i = 1,2 \tag{3}$$

Layer 3: Here the *i*-th node is calculated by the ratio of the *i*-th rules firing strength to the sum of the rule's firing strengths.

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2} \tag{4}$$

Where $\overline{w_i}$ is referred to as the normalized firing strength.

Layer 4: The total output is calculated in this layer as the summation of all incoming signals given in the Eq. (5).

$$O_{4,i} = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i) \tag{5}$$

Where $\overline{w_i}$ is the output of layer 3, and $(p_i x + q_i y + r_i)$ is the parameter set. Parameters in this layer are mentioned to as the consequent parameters.

Layer 5: The single node in this layer calculates the total output as the number of all input signals, which is stated as

$$O_{5,i} = \sum_i \overline{w_i} f_i = \sum_i \frac{w_i f_i}{w_i} \quad (6)$$

b. Hybrid-learning Algorithm

The ANFIS has two classes of tuning criteria, the first layer called the premise and consequent parameters in the fourth layer [25]. Throughout the learning process, these parameters are attuned until the wanted response of the FIS is accomplished. In this paper, FIS was trained using the hybrid learning algorithm, where it combines the backpropagation (BP) algorithm and the least square method (LSM) [25], The ANFIS output can be written as:

$$f = \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 = \overline{w_1} f_1 + \overline{w_2} f_2 = (\overline{w_1} x)p_1 + (\overline{w_1} y)q_1 + (\overline{w_1})r_1 + (\overline{w_2} x)p_2 + (\overline{w_2} y)q_2 + (\overline{w_2})r_2 \quad (7)$$

4.1.2. Shapely values review

The Shapley Additive explanations (SHAP) approach is founded on the Shapley value principle from game theory [26,27]. The Shapley value (SHAP) principle was initially established to predict the significance of an individual player in a cooperative team. This conception is designed to distribute the total benefit or payoff between players and is based on the comparative prominence of their contributions to the result of a game. Shapley values deliver a solution to each player's allocation of a fair or rational reward and represent a distinctive result characterized by the following natural characteristics or axioms: local precision (additivity), consistency (symmetry) and non-existence (null effect) [27].

Features that contribute to the production or prediction of the model with distinct magnitudes and signs are accounted for by Shapley values. Therefore, Shapley values reflect estimates of the significance of the function (magnitude of the contribution) as well as the direction (sign). Features in this study represent the hematological parameters in the collected blood tests where positive SHAP values indicate that the feature is "helping" the positive class (i.e., pushing the prediction to be "infected"), whereas negative SHAP values indicate that the feature is pushing the prediction to be "not-infected". In specific, the importance of a feature i is defined by the Shapley value in Eq. (8)

$$\emptyset_i = \frac{1}{|N|!} \sum_{SCN\backslash\{i\}} |S|!\,(|N| - |S| - 1)! \left[f\left(S \cup \{i\} - f(S)\right)\right] \quad (8)$$

Here f (S) relates to the ANFIS model output which can be clarified by a set of features S, and N is the whole set of all features. The ultimate contribution of Shapley value of feature $i\emptyset_i$ is determined as the middling of its contributions across all possible permutations of a feature set. Features are then individually applied to the collection and their importance is exposed by the shift in model performance. Importantly, this formalism takes into account feature orderings that affect the observed modifications in the performance of a model in the presence of correlated features.

4.2 Proposed approach

In this paper, we defined a machine learning method as a computational approach for the rapid detection of COVID-19 patients that learns from a predefined bag of features and goes through several different steps. These steps are illustrated in Fig. 2 which consisted of the following stages: (1) Dataset preparation, (2) Feature Selection, (3) Classification, and (4) Interpreting Model Predictions.

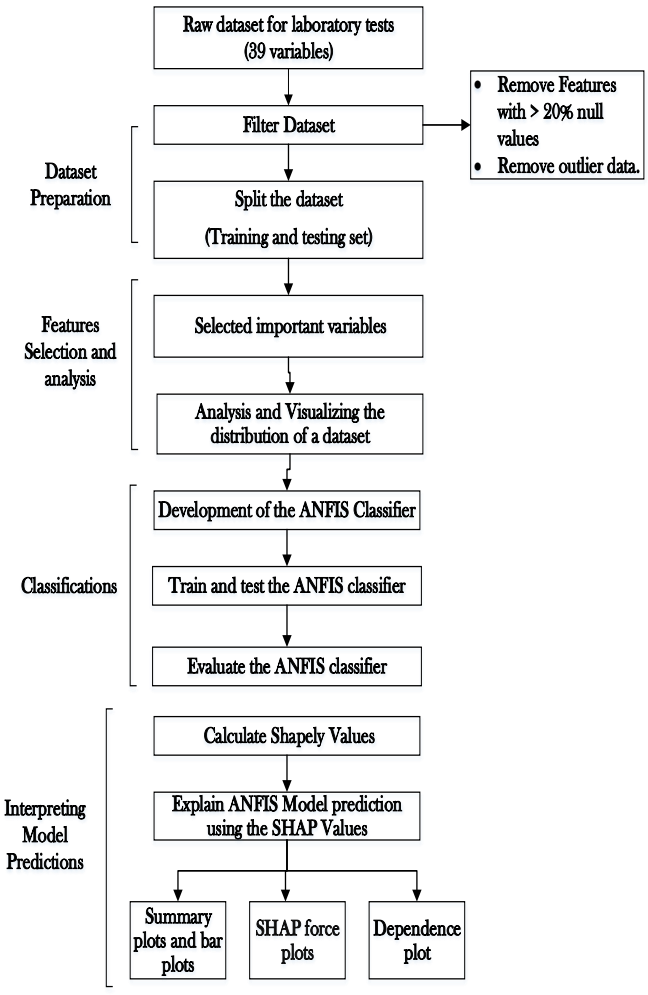


Figure. 2 Overall proposed approach steps

4.2.1. Filtering the dataset and splitting:

The main purpose of filtering the dataset is to remove the artifacts (variables that have many missing data points and outlier data) and the splitting step is to provide the classifier model with an accurate dataset and also to split the dataset into two parts. One part is used for training the classifier while the other part is used for the final evaluation of the classifier. Filtering the dataset and splitting steps are shown in the scheme:

a. All information was anonymized in compliance with the current international practices and guidelines.
b. In order to have a mean of zero and a unit standard deviation, all clinical data was standardized.
c. It was observed from the data that most of the variables have a very high percentage of missing values. Hence, variables that had too many missing data points (> 20 %) were removed from the dataset.
d. Samples that were too sparse (outlier data) in laboratory data were also removed.
e. We chose to keep negative samples that have at least 10 variables with data points available. This is performed to avoid an overfit scenario where a few samples (sparse but positive) may have an undue influence on the predictive model.
f. Splitting the dataset into 80% as the training set, 20 % as the testing set

4.2.2. Selected features

After filtering the dataset as previously mentioned in the latter section, the data set contained 16 variables which included WBC count, Platelet Count, Patient age, HCT, Hgb, MPV, RBC Count, Basophils count, Absolute Eosinophil Count, Lymphocyte Count, MCHC, MCH, MCV, Absolute Monocyte Count, RDW and the presence of chronic disease. The selected variables or features represent the hematologic parameters that are present in a complete blood cell analysis of suspected patients. The target outcome variable indicates whether the patient is tested positive or negative for COVID 19.

4.2.3. Analysis and visualizing the distribution of a dataset

The most essential step before developing the classifier is the analysis and understanding of the relations between each selected feature and target (0 or 1). Creating a histogram provides a visual representation of data distribution.

4.2.4. Development of a classification model

The classification model for the COVID-19 diagnosis is achieved by the Adaptive Neuro-Fuzzy Inference System (ANFIS). The classification schemes employed in our experiments include the following steps:

a. ANFIS Classifier Design

For the classification of COVID-19 patients, a multilayer ANFIS classifier consisting of an input layer, three hidden layers, and an output layer is used in this article. A number of neurons are installed on the input layer of the ANFIS classifier, which is proportional to the size of the features extracted. The hidden layer is checked with a different number of neurons and, eventually, four hidden layers are designed to achieve the optimum rate of classification. Four hidden layers and each hidden layer with 16 neurons are configured in this article. The output layer consists of a single neuron and thus produces a single binary output 0 (negative) or 1 (positive). Training the ANFIS model was accomplished using the Hybrid- Learning algorithm.

b. Model performance

We employed the selected features to train the ANFIS classifier then the classifier performance was evaluated in terms of Sensitivity, Specificity, and Accuracy. These terms are calculated as shown in Eqs. (9) -(11). Accuracy is the probability that the test will deliver precise outcomes, that is, negative in healthy patients and positive in infected patients. Hence, it is the probability of the true positives and true negatives between all the results. The sensitivity is the rate of true positives and determines the classifier's ability to detect correctly people with COVID-19. Specificity is the capability of categorizing healthy patients as negatives. It is the rate of accurate negatives. The explanation of the terms of TP, FP, TN, and FN are shown in Table 1.

$$Sensitivity = \frac{TP}{TP+FN} \times 100\% \quad (9)$$

$$Specifity = \frac{TN}{FP+TN} \times 100\% \quad (10)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (11)$$

In addition, To evaluate our prediction ANFIS model classifier, other state of Art classifiers KNN (k-Nearest Neighbor), and RF (Random Forest) were implemented and compared them with the proposed model.

Table 1. Definitions of terms used for measuring the ANFIS model performance

Term	Meaning
True Positives (TP)	The number of positively tested COVID-19 cases that are properly categorized as positive cases.
False Positives (FP)	The number of positively tested COVID-19 cases that are wrongly categorized as negative cases.
True Negatives (TN)	The number of negatively tested COVID-19 cases that are properly categorized as negative cases.
False negatives (FN)	The number of negatively tested COVID-19 cases that are wrongly categorized as positive cases.

4.2.5. Interpreting the ANFIS Model prediction using the SHAP Values

Model explanations are important because they can be employed to improve medical decision-making and guide policy-making initiatives. In this study, the SHAP method was employed to comprehend the significance of the various clinical variables and its effect on the model output hence pointing out the best indicators for predicting patients that are infected with the COVID-19 disease. A main benefit of the associated SHAP values is that they add interpretability to complicated models such as ANFIS. The model interpretability was evaluated both locally and globally by looking at the relative importance of the variables and their impact on the ANFIS model's prediction. Global interpretability helps to understand the entire structure of the model and it can be obtained through summary plots and bar plots that show the global importance of the features. The SHAP summary plot shows how much each predictor contributes, either positively or negatively, to the target outcome variable (whether the patient is tested positive or negative for COVID 19). Features are organized by the sum of the magnitudes of the SHAP values in all the samples, i.e. by their global impact $\sum_{j=1}^{N}\left|\emptyset_i{}^{(j)}\right|$. SHAP values $\emptyset_i{}^{(j)}$ are drawn horizontally. Local explanations focus on explaining each prediction and it can be accomplished through SHAP force plots. Force plots show how features contribute to pushing the ANFIS prediction output from the base value to the model output. The output value of the model is the prediction for that observation. The base value is the average model output over the training dataset. The prediction starts from the base value.

5. Type-style and fonts

In the first part of this section, Analysis and Visualizing of the distribution of the dataset will be discussed to understand the relations between the selected features and the target outcome (1 for positively tested COVID-19 patients and 0 for negatively tested patients). Secondly, an evaluation of the ANFIS model performance will be presented to investigate the prediction ability of the proposed model classifier. Interpreting the ANFIS model will be further discussed using SHAP values.

The distribution curves for the selected features with target values are shown in Fig. 3. It was observed that several features had pronounced effects on the diagnosis of the COVID-19 disease while others had no significant effect. It was clear that Patient age Quantile, Platelet count, WBC count, Basophils count, Absolute Eosinophil count, Monocytes, and the existence of chronic disease had a pronounced effect on the diagnosis of COVID-19 patients where other values of other parameters like RBC count, HCT, Hgb, MPV, MCHC, Lymph count and RDW counts were not significantly altered between positively and negatively tested patients with COVID-19.

To evaluate the ANFIS model performance, a confusion matrix was developed and shown in Fig. 4 to show the predicted values produced from the model. The testing data set had a total of 121 cases; it was obvious from the confusion matrix that the trained ANFIS classifier could correctly identify 9 cases as positively tested COVID-19 patients (TP) and 106 cases are correctly identified as negatively tested COVID-19 patients (TN). 3 Positively tested cases were wrongly identified as negatively tested cases (FP) and 3 negatively tested cases were wrongly identified as positively tested (FN).

Table 2 shows the computed Accuracy, Sensitivity, and Specificity for testing data of the ANFIS model classifier. The results revealed that the ANFIS model could classify between positive and negative COVID-19 patients by achieved Accuracy, Sensitivity, and Specificity rates of 95%, 75%, and 97.25% respectively. It is also noticeable that the ANFIS classifier attained a significantly higher accuracy than the corresponding counterpart classifiers.

Table 2. Comparison between ANFIS model and state of art methods

Performance Metrics	**ANFIS**	**KNN**	**RF**
Accuracy	0.95	0.892	0.90
Sensitivity	0.75	0.895	0.71
Specificity	0.97	0.882	0.91

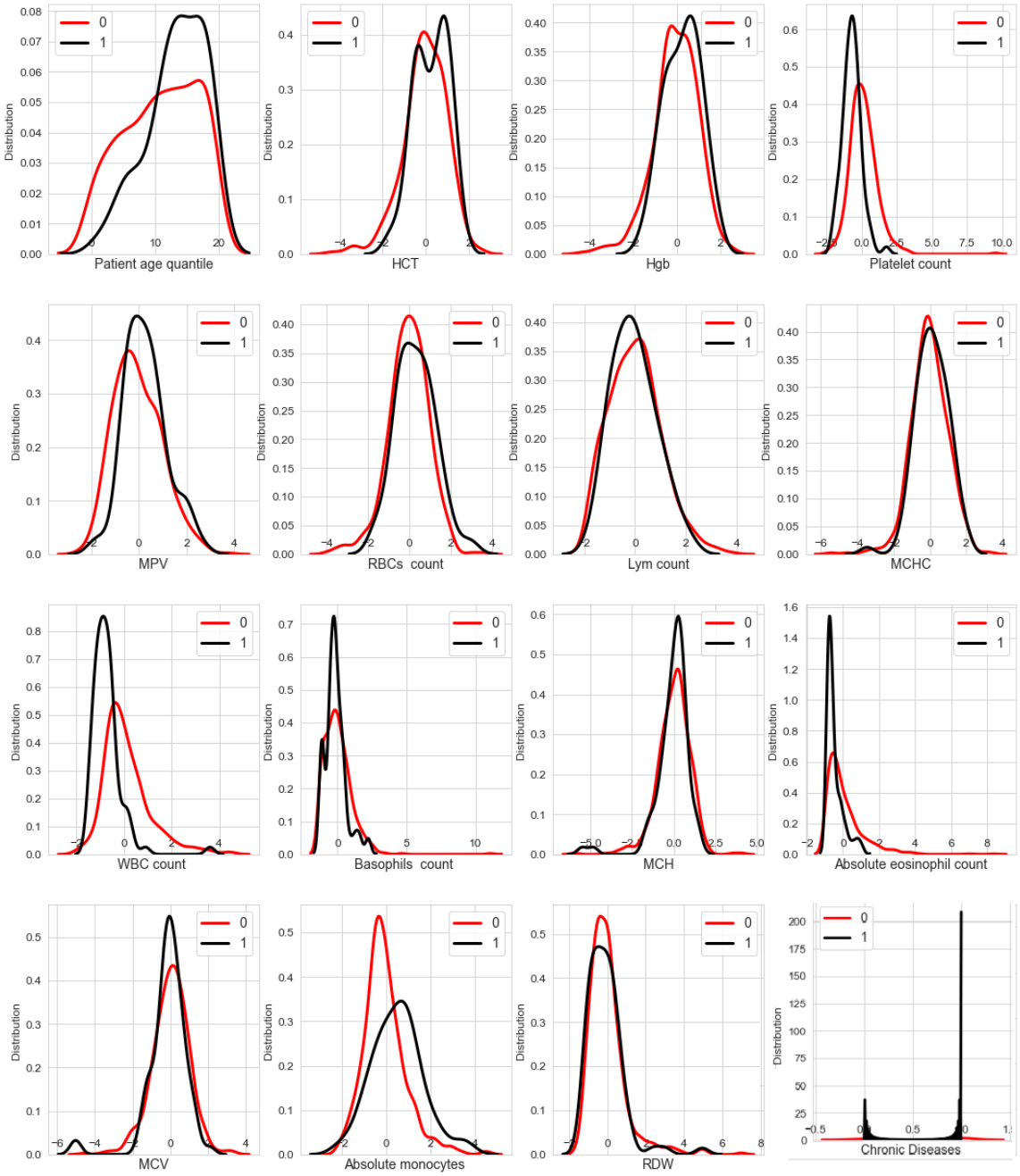


Figure. 3 Distribution curves for selected features with target values

In general, the high accuracy of the ANFIS classifier reveals its high ability of classification hence providing a useful and efficient diagnosis of COVID-19 cases using commonly available laboratory blood test data. The high specificity achieved by the model indicates high true negative rates due to the high number of negatively tested COVID-19 patients that are available in the dataset which can help to train the classifier to distinguish the healthy cases. However, the low sensitivity of the ANFIS classifier (75%) means that there are high false-negative results and thus reducing the ability of the classifier to discern between the cases of the disease. This can be attributed to the fact that 10% only of the dataset was used to train the classifier on positive cases of COVID 19. Accordingly, to improve the performance of the classifier we should increase the number of patients and therefore can help in discriminating between positive and negative COVID-19 patients.

To Interpret the ANFIS model and to show the relative importance of each feature and its effect on the predicted diagnoses of patients, an aggregate bar graph was performed and shown in Fig. 5. The bar graph plots the mean absolute SHAP value for each feature. Furthermore, a SHAP summary plot shown in Fig. 6 was developed that provided more context than the bar chart and also their range of effects over

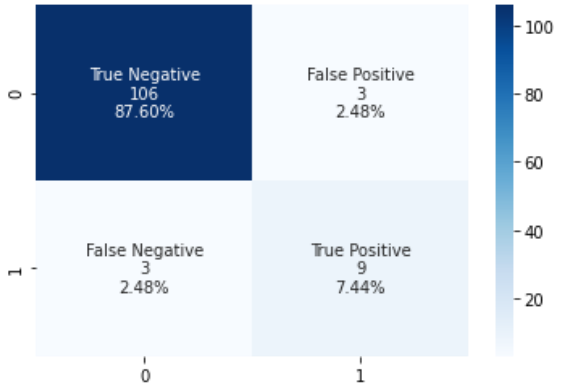


Figure. 4 Confusion matrix of the ANFIS classifier

the dataset. Each point in the SHAP summary plot represents a row of the data set. This is like the bar plot but it can show the positive or negative relationship for each variable with the target. Features are sorted descendingly according to their importance. The horizontal location in the SHAP summary plot shows whether the effect of that value is associated with a higher or lower prediction.

The x-axis points indicate the effect of the feature on the estimation of a specific patient. Color refers to either high (red) or low (blue) relative variables. Positive SHAP values indicate that the model predicted positively tested cases with COVID-19 while a negative SHAP value indicates negatively tested patients. SHAP values farther away from zero means a bigger impact for a certain feature.

It was noticed from Figs. 5 and 6 that the top five most important hematological parameters that had a significant effect on the ANFIS model's prediction are the WBC count, Platelet count, existence of the chronic disease, Basophils count & patient age quantile. It was also clear that WBC and platelet count had the maximum impact of the prediction where low values of these two variables are associated with clear positive impacts on the model prediction. This indicates that patients having low values of WBC and platelet counts are most likely of being infected with COVID-19. This result can lead us to conclude that there is a significant association between low WBC and platelet counts and COVID-19 patients which is similar to previous findings [17], [19], and [20]. The interrelation between low WBC count and COVID-19 cases was also proved by other previous studies [4], [14] and [18]. Xiaofang Zhao et al. [21] also found that that there was a decrease in the platelet count of COVID-19 patients in the early stages of the disease and that the progression of the disease is related to the risk of lowering the platelet count.

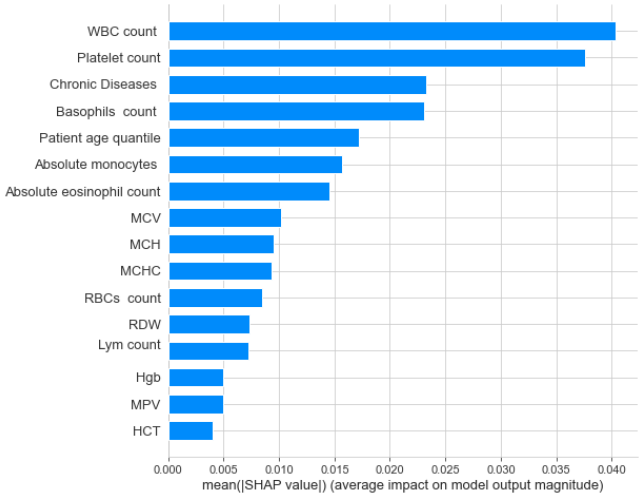


Figure. 5 Bar plot to show the global importance for each feature

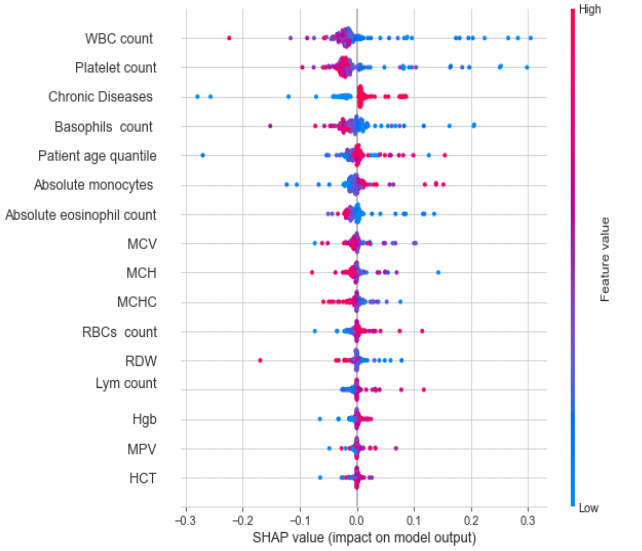


Figure. 6 SHAP summary plot of the ANFIS model

In this study, a possible explanation of the low WBC and Platelet counts in COVID-19 patients is that the erosion of the virus caused excessive destruction of white blood cells leading to a decrease in the number of white blood cells in the peripheral blood of the patients. However, with the progression of the virus and the severity of the patient's conditions, it could lead to an increase in the white blood cells. Furthermore, the virus infection may bring about immune damages to platelets by inducing auto-antibodies and immune complexes.

Another relevant observation from the SHAP summary plot that the presence of chronic disease had a positive impact on the ANFIS prediction target which implies that patients having a chronic disease or illness are more likely to be infected with COVID-19 which can be due to the disorders in the immune system of the body caused by the presence of the chronic disease and the lower ability to resist any

other infections or viruses. As also observed from Fig. 4 that low values of Basophils and Eosinophil counts had a positive impact on the target variable where low values of theses magnitudes are expected to be present in COVID-19 patients. This finding is similar to other previous studies[16,20,28] . On the other hand, Patient age quantile did not show a clear effect on the prediction of the model so in our study it is not considered as a marker for diagnosis of the COVID-19 disease. This bias can be due to the limited number of positively tested COVID-19 patients in the dataset. Moreover, the plot showed that an increase in the monocyte count had positive SHAP values which meant that the model predicted positively tested cases having an increased monocyte count. This result is similar to another study, which showed that COVID-19 patients had a greater number of monocytes than healthy people, but still within the normal range. [28]. In regards to the seriousness of the disease, pro inflammatory monocyte activation, particularly for elderly people on early diagnosis have also been shown to be linked to the severe condition [29]. Activated monocytes are therefore not generally available for routine analyzers in the form of parameters.

In conclusion, as observed from Figs. 5 and 6 that cases having low WBC, Platelet, Basophils, and Eosinophil counts and also having a state of illness or chronic disease are more likely to be infected with the COVID-19 virus. Since the WBC and Platelets count had the maximum impact on the ANFIS model predictions; therefore WBC & platelet counts can act as relevant and significant indicators for the diagnosis of COVID-19 patients.

In order to get a deeper insight into the interactions between variables, we further developed SHAP dependence plots. They show the marginal effect that one or two variables have on the predicted outcome of the proposed ANFIS model. Showing how the model output changes as the features change helps us to explain how the model depends on that feature. They plot a feature's value versus the SHAP value of that feature across various samples.

Each dot is a single forecast (row) from the dataset. The value of the feature appear on the x-axis and the y-axis is the SHAP value for that feature, which characterizes how much the feature's value changes the output of the model for that sample's prediction. The color relates to a second feature that may have a collaboration effect with the feature that has been plotted. The second feature is selected automatically.

Fig. 7 shows dependence plots between different variables and their effect on the ANFIS predicted outcome. Fig. 7 (a) shows the SHAP values of WBC count where Platelet count was chosen to show the interaction between them. It was clear that there was a downward trend between the WBC count and the target variable. The lower the WBC count the higher the ANFIS model prediction for the existence of COVID-19.

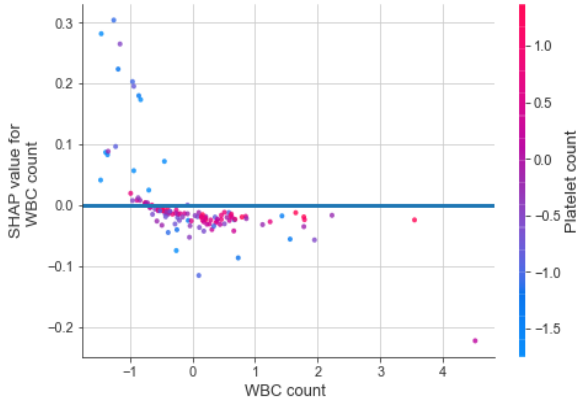
Moreover, low WBC counts accompanied by low Platelet counts increased the risk of being infected with the virus. SHAP values for the existence of the chronic disease in suspected patients were plotted in Fig. 7 (b) where a value of 1 and 0 indicates the existence and non-existence of chronic disease. It was obvious that the presence of a chronic disease or a state of illness in patients showed positive SHAP values indicating higher probabilities of infection with COVID-19.

Fig. 7 (c) and 7 (d) revealed that low values of Basophil and Eosinophil counts had a higher impact on the prediction along with low values of WBC count increases the risk of COVID-19. Furthermore, the appearance of monocytes in the blood tests of patients as shown in Fig. 7 (e) were associated with the high prediction of the ANFIS model and thereby increased the probability of COVID-19.

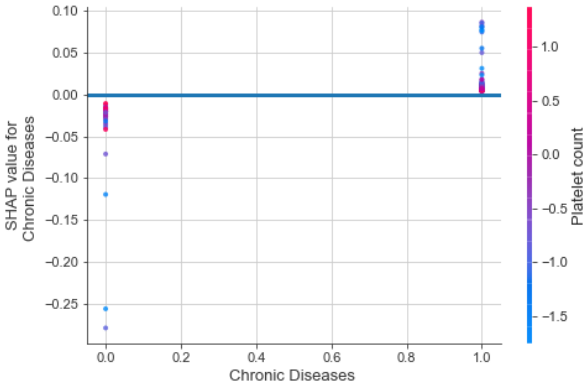
To evaluate the model interpretability locally, SHAP force plots were performed for two observations or in other words two cases. In the plot, each Shapley value is an arrow that pushes to increase or decrease the prediction. Feature values that push the model towards a higher prediction of 1 (infected with COVID-19) are in red, and those reducing the prediction to 0 (not infected with COVID-19) are in blue and the length of the region shows how much the feature contributes to this effect. These plots explain why the model output takes a given value for each observation and it can also determine the observations where a certain variable or a set of variables have a greater or lower impact on the model's prediction. Figs. 8 and 9 show the SHAP force plots for two patients from the data set.

The force plot shown in Fig. 8 reveals that for this particular patient, the ANFIS model output has a high prediction of 1 which is higher than the base value (0.08264). This indicates that this patient is infected with COVID-19. In this observation, features like Basophil count, Absolute Monocytes, and the existence of chronic disease pushed the model prediction to the right towards a value of 1 which increased the patient's predicted probability of being infected with COVID-19. Since in this study, the patient age wasn't considered as an indicator to differentiate between positive and negative cases of the virus; therefore it is not taken into account.

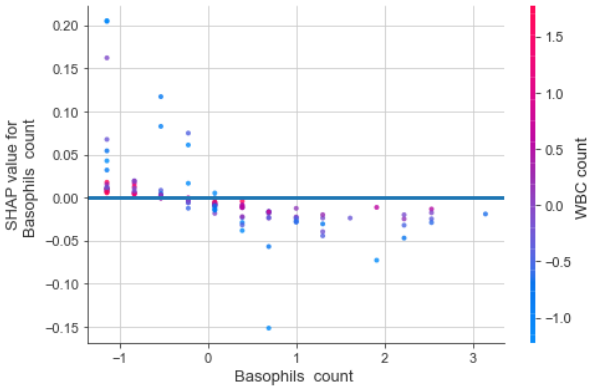
Fig.9 shows the force plot of the second patient where the ANFIS output prediction is 0 which means



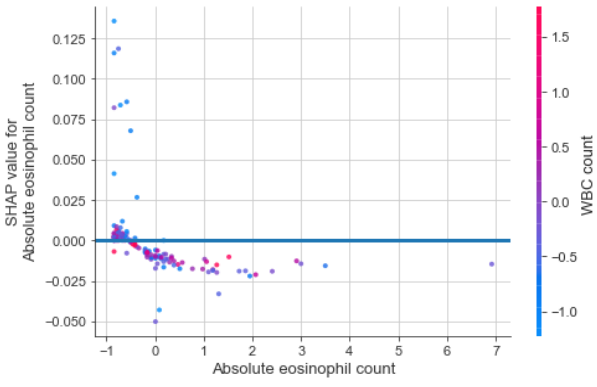
(a)



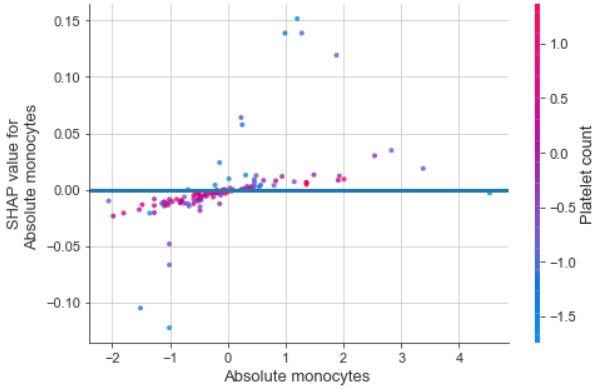
(b)



(c)



(d)



(e)

Figure. 7 Dependence plot that shows SHAP values of: (a) WBC count, (b) chronic disease, (c) basophils count, (d) absolute eosinophil count and (e) absolute monocytes

that the model classifies this observation as a negatively tested COVID-19 patient. As observed from the plot that the base value is higher than the model output. Features increasing the prediction of being infected with COVID-19 like the Basophil counts and the existence of chronic disease are offset by forces driving the prediction lower to a value of 0.

The biggest impact that drives the prediction lower as seen from the plot is Platelet count and WBC count.

6. Conclusion

In this paper, an Adaptive Neuro-Fuzzy Inference System (ANFIS) approach was proposed to rapidly detect COVID-19 cases using commonly available laboratory blood test data. The use of blood tests has the advantages of being less expensive and less time consuming than other diagnosis methods and thus providing a more accessible system. Hence, the present study aims to provide an efficient and reliable predicting model to analyze blood tests and identify significant hematological indicators to support COVID-19 diagnosis. The study was applied to the results created by the Hospital Israelita Albert

Einstein in São Paulo, Brazil. The ANFIS model classifier had the capability of discriminating between positive and negative COVID-19 patients with an Accuracy, Sensitivity, and Specificity rates of 95%, 75%, and 97.25% respectively although 10% only of the data was positive for COVID 19.

The ANFIS model performance was interpreted by shapely values in order to identify the most important and decisive hematological parameters that could assist clinicians in making effective patient management decisions. The results revealed that patients having low WBC and Platelet counts are more likely to be infected with COVID-19 patients

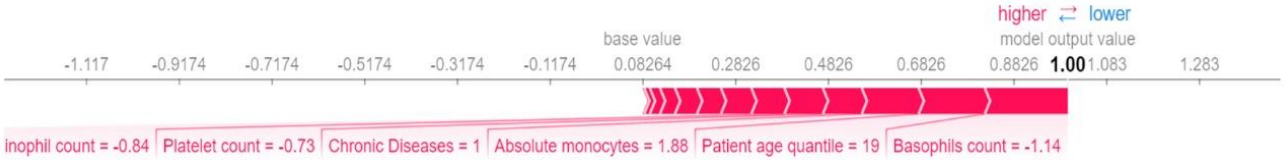


Figure. 8 SHAP force plot to show the prediction of the ANFIS model for the 10t^{h} observation

Figure. 9 SHAP force plot to show the prediction of the ANFIS model for the 50th observation

and that they had the maximum impact on the ANFIS model predictions. Hence, it was concluded that WBC & platelet counts could act as relevant and significant indicators for the diagnosis of COVID-19 patients. The high classification ability of the ANFIS model proves that it is a useful, efficient, and rapid diagnosis system of COVID-19. Therefore, a stable, reliable, and readily available method for diagnosing COVID-19 can be provided by combining available blood test results with analysis based on the proposed ANFIS model.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Mohanad Deif, Ahmed Solyman and Rania Hammam; methodology, Mohanad Deif; software, Mohanad Deif; validation, Rania Hammam; resources, Mohanad Deif; writing—original draft preparation, Mohanad Deif; writing—review and editing, Ahmed Solyman; visualization, Rania Hammam; supervision, Ahmed Solyman.

References

[1] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, E. H. Y. Lau, and J. Y. Wong, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia", *N. Engl. J. Med.*, 2020.

[2] J. A. Backer, D. Klinkenberg, and J. Wallinga, "Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020", *Eurosurveillance*, Vol. 25, No. 5, p. 2000062, 2020.

[3] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, and Y. Wei, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study", *Lancet*, Vol. 395, No. 10223, pp. 507–513, 2020.

[4] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, and X. Gu, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China", *Lancet*, Vol. 395, No. 10223, pp. 497–506, 2020.

[5] L. Peñarrubia, M. Ruiz, R. Porco, S. N. Rao, M. Juanola-Falgarona, D. Manissero, M. López-Fontanals, and J. Pareja, "Multiple assays in a real-time RT-PCR SARS-CoV-2 panel can mitigate the risk of loss of sensitivity by new genomic variants during the COVID-19 outbreak", *Int. J. Infect. Dis.*, 2020.

[6] Z. Li, Y. Yi, X. Luo, *et al.*, "Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis", *J. Med. Virol.*, p. 10.1002/jmv.25727, 2020, doi: 10.1002/jmv.25727.

[7] G. Lippi, A.-M. Simundic, and M. Plebani, "Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19)", *Clin. Chem. Lab. Med.*, Vol. 1, No. ahead-of-print, 2020.

[8] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, and J. Liu, "Chest CT for Typical Coronavirus Disease 2019 (COVID-19) Pneumonia: Relationship to Negative RT-PCR Testing.", *Radiology*, Vol. 296, No. 2, pp. E41–E45, 2020, doi: 10.1148/radiol.2020200343.

[9] T. Ai, Z. Yang, and H. Hou, "Correlation of chest CT and RT-PCR testing in coronavirus Disease 2019 (COVID-19) in China: a report of 1014 cases [e-pub ahead of print]. Radiology." .

[10] A. I. L. D. Burog, C. P. R. C. Yacapin, R. R. O. Maglente, A. A. Macalalad-Josue, E. J. B. Uy, A. L. Dans, and L. F. Dans, "Should IgM/IgG rapid

test kit be used in the diagnosis of COVID-19?," *Asia Pacific Cent. Evid. Based Healthc.*, Vol. 4, pp. 1–12, 2020.

[11] R. Patel, E. Babady, E. S. Theel, G. A. Storch, B. A. Pinsky, K. S. George, T. C. Smith, and S. Bertuzzi, "Report from the American Society for Microbiology COVID-19 international summit, 23 march 2020: value of diagnostic testing for SARS–CoV-2/COVID-19", Am Soc Microbiol, 2020.

[12] M. Döhla, C. Boesecke, B. Schulte, C. Diegmann, E. Sib, E. Richter, M. Eschbach-Bludau, S. Aldabbagh, B. Marx, and A.-M. Eis-Hübinger, "Rapid point-of-care testing for SARS-CoV-2 in a community screening setting shows low sensitivity", *Public Health*, 2020.

[13] W. Egner, D. C. C. Beck, G. Davis, C. Dayan, T. El-shanawany, E. Griffiths, D. Kumararatne, L. Sandle, R. Sargur, and S. Savić, "Statement from RCPath's immunology specialty advisory committee on COVID-19/SARS CoV2 antibody evaluation", *Inst. Clin. Epidemiol. Natl. Institutes Heal. Manila Asia-Pacific Cent. Evid. Based Healthc. Inc*, 2020.

[14] D. Ferrari, A. Motta, M. Strollo, G. Banfi, and M. Locatelli, "Routine blood tests as a potential diagnostic tool for COVID-19", *Clin. Chem. Lab. Med.*, Vol. 1, No. ahead-of-print, 2020.

[15] T. A. Khartabil, H. Russcher, A. van der Ven, and Y. B. de Rijke, "A summary of the diagnostic and prognostic value of hemocytometry markers in COVID-19 patients", *Crit. Rev. Clin. Lab. Sci.*, pp. 1–17, 2020.

[16] Z. Cheng, Y. Lu, Q. Cao, L. Qin, Z. Pan, F. Yan, and W. Yang, "Clinical features and chest CT manifestations of coronavirus disease 2019 (COVID-19) in a single-center study in Shanghai, China", *Am. J. Roentgenol.*, pp. 1–6, 2020.

[17] B. E. Fan, V. C. L. Chong, S. S. W. Chan, G. H. Lim, K. G. E. Lim, G. B. Tan, S. S. Mucheli, P. Kuperan, and K. H. Ong, "Hematologic parameters in patients with COVID-19 infection", *Am. J. Hematol.*, Vol. 95, No. 6, pp. E131–E134, 2020.

[18] W. Guan, Z. Ni, Y. Hu, W. Liang, C. Ou, J. He, L. Liu, H. Shan, C. Lei, and D. S. C. Hui, "Clinical characteristics of coronavirus disease 2019 in China", *N. Engl. J. Med.*, Vol. 382, No. 18, pp. 1708–1720, 2020.

[19] H. Yun, Z. Sun, J. Wu, A. Tang, M. Hu, and Z. Xiang, "Laboratory data analysis of novel coronavirus (COVID-19) screening in 2510 patients", *Clin. Chim. Acta*, 2020.

[20] X. Zhao, K. Wang, P. Zuo, Y. Liu, M. Zhang, S. Xie, H. Zhang, X. Chen, and C. Liu, "Early decrease in blood platelet count is associated with poor prognosis in COVID-19 patients—indications for predictive, preventive, and personalized medical approach", *EPMA J.*, p. 1, 2020.

[21] G. Gunčar, M. Kukar, M. Notar, M. Brvar, P. Černelč, M. Notar, and M. Notar, "An application of machine learning to haematological diagnosis", *Sci. Rep.*, Vol. 8, No. 1, pp. 1–12, 2018.

[22] V. A. de F. Barbosa, J. C. Gomes, M. A. de Santana, J. E. de A. Albuquerque, R. G. de Souza, R. E. de Souza, and W. P. dos Santos, "Heg.IA: An intelligent system to support diagnosis of Covid-19 based on blood tests", *medRxiv*, p. 2020.05.14.20102533, 2020, doi: 10.1101/2020.05.14.20102533.

[23] "Diagnosis of COVID-19 and its clinical spectrum. Kaggleitle", *Kaggle*, 2020. https://www.kaggle.com/einsteindata4u/covid19.

[24] R. Jyh-Shing and Jang, "ANFIS: adaptive-network-based fuzzy inference system", *Comput. methods programs Biomed. IEEE Trans. Syst.*, Vol. 23, No. 3, pp. 665–685, 1993, doi: 10.1109/21.256541.

[25] R. Rodríguez-Pérez and J. Bajorath, "Interpretation of compound activity predictions from complex machine learning models using local approximations and shapley values", *J. Med. Chem.*, 2019.

[26] L. S. Shapley, "A Value for n-Person Games, Contributions to the Theory of Games (Kuhn, HW, Tucker, AW, eds.), 307-317", *Ann. Math. Stud.*, Vol. 28, pp. 275–293, 1953.

[27] G. Lippi and M. Plebani, "Laboratory abnormalities in patients with COVID-2019 infection", *Clin. Chem. Lab. Med.*, Vol. 58, No. 7, pp. 1131–1134, 2020.

[28] H. Yu, D. Li, Z. Deng, Z. Yang, J. Cai, L. Jiang, K. Wang, J. Wang, W. Zhou, and X. Wei, "Total protein as a biomarker for predicting coronavirus disease-2019 pneumonia", 2020.

[29] Y. Lin, C. Ji, W. Weng, P. Xu, Y. Hu, W. Liang, J. Pu, and H. Zhang, "Epidemiological and Clinical Characteristics of 124 Elderly Outpatients with COVID-19 in Wuhan, China", *China (2/20/2020)*, 2020.