

Data Collected and Analysis of COVID-19 Infection Status: Case Study of Iraqi Hospital Patients in Diwaniyah and Najaf Governorates

Haitham Abdulaaïma Aswad 

Informatics Institute for Postgraduate Studies, Iraqi Commission for Computer and Informatics,
Baghdad, Iraq

ms202130667@iips.icci.edu.iq

Ali Hasan Taresh

Informatics Institute for Postgraduate Studies, Iraqi Commission for Computer and Informatics,
Baghdad, Iraq

Abstract— In recent years, the coronavirus (COVID-19) spread in a dangerous and rapidly. It first appeared in the Chinese city of Wuhan in early December 2019, causing many cases of infection and death because of the rapid spread of the virus throughout the world. The conversion of patient information from paper to electronic helps the medical staff and researchers analyze and retrieve information faster. In this paper, we gathered data about patients infected with COVID-19 who were registered in the Iraqi hospitals of two governorates (Diwaniyah and Najaf) from 2020 to 2023. In addition, we used machine learning algorithms that protect the infection of the data entered for the patient. The highest accuracy (average in both datasets of Diwaniyah and Najaf) are achieved by the both algorithms Nave Bayes (90%) and Decision tree (90%), while the K nearest neighbor (89%), Random forest() and artificial neural network (84).

Keywords— COVID-19 Dataset; Data analysis; Electronic documentation; Machine learning; information retrieval.

1 Introduction

The recent epidemiological development represents an emergency at the national and international levels[1]. The Chinese city of Wuhan announced in early December 2019 that it was infected with the Coronavirus (Covid-19). Due to the population density of Wuhan and the government's inability to impose a complete lockdown, the virus spread throughout China[2]. Moreover, due to trade, tourism, and travel between countries, cases of infection by the virus began to appear in many European countries and then the countries of the Middle East. What caused global horror was the rapid spread of the virus until the World Health Organization was forced to declare it a pandemic on March 11, 2020. This rapid spread of the virus caused panic, fear, and terror among the world's population[1], [3]. It also increased the cases of infection and deaths and revealed the inability of governments and health organizations to contain the virus and prevent its spread[4], [5]. This was due to the weakness in health institutions and their

lack of complete information about the virus, its transmission, the causes of death, and the lack of appropriate drug capacity to treat patients. Figures 1 and 2 illustrate the number of new cases and deaths worldwide[6].

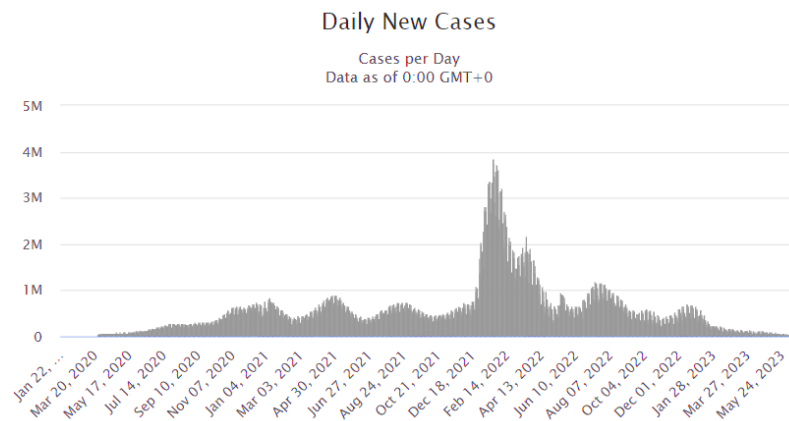


Fig. 1. Total New cases of COVID-19 (daily) over the world[6]

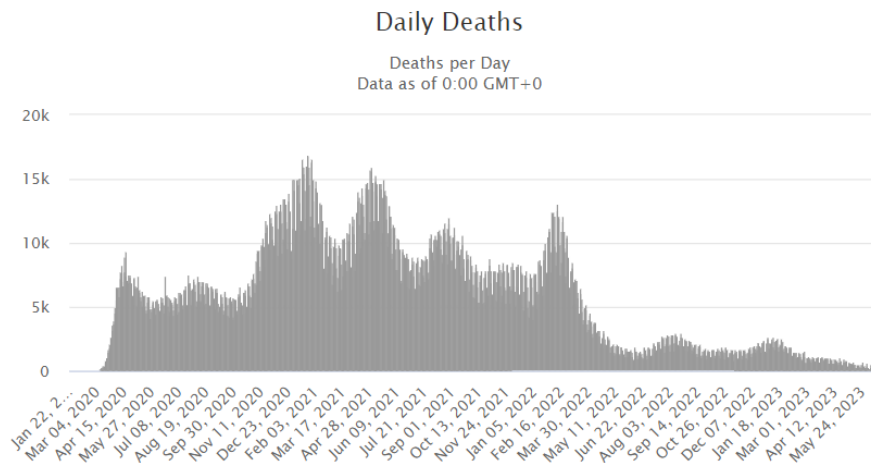


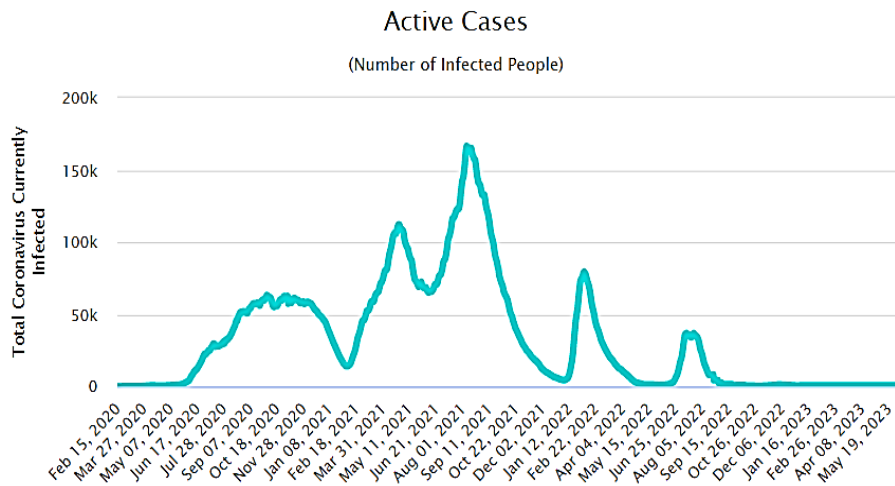
Fig. 2. Total death cases of COVID-19 (daily) over the world[6]

Table 1: shows the number of injuries and deaths of COVID-19 patients in Iraq and the surrounding countries.

Table 1. Table 1: Death and Confirmed Cases in Iraq and surrounding countries[6]

Country	Deaths	Confirmed Cases
Iraq	25,375	2,465,545
Turkey	102,174	17,232,066
Iran	146,289	7,612,234
Kuwait	2,570	665,955
Jordan	14,122	1,746,997
Saudi Arabia	9646	841,469
Syria	3,165	57,638

Figure 3 shows a visual representation of the cumulative number of active coronavirus cases (COVID -19) in Iraq [6]. The graph illustrates the total number of individuals currently infected with the virus over a period. It provides a comprehensive overview of the current prevalence of COVID -19 in Iraq and allows a comparative analysis of the progression of the disease. Using the curve shown in Figure 3, information on the increase, decrease, or stabilization of active cases can be obtained, providing valuable insight into the impact and measures taken to control the virus in Iraq.

**Fig. 3.** Total active cases of COVID-19 in Iraq from 15- Feb 2020 to 19-May 2023[6]

1.1 Motivation

The rapid spread of the virus and the variety of disease cases have led us to present proposals to assist medical personnel. The motive can be summarized as follows.

- The coronavirus (COVID-19) has spread rapidly worldwide and raised great concern. This paper examines and locates the information on COVID-19-infected patients who received treatment in Iraqi hospitals.
- The paper also highlights the benefits of converting patient information from paper to electronic format. Digitizing data allows medical professionals and researchers to access and analyze information quickly.
- This conversion should enable faster retrieval of patient records, better data management, and the application of machine learning algorithms to analyze and predict COVID-19 infections.

1.2 Contributions

The contribution of this paper is to provide insight into the prevalence and impact of COVID-19 in specific regions, advocate for the shift to electronic records, and demonstrate the use of machine learning algorithms to analyze and protect patient data.

- **Data collection:** this work collected data on patients infected with COVID-19. This data collection helps to understand the spread and impact of COVID-19 in these specific regions during the mentioned period.
- **Electronic data conversion:** The conversion allows for faster patient records access, analysis, and management.

1.3 Coronavirus (COVID-19)

The Coronavirus (COVID-19) is a highly contagious virus that has caused a global pandemic[7]. It has resulted in numerous infections and deaths worldwide. The ribosomal structure of the virus consists of four proteins similar to human proteins, which contributed to its spread[8]. These structural proteins protect the virus outside the host cell[3], [9]. When the virus enters the human body, it adheres to the surface of the cell it infects. The viral envelope of the protein attaches to the cell membrane of the infected cell and then enters the cell[10], [11]. In the infected person's body, the virus organizes the inner membrane to create replication organelles through which the virus multiplies. The immune system of the infected person tries to defend the body and fight the virus, but the virus has ways and means to overcome the immune defenses of the human body. The strength of the body's defenses varies from person to person. It was found that some of the infected people did not show any symptoms, although they were able to transmit the virus[12]. This was one of the reasons for the spread of the virus. They are considered one of the most dangerous infected people because they did not show symptoms and unintentionally contributed to the spread of the virus. One of the factors contributing to the inability of healthcare facilities to archive data and patient profiles electronically This prevents researchers from developing systems to help healthcare professionals confront and diagnose the virus early[13].

1.4 Related word

In this section, we will delve into the literature review, which explores various studies that have proposed different models for classification and regression, aiming to develop machine learning models.

The study [14] focused on early detection and control of Covid-19 using the FCNB classification strategy. However, limitations were observed in the preprocessing stage with the FSGA method due to stagnation issues and reliance on random metaheuristic algorithms. The model's applicability was restricted to specific medical data, lacking generalizability. Prakash et al. [15] employed machine learning algorithms to analyze the effects of COVID-19 on various age groups. However, criticisms regarding the study include concerns about the limited factors used for dataset construction, absence of essential medical data, and ambiguity in treatment approach and potential data redundancy affecting algorithm performance. Dubey et al.[16] employed various machine learning algorithms, including linear and logistic regression, decision tree, K-nearest neighbour, and support vector machine, but the study neglected important medical data related to individuals with Covid-19, leading to inconsistent and unreliable outcomes in developing a model to aid healthcare professionals in addressing the epidemic. Singh et al. [17] introduced a Li-MuLi-Poly model utilizing regression techniques to predict COVID-19 mortality rates, but its suitability is limited by a lack of knowledge, introducing uncertainties in its application for predicting COVID-19 deaths.

2 Dataset Coronavirus (COVID-19)

The Dataset on Coronavirus provides valuable information regarding the impact of the virus on a global scale[18]. It includes infection rates, death tolls, and recovery statistics across countries. This dataset serves as a crucial resource for researchers and policymakers in understanding and combating the virus. We gathered data about patients infected with COVID-19 who were registered in the Iraqi hospitals of two governorates (Diwaniyah and Najaf) from 2020 to 2023. The data set was collected from Diwaniyah Governmental Hospital and Najaf Governmental Hospital, affiliated with the Iraqi Ministry of Health, where the patient's profile was recorded and archived electronically for those infected with Covid-19. Data collection was collected for (560 records) of Diwaniyah Hospital and containing 43 features, including (171 deaths, 95 injured, and 294 suspects) and data was collected (from 380 records) for Najaf Hospital

and consisting of 59 attributes, including (88 deaths, 84 injured, and 208 suspected of being infected). Figure 4 shows the scenario of collecting data related to COVID-19.

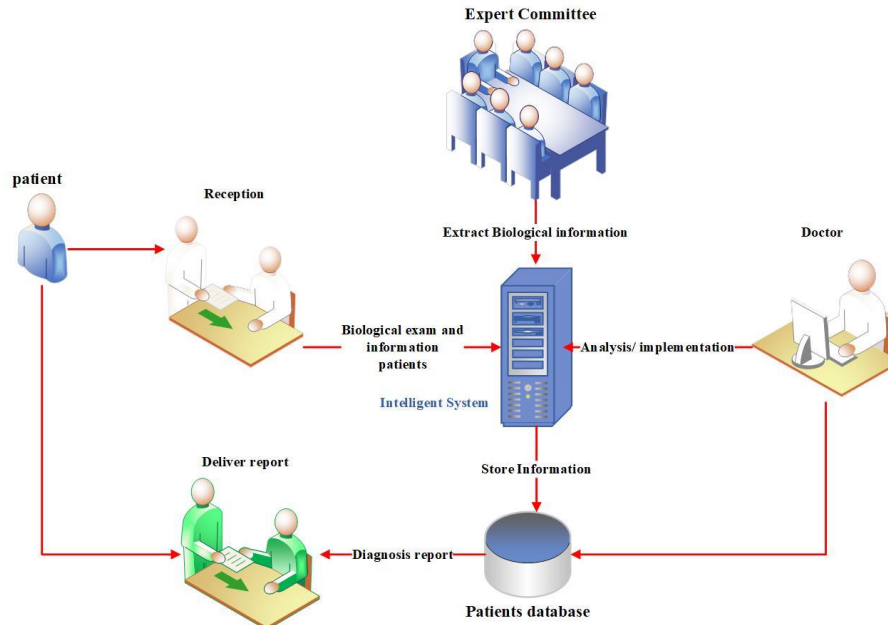


Fig. 4. The scenario of collecting data about COVID-19

Table 2. Table 2 shows the ratio of males to females in the above cases for Al-Diwaniyah and Najaf :

Region	Dead	Infected	Not Infected	Total	No of features
Al-Diwaniyah	171	95	294	560	43
Najaf	88	84	208	380	59

Table 3 illustrates the dataset's features with the description of each element with the data type. In the type column, N refers to the numerical value, whereas the C refers to the Categorical value

Table 3. Table 3 illustrates the features description of dataset

Feature Name	Description	Value type
Age	Age of patient	N
Gender	Gender of the patient (Male, Female)	C
Total days of hospital stay	Total days of hospital stay	N
Diabetes mellitus (DM)	Is a chronic disease that occurs when the pancreas cannot produce enough insulin	N

Random Blood Sugar (R.B.S.) admission	A random blood test to check glucose (sugar) levels	C
Hypertension (H.T)	The blood pressure in the arteries is high in a chronic condition. Blood pressure is made up of two numbers: systolic pressure and diastolic pressure	C
Highest D-dimer	A blood test that can be used to detect the presence of a dangerous blood clot	N
Chronic kidney disease (CKD)	Gradual loss of kidney function, which causes the prevention of water absorption from the urine to reduce its volume and concentration	C
Urea admission	They are natural waste products produced by the human body after eating, produced from the liver's analysis of food protein, and sent to the kidneys to be eliminated through the urine	N
Creat-admission	A chemical compound left over from the energy production processes in your muscles	N
Highest ferritin	Blood protein containing iron	N
Heart- diseases	A group of diseases that affect the heart , including cardiovascular diseases. Arrhythmia. Diseases of congenital heart defects	C
Heart failure (H.F)	Heart failure is a condition in which the heart is unable to pump enough blood for the body's needs	C
Chronic obstructive pulmonary disease (COPD)	Damage to the small air sacs at the tip of the bronchi in the lungs	C
Asthma	It is an inflammatory disease that affects the airways in the lungs, with bronchospasm and obstruction of the airways.	C
Platelets	Colorless blood cells that help blood clot	N
Alanine aminotransferase (A.L.T.)	A transporter enzyme used to diagnose chronic liver disease	N
Highest C.R.P. C-Reactive Protein	It measures the amount of protein produced by the liver in the blood to detect any inflammation in the body	N
hemoglobin (H.B.)	The hemoglobin test measures the level of hemoglobin in the blood	N
S.A.T. on admission	A term referring to the concentration of oxygen in the blood	N

Anti-viral	A factor that kills the virus or limits its ability to reproduce, thus inhibiting its ability to reproduce	C
Antibiotics	A substance or compound that kills or inhibits the growth of germs caused by microorganisms including fungi and parasites	C
Anticoagulants	These are chemicals that prevent or reduce blood clotting	C
Antiplatelet	A drug that prevents platelet aggregation and inhibits clot formation	C
Inotropic	A factor modifies the strength or energy of muscle contractions	C
Anti-inflammatory	It is a group of medicines directed to treat inflammatory conditions and the diseases caused by them	C
Plasma therapy	The convalescent plasma therapy technique involves using the blood of people who have recovered. The blood cells are removed, leaving the liquid (plasma) and antibodies	N
Vitamin	Organic compounds that are free of energy, which the human body needs in very small quantities	C
Dialysis	It is the artificial process through which the blood is purified from waste, toxic substances and excess salts	C
Continuous Positive Airway Pressure (CPAP)	It is a common treatment for obstructive sleep apnea	C
Acute Kidney Injury (A.K.I.)	It is a sudden damage to the kidneys, which leads to a defect in their functions, and may progress to complete kidney failure	C
Respiratory failure	A condition in which the level of oxygen in the blood becomes dangerously low, or the level of carbon dioxide in the blood becomes dangerously high	C
Septic shock	A sharp drop in blood pressure, which leads to the body's organs not getting the oxygen they need in the blood	C
C.K.MB (Creatine Kinase-MB)	Myocardial creatine, meaning creatine kinase and myoglobin binding, is used in the diagnosis of acute angina pectoris	C
Troponin I	It refers to the amount of protein that regulates the contraction of the heart muscle and	C

	skeletal muscles, which in turn prevents muscle contraction in the event of a high calcium deficiency.	
L.D.H. (Lactate Dehydrogenase)	Is the enzyme lactate dehydrogenase that the body uses to convert sugar into energy and is found in the heart, brain and blood cells, and its rise leads to heart problems	N
M.Y.O. (Myoglobin)	It measures the amount of myoglobin present in the skeletal muscles and the heart, and its height indicates damage to the heart muscles	C
T.S.Protein (Total Serum Protein)	It is an analysis used to measure the amount of total protein in the blood (albumin and globulin).	N
Albumin	It is a test used to measure the amount of albumin in the blood	N
GOT (Glutamic-Oxaloacetic Transaminase)	It is one of the analyzes of liver enzymes to determine the efficiency of liver functions. It is used as a carrier of glutamic amine, oxaloacetic acid or aspartate enzyme. Its height indicates the presence of pathological problems in terms of damage to the tissues of the liver or heart.	N
A.L.P. (Alkaline Phosphatase)Or ALK	One of the tests for liver enzymes responsible for the transfer of alkaline phosphatase	N
Bill indirect	It is one of the tests that measure the percentage of indirect bilirubin in the blood	N
Bill direct	It is one of the tests that measures the level of direct bilirubin in the blood	N
HBA1C (glyco-hemoglobin, hemoglobin, A1C or A1c	It is glycosylated hemoglobin and is used to measure cumulative blood sugar for the past three months	N
C. Peptide (connecting peptide)	An analysis used to find out the level of C-peptide in people, to know whether they have type 1 or type 2 diabetes, and to know the amount of insulin secreted by the body	N
S.Insulin (serum Insulin (human))	An analysis used to determine the level of insulin in the blood	N
CHOLI (Cholesterol)	An analysis used to find out the amount of waxy substance and fats in the blood	N

TRIG (Triglycerides)	A blood test that measures the level of triglycerides in the blood	N
P.T (Prothrombin time)	One of the tests used to measure the speed of clotting in the blood	N
PTT (partial thromboplastin time)	One of the tests used to measure the time it takes for blood to clot	N
INR (International normalized ratio)	One of the analyzes used to measure blood fluidity	N
E.S.R (Erythrocyte sedimentation rate)	One of the tests used to find out the presence of rheumatoid arthritis	N
R.F (R.H factor titer or Rheumatoid Factor)	It is a test that determines the percentage of rheumatoid in the blood, which are antibodies made by the immune system to attack germs	C
PCV (Packed cell volume)	It is a test that determines the percentage of red blood cells in the body	N
W.B.C (White blood cells)	It is a test that determines the percentage of white blood cells in the body to determine the rate of inflammation.	N
IL6 (interleukin-6 test)	It is an analysis used to find out the number of damaged cells in the lung	N
Sputum for AFB (Sputum acid fast bacilli)	Sputum analysis of patients to identify patients with pulmonary tuberculosis and pneumonia	C
Ventolin-nebulizer	Bronchodilator treatment	C
Paracetamol - Ibuprofen	A remedy used to lower body temperature	C
Furosemide	A diuretic is used to treat high blood pressure	C
oxygen bottle	Treat as a temporary measure to compensate for hypoxia	C
Anti (Betamethasone , Tocilizumab)	An antiallergic treatment is given to those suffering from rheumatoid arthritis and arteritis	C
(Rose Vastatin , Atorvastatin , Lipid)	A treatment used to treat high fat in vessels and arteries	C

3 Artificial intelligence

Artificial intelligence is a modern technology that simulates human mind functions and is used for complex tasks[19]. It encompasses machine learning and deep learning, utilizing digital systems and advanced technologies[20]. A.I.'s key features include learning, reasoning, and application in various domains such as programming, healthcare, and research centers. It aims to enhance performance by automating processes and tasks[21]. A.I. has become a significant product of the industrial revolution, finding applications in military, economic, and industrial sectors[22], [23]. The focus on A.I. is expected to drive progress and growth in the coming years, enabling companies to achieve substantial profits through solutions that address their challenges. Companies increasingly rely on A.I. programs, leading to reduced labor, fewer errors, improved product quality, and advancements in military research and development[24]. Since its inception in 1956, A.I. has been employed in the medical field to design programs for data analysis, diagnostic tools, and remotely guided surgery devices. Machine learning algorithms have been used to analyze data related to COVID-19, predict infections, identify causes of death, and aid healthcare professionals in disease management[25], [26].

The machine learning algorithms that are used to predict the classification of patients who's infected with Covid 19 are:

3.1 K nearest neighbor (K.N.N.):

It is one of the supervised algorithms used to classify data by identifying random samples and calculating the data closest to it using the law of distance measurement, such as the Euclidean distance and other laws, and it is used to solve classification and regression problems and is ineffective with high-dimensional data[27].

3.2 Decision tree (D.T.):

It is a supervised learning model that is used to solve classification and regression problems, where the data represents the form of a decision tree. Each internal node represents a feature of the data. The algorithm works by dividing the primary node into smaller nodes to provide prediction according to specific entries, which works with categorical and numerical data[28].

3.3 Random forest (R.F.):

It is one of the supervised learning algorithms, which are used to solve the problems of regression and classification, and the mechanism of its work is to train decision trees on random sets of training data and determine the best division in each node to reduce the difference between the attributes to obtain the final prediction and this algorithm is characterized by working with big data and efficiently and with the presence of outliers.[5], [29].

3.4 Naïve bayes (N.B.):

It is one of the supervised machine learning algorithms that predict the category of query data based on the basic concepts of Bayes' theory and is widely used with classification problems, and this algorithm is characterized by simplicity and efficiency with small training data[30].

3.5 Artificial neural network (ANN):

It is one of the algorithms that are called the neural network is inspired by the biological neural networks of humans; where this algorithm is made by dividing the data into layers of nodes for input, hidden, and output, where the input nodes perform mathematical and arithmetic operations on the data to predict the output nodes and are used in machine learning and deep learning[31], [32].

4 Evaluation Metrics

This experiment uses a set of evaluation metrics for performance evaluation based on the confusion matrix, as shown in Figure (5).

		Actual Class	
		Positive (P)	Negative (N)
Predicted Class	Positive (P)	True Positive (TP)	False Positive (FP)
	Negative (N)	False Negative (FN)	True Negative (TN)

Fig. 5. Convulsion Matrix

The performance evaluation of this algorithm was assessed using a set of metrics ranging from 1 to 14. These metrics provide valuable insights into the algorithm's effectiveness and capabilities. We can determine accuracy, precision, recall, F1 score, and other relevant indicators by employing a comprehensive evaluation framework. These metrics enable us to measure the algorithm's performance across different dimensions and validate its effectiveness in addressing the given problem. The evaluation process assists in identifying strengths and areas for improvement, ensuring the algorithm's robustness and reliability. By meticulously examining these metrics, we gain a holistic understanding of the algorithm's performance and potential impact on real-world applications.

$$TP\ Rate = \frac{TP}{TP+FN} \quad (1)$$

$$FP\ Rate = \frac{FP}{FP+TN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F - Measure = \frac{2 * precision * Recall}{precision + Recall} \quad (5)$$

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP) * (TN + FP) * (TN + FN)}} \quad (6)$$

$$Roc\ Area = \frac{(TPR + TPR_p) * (FPR_p - FPR)}{2} \quad (7)$$

$$PRC\ Area = \frac{(Precision + precision_p) * (Recall_{prev} - Recall)}{2} \quad (8)$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

5 Result analysis

We used machine learning algorithms to analyze the data, and the results were as follows from Table No. 4 to Table No. 8 regarding the data of Al-Diwaniyah Hospital:

Table 4. The result of K nearest neighbor (K.N.N.) for the Al-Diwaniyah Hospital data set

Metrics	Dead	Alive	Good	Weighted Avg
T.P. Rate	0.895	0.611	0.922	0.861
FP Rate	0.075	0.056	0.086	0.078
Precision	0.841	0.69	0.922	0.858
Recall	0.895	0.611	0.922	0.861
F-Measure	0.867	0.648	0.922	0.859
MCC	0.806	0.583	0.835	0.784
ROC Area	0.955	0.879	0.962	0.946
PRC Area	0.904	0.648	0.96	0.89

Table 5. The result of the Decision tree (D.T.) for Al-Diwaniyah Hospital data set

Metrics	Dead	Alive	Good	Weighted Avg
T.P. Rate	0.906	0.789	0.946	0.907
FP Rate	0.039	0.045	0.06	0.051
Precision	0.912	0.781	0.946	0.907
Recall	0.906	0.789	0.946	0.907
F-Measure	0.909	0.785	0.946	0.907
MCC	0.869	0.741	0.885	0.856
ROC Area	0.932	0.853	0.936	0.921
PRC Area	0.834	0.616	0.905	0.834

Table 6. The result of Random forest (R.F.) for Al-Diwaniyah Hospital data set

Metrics	Dead	Alive	Good	Weighted Avg
T.P. Rate	0.854	0.684	0.915	0.857
FP Rate	0.057	0.075	0.086	0.075
Precision	0.869	0.65	0.921	0.859
Recall	0.854	0.684	0.915	0.857
F-Measure	0.861	0.667	0.918	0.858
MCC	0.801	0.597	0.828	0.781
ROC Area	0.935	0.843	0.944	0.924
PRC Area	0.822	0.78	0.96	0.887

Table 7. The result of Naïve Bayes (N.B.) for Al-Diwaniyah Hospital data set

Metrics	Dead	Alive	Good	Weighted Avg
T.P. Rate	0.906	0.779	0.949	0.907
FP Rate	0.039	0.043	0.064	0.053
Precision	0.912	0.787	0.943	0.907
Recall	0.906	0.779	0.949	0.907
F-Measure	0.909	0.783	0.946	0.907
MCC	0.869	0.739	0.885	0.856
ROC Area	0.94	0.856	0.95	0.931
PRC Area	0.838	0.698	0.944	0.87

Table 8. The result of Artificial neural network (ANN) for Al-Diwaniyah Hospital data set

Metrics	Dead	Alive	Good	Weighted Avg
T.P. Rate	0.854	0.6	0.918	0.845
FP Rate	0.077	0.067	0.098	0.086
Precision	0.83	0.648	0.912	0.842
Recall	0.854	0.6	0.918	0.845
F-Measure	0.841	0.623	0.915	0.843
MCC	0.771	0.55	0.821	0.76
ROC Area	0.96	0.903	0.964	0.952
PRC Area	0.9	0.727	0.976	0.91

and note that the Naïve Bayes (N.B.) algorithm achieved the highest accuracy=90.7143 with TP Rate=0.949 in label (Good) and the accuracy of algorithm Decision tree (D.T.) =90.7143 with T.P. Rate = 0.946 in label (Good) and the accuracy of the algorithm Random Forest (R.F.) = 86.7857 with T.P. Rate = 0.922 in label (Good) and the accuracy of the algorithm K nearest neighbor (K.N.N.)= 86.4286 with T.P. Rate = 0.942 in label (Good) and the accuracy of the algorithm Artificial neural network (ANN) = 84.2857 with T.P. Rate 0.898 in label (Good)

And the Result analysis of Najaf Hospital data by using machine learning algorithms to analyze the data, and the results were as follows from Table No. 9 to Table No. 13 regarding:

Table 9. The result of K nearest neighbor (K.N.N.) for the Najaf Hospital data set

Metrics	Dead	Alive	Good	Weighted Avg
T.P. Rate	0.875	0.798	0.947	0.897
FP Rate	0.058	0.034	0.07	0.059
Precision	0.819	0.87	0.943	0.898
Recall	0.875	0.798	0.947	0.897
F-Measure	0.846	0.832	0.945	0.897
MCC	0.799	0.789	0.878	0.84
ROC Area	0.947	0.945	0.964	0.956
PRC Area	0.912	0.879	0.96	0.931

Table 10. The result of the Decision tree (D.T.) for Najaf Hospital data set

Metrics	Dead	Alive	Good	Weighted Avg
T.P. Rate	0.864	0.881	0.947	0.913

FP Rate	0.031	0.047	0.058	0.049
Precision	0.894	0.841	0.952	0.914
Recall	0.864	0.881	0.947	0.913
F-Measure	0.879	0.86	0.949	0.913
MCC	0.843	0.82	0.889	0.863
ROC Area	0.916	0.906	0.923	0.917
PRC Area	0.795	0.759	0.906	0.847

Table 11. The result of Random Forest (R.F.) for Najaf Hospital data set

Metrics	Dead	Alive	Good	Weighted Avg
T.P. Rate	0.83	0.821	0.933	0.884
FP Rate	0.055	0.051	0.076	0.065
Precision	0.82	0.821	0.937	0.885
Recall	0.83	0.821	0.933	0.884
F-Measure	0.825	0.821	0.935	0.884
MCC	0.772	0.771	0.857	0.818
ROC Area	0.91	0.921	0.937	0.927
PRC Area	0.897	0.888	0.925	0.911

Table 12. The result of Naïve Bayes (N.B.) for Najaf Hospital data set

Metrics	Dead	Alive	Good	Weighted Avg
T.P. Rate	0.886	0.881	0.952	0.921
FP Rate	0.034	0.034	0.058	0.047
Precision	0.886	0.881	0.952	0.921
Recall	0.886	0.881	0.952	0.921
F-Measure	0.886	0.881	0.952	0.921
MCC	0.852	0.847	0.894	0.874
ROC Area	0.935	0.914	0.939	0.932
PRC Area	0.835	0.841	0.919	0.883

Table 13. The result of the Artificial neural network (ANN) for Najaf Hospital data set

Metrics	Dead	Alive	Good	Weighted Avg
---------	------	-------	------	--------------

T.P. Rate	0.807	0.798	0.918	0.866
FP Rate	0.062	0.047	0.11	0.085
Precision	0.798	0.827	0.91	0.865
Recall	0.807	0.798	0.918	0.866
F-Measure	0.802	0.812	0.914	0.866
MCC	0.742	0.76	0.809	0.783
ROC Area	0.939	0.916	0.953	0.942
PRC Area	0.892	0.857	0.932	0.906

Note that the Naïve Bayes (N.B.) algorithm achieved the highest accuracy=92.1053 with TP Rate=0.952 in label (Good) and the accuracy of algorithm Decision tree (D.T.) =91.3158 with T.P. Rate = 0.947 in the label (Good) and the accuracy of the algorithm K nearest neighbor (K.N.N.) = 89.7368 with T.P. Rate = 0.947 in the label (Good) and the accuracy of the algorithm Random Forest (R.F.) = 88.4211 with T.P. Rate = 0.933 in label (Good) and the accuracy of the algorithm Artificial neural network (ANN) = 86.5789 with T.P. Rate 0.918 in label (Good)

6 Conclusion

The COVID-19 pandemic has caused widespread infection and death worldwide, with the virus first appearing in Wuhan, China, in December 2019. The World Health Organization declared the pandemic on March 11, 2020, highlighting the inability of governments and health organizations to contain and prevent its spread. The COVID-19 dataset, collected from Diwaniyah Governmental Hospital and Najaf Governmental Hospital, includes 43 features, including 171 deaths, 95 injured, 294 suspects, and 380 records for Najaf Hospital. It will be useful for analyzing the health situation, recognizing patterns, and making informed decisions for effective health management and resource allocation. A.I., a modern technology, has applications in military, economic, and industrial sectors, including medical data analysis, diagnostic tools, and remotely guided surgeries. Machine learning algorithms like K.N.N., decision tree, Nave Bayes, and artificial neural networks have shown high accuracy in predicting COVID-19 patients, providing valuable insights into the virus's spread and potential impact on healthcare professionals.

7 Acknowledgment

I extend my thanks, pride, and appreciation to Dr. Atheer Mahmoud Ali, a specialist in the chest and respiratory diseases at Al-Diwaniyah Hospital, for all the assistance he provided me in collecting and analyzing data.

8 References

- [1] S. S. Mohamed Ali, A. H. Alsaeedi, D. Al-Shammary, H. H. Alsaeedi, and H. W. Abid, "Efficient intelligent system for diagnosis pneumonia (SARSCoVID19) in X-ray images empowered with initial clustering," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 22, no. 1, pp. 241–251, 2021, doi: 10.11591/ijeecs.v22.i1.pp241-251.
- [2] X. Wang, T. Okoshi, and J. Nakazawa, "Lower Face Inpainting Aiming at Face Recognition under Occlusions," *2022 IEEE Int. Conf. Pervasive Comput. Commun. Work. other Affil. Events, PerCom Work. 2022*, pp. 62–65, 2022, doi: 10.1109/PerComWorkshops53856.2022.9767220.
- [3] S. S. Mohamed Ali, A. H. Alsaeedi, D. Al-Shammary, H. H. Alsaeedi, and H. W. Abid, "Efficient intelligent system for diagnosis pneumonia (SARSCoVID19) in X-ray images empowered with initial clustering," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 22, no. 1, pp. 241–251, 2021, doi: 10.11591/ijeecs.v22.i1.pp241-251.
- [4] V. Aswal, O. Tupe, S. Shaikh, and N. N. Charniya, "Single Camera Masked Face Identification," *Proc. - 19th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2020*, pp. 57–60, 2020, doi: 10.1109/ICMLA51294.2020.00018.
- [5] A. H. Al-saeedi, "Binary Mean-Variance Mapping Optimization Algorithm (BMVMO)," *J. Appl. Phys. Sci.*, vol. 2, no. 2, pp. 42–47, 2016, doi: 10.20474/japs-2.2.3.
- [6] Worldometer, "COVID-19 Coronavirus Pandemic," [Online]. Available: <https://www.worldometers.info/coronavirus/#coronavirus-cases-linear>. [Accessed: June 30, 2023].
- [7] S. Nematzadeh, F. Kiani, M. Torkamanian-Afshar, and N. Aydin, "Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases," *Comput. Biol. Chem.*, vol. 97, p. 107619, 2022, doi: <https://doi.org/10.1016/j.compbiolchem.2021.107619>.
- [8] R. R. Nuiiaa, S. Manickam, and A. H. Alsaeedi, "A Comprehensive Review of DNS-based Distributed Reflection Denial of Service (DRDoS) Attacks: State-of-the-Art," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 12, no. 6, pp. 2452–2461, 2022, doi: 10.18517/ijaseit.12.6.17280.
- [9] Y. Shi et al., "An overview of COVID-19," *J. Zhejiang Univ. Sci. B*, vol. 21, no. 5, p. 343, 2020.
- [10] B. F. Haynes et al., "Prospects for a safe COVID-19 vaccine," *Sci. Transl. Med.*, vol. 12, no. 568, p. eabe0948, 2020.
- [11] S. J. Daniel, "Education and the COVID-19 pandemic," *Prospects*, vol. 49, no. 1, pp. 91–96, 2020.
- [12] C. Wang, Z. Li, M. Clay Mathews, S. Praharaj, B. Karna, and P. Solís, "The spatial association of social vulnerability with COVID-19 prevalence in the contiguous United States," *Int. J. Environ. Health Res.*, vol. 32, no. 5, pp. 1147–1154, 2022.
- [13] F. Saberi-Movahed et al., "Decoding clinical biomarker space of COVID-19: Exploring matrix factorization-based feature selection methods," *Comput. Biol. Med.*, vol. 146, pp. 1–51, 2022, doi: 10.1016/j.compbiomed.2022.105426.
- [14] Accurate detection of Covid-19 patients based on Feature Correlated Naïve Bayes (FCNB) classification strategy," in *Proc. IEEE International Conference on [Conference Name]*, 2023 pp

151–153 doi: 10.11591/hamj.v2.i1.pp151–1531.

- [15] Prakash, Kolla Bhanu, et al. "Analysis, prediction and evaluation of covid-19 datasets using machine learning algorithms." *International Journal* 8.5 (2020): 2199-2204.
- [16] Dubey, Ashutosh Kumar, et al. "Performance estimation of machine learning algorithms in the factor analysis of COVID-19 dataset." *Computers, Materials and Continua* (2020).
- [17] Singh, Hari, and Seema Bawa. "Predicting COVID-19 statistics using machine learning regression model: Li-MuLi-Poly." *Multimedia Systems* 28.1 (2022): 113-120.
- [18] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic," *Meas. J. Int. Meas. Confed.*, vol. 167, no. May 2020, 2021, doi: 10.1016/j.measurement.2020.108288.
- [19] S. M. Hadi *et al.*, "Trigonometric words ranking model for spam message classification," *IET Networks*, no. August, 2022, doi: 10.1049/ntw2.12063.
- [20] D. Al-Shammary, A. L. Albukhnefi, A. H. Alsaeeedi, and M. Al-Asfoor, "Extended particle swarm optimization for feature selection of high-dimensional biomedical data," *Concurr. Comput. Pract. Exp.*, vol. 34, no. 10, p. e6776, 2022, doi: <https://doi.org/10.1002/cpe.6776>.
- [21] V. Meshram, K. Patil, V. Meshram, D. Hanchate, and S. D. Ramkteke, "Machine learning in agriculture domain: A state-of-art survey," *Artif. Intell. Life Sci.*, vol. 1, no. August, p. 100010, 2021, doi: 10.1016/j.aills.2021.100010.
- [22] A. H. Jabor and A. H. Ali, "Dual Heuristic Feature Selection Based on Genetic Algorithm and Binary Particle Swarm Optimization," *J. Univ. BABYLON Pure Appl. Sci.*, vol. 27, no. 1, pp. 171–183, 2019, doi: 10.29196/jubpas.v27i1.2106.
- [23] R. R. Nuiaa, S. Manickam, and A. H. Alsaeeedi, "Distributed reflection denial of service attack: A critical review," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 6, pp. 5327–5341, 2021, doi: 10.11591/ijece.v11i6.pp5327-5341.
- [24] O. Almomani, "A feature selection model for network intrusion detection system based on pso, gwo, ffa and ga algorithms," *Symmetry (Basel)*, vol. 12, no. 6, pp. 1–20, 2020, doi: 10.3390/sym12061046.
- [25] S. Manickam, R. Rahef Nuiaa, A. Hakem Alsaeeedi, Z. A. A. Alyasseri, M. A. Mohammed, and M. M. Jaber, "An enhanced mechanism for detection of Domain Name System-based distributed reflection denial of service attacks depending on modified metaheuristic algorithms and adaptive thresholding techniques," *IET Networks*, vol. 11, no. 5, pp. 169–181, 2022.
- [26] A. Bassel, A. B. Abdulkareem, Z. A. A. Alyasseri, N. S. Sani, and H. J. Mohammed, "Automatic Malignant and Benign Skin Cancer Classification Using a Hybrid Deep Learning Approach," *Diagnostics*, vol. 12, no. 10, 2022, doi: 10.3390/diagnostics12102472.
- [27] A. S. Alfoudi *et al.*, "Hyper clustering model for dynamic network intrusion detection," *IET Commun.*, 2022, doi: 10.1049/cmu2.12523.
- [28] S. M. Hadi, A. H. Alsaeeedi, M. I. Dohan, R. R. Nuiaa, S. Manickam, and A. S. D. Alfoudi, "Dynamic Evolving Cauchy Possibilistic Clustering Based on the Self-Similarity Principle (DECS) for Enhancing Intrusion Detection System," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 5, pp. 252–260, 2022, doi: 10.22266/ijies2022.1031.23.

- [29] A. S. Alfoudi, A. H. Alsaedi, M. H. Abed, A. M. Otebolaku, and Y. S. Razooqi, "Palm Vein Identification Based on Hybrid Feature Selection Model," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 5, pp. 469–478, 2021, doi: 10.22266/ijies2021.1031.41.
- [30] R. R. Nuiiaa, S. Manickam, A. H. Alsaedi, and E. S. Alomari, "A new proactive feature selection model based on the enhanced optimization algorithms to detect DRDoS attacks," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 2, pp. 1869–1880, 2022, doi: 10.11591/ijece.v12i2.pp1869-1880.
- [31] A. L. A. A. H. A. A. H. A. Mehdi Ebady Manna, "A proactive metaheuristic model for optimizing weights of artificial neural network," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 20, no. 2, pp. 976–984, 2020, doi: 10.11591/ijeecs.v20.i2.pp976-984.
- [32] P. Boonyopakorn, "The Optimization and Enhancement of Network Intrusion Detection through Fuzzy Association Rules," *2019 6th Int. Conf. Tech. Educ. ICTechEd6 2019*, no. August, 2019, doi: 10.1109/ICTechEd6.2019.8790881.

Article submitted 1 Jun 2023. Accepted at 2 July 2023

Published at 30 September 2023.