Al-Rafidain Journal of Computer Sciences and Mathematics (RJCM)

www.csmj.mosuljournals.com

Prediction of COVID-19 Effect on Patients during Six Month After Recovery, by Using AI Algorithm

Hanan Anas Aldabagh1,* Ghayda Abdulaziz Altalib2

The General Directorate of Education in Nineveh Governorate, Mosul, Iraq 1 , Department of Computer Science, College of computer science and mathematics, Mosul University, Mosul, Iraq 2
**Corresponding author. Email: hanan.csp9@student. uomosul.edu.iq1*

Article information

Article history:
Received : 14/9/2022
Accepted : 17/1/2023
Available online :

Abstract

People all over the world have experienced severe panic as a result of the novel coronavirus (COVID-19). To secure their future, it is crucial to undertake thorough evaluations in their psychological, physical, and social domains, comprehend the potential outcomes of patients recovering from it, and ascertain whether they have any other harmful diseases. This possible outcome for people recovered from Covid 19 was predicted by collecting data from people who had previously been infected with this virus to determine the effects they had, using intelligent techniques. The GSO algorithm was used for feature selection, and for hyper-parameter tuning for the Random Forest (RF) algorithm used in the prediction process in order to make predictions to identify the effects that may accrue on recoveries persons. This model was evaluated using different metrics after performing multiple processing operations on the data and using the GSO algorithm to perform the feature selection process in order to obtain the important features. Good results were obtained for each expected effect, as the highest AUC was obtained when predicting the impact of the gastrointestinal tract of recovered persons, which is 0.91. This will then reveal the effects that Covid-19 has had on people after they have recovered. This will assist in anticipating possible results to provide counseling and psychological support, as well as some recommended guidelines for healing patients and the community to return to a normal life.

Keywords:
Prediction, COVID-19, physical activity, Artificial Intelligence, Machine Learning, Glowworm Swarm Optimization.

Correspondence:
Author :Hanan Anas Aldabagh
Email:hanan.csp9@student.uomosul.edu.iq

1. INTRODUCTION

A significant global issue was brought on by the Coronavirus (COVID-19) outbreak in late 2019 [1]. Just one month after it started to spread, the WHO had to declare the epidemic to be a pandemic due to its extreme severity. The spread of the virus is inflicting a shock on the global economy by severely disrupting many industries and public functions, including the supply chain, payments, public transportation, and the financial system [2].

Artificial intelligence mimics human intelligence. Through automatic driving, fraud detection, robotics, computer vision, and internet advertising all uses AI. With its effectiveness in diagnosis, care, patient monitoring, medication research, epidemiology, etc., AI may become a vibrant field of study to answer humanity's concerns [3].

Corona virus infection leads to long-term effects, although these effects appear on some survivors only, and this suggests the need to study and follow up on the long-term effects that may get to those recovering from Covid-19[4]. It is necessary to determine the prevalence of these long-term outcomes to facilitate timely preparations for the management of survivors.

2. Related Works

Claudia et al. presented research to describe clinical progression and predict symptom continuity during 2-month follow-up in adults with non-critical COVID-19 in [5]. 150 patients were followed up in Tours Hospital, Italy, from 17/3 to 3/4. Hong et al. also presented research to study the long-term pulmonary function and related physiological characteristics of COVID-19 survivors in [6]. COVID-19 survivors were recruited to undergo chest and lung HRCT and IgG antibody tests 3 months after discharge. In [7] the long-term health consequences of hospitalized patients with COVID-19 and their associated risk factors are described by Huang et al. A study was conducted on COVID-19 patients who were discharged from Jin Yin-tan Hospital (Wuhan) between 7/1 and 29/5 and all patients were interviewed through a series of questionnaires to assess symptoms, health-related quality of life, physical examinations and a 6-minute walking test, Blood and lung examinations and a high-resolution CT scan of the chest were performed.

Each day brings new AI concepts, applications, and technology, so it has been useful in identifying corona disease outbreaks, diagnosing patients, and disinfecting regions [8]. "Prediction" refers to statistical and probabilistic data from past observations. AI algorithms have been used to disease spread [9], inventory valuation [10], weather [11] and sales [12] in recent years. Prediction methods are also useful in healthcare. Sarkar et al. in [13] used 430 instances of COVID-19 infection gathered from Kaggle to apply the RF (random forest) classification method to a clinical data set in identifying critical variables and their impact on death. The model has been validated in the test data set using the AUC-ROC (Areas Under the Curve and the Receiver Operating characteristic Curve). Yan et al. [14] used general ML techniques and the XGBoost (eXtreme Gradient Boosting) classifier to pick three biomarkers (lymphocytes, lactic dehydrogenase and high sensitivity C-reactive protein) that predicted survival with over 90% accuracy. Clinical information of 485 patients was used. It has been found that the severity of COVID-19 patients can be discerned when LDH levels are increased. In [15], Ogundokun et al. proposed a simple aggregated ML strategy that was created to pridict the length, number, and size of COVID-19 cases in India, employing three regression techniques as an ensemble method: LR, SVM, and NN. An online dataset of 408,658 samples of patients in India was used. When employing RMSE and MSE metrics to evaluate performance, the greatest results were obtained when these techniques were combined rather than employed independently. Iwendi et al. showed in [16] how they used general machine learning algorithms to come up with a model that could predict the severity of the condition and also the likely outcome by using RF with the AdaBoost algorithm and data from COVID-19 patients including their health, travel, and demographics. The model was tested using the F1 Score, accuracy, precision, and recall, and it was found to be 94% accurate because there was a link between the sex of the patients as well as their deaths that could be seen in the data.

This paper aims to know and proactively identify the effects that can take place on those having recovered from Covid 19 in order to support the infected during the recovery phase with the objective of restoring their health and early identification of the effects that may encourage people to follow them quickly. Including warning the infection and medical staff by the risk of persistent symptoms in the injured, regardless of the severity of their damage. Also performing pulmonary rehabilitation on patients with chronic respiratory disorders in order to increase their exercise capacity and eliminate shortness of breath. Community Respiratory Teams will play a key role in the early and long-term care of discharged patients in order to identify recovery needs, control breathing, and examine physical health and physical activity. This goal was achieved by using a proposed model that uses intelligent algorithms to predict the effects that may have on people recovering from Covid-19 and test the accuracy of this prediction using different measures.

3. Material And Methods

3.1. Data Acquisition

Due to the lack of clinical information and laboratory data to follow up on the state of patients after their recovery a Google-hosted questionnaire form was developed and distributed via social media platforms including WhatsApp, Messenger, Facebook, and others. Where the form was prepared exactly based on published medical studies and a specialist doctor was consulted to confirm useful questions (Juli Evangelou Strait, 2021; World Health organization, 2020). As the questionnaire asks about the person's age and gender, chronic illnesses like diabetes, asthma, kidney disease, and other conditions, as well as whether or not they smoke, as well as information about the person's health at the time of the infection, such as date of infection, length of the infection, oxygen saturation, whether or not they might need to be admitted to an intensive care unit, and the symptoms of the infection. The general framework of the proposed model is shown in Figure 1.

3.2. Data pre-processing

The following steps are included in the process of managing the collected data in order to be suitable for the proposed model:

A. **Data Cleaning**
   The responses that represent the period of follow-up were taken in the first six months after the infection, were covered by 457 people. The data was cleaned and processed by discarding the responses that lacked age and that were less than a month old after the infection. Some of the responses were processed with the aid of a medical practitioner.

B. **Data Aggregation**
   In order to infer the effect of Corona virus on each organ or system in the human body separately, we aggregate the answers related to the effect of a particular organ or system to obtain one answer placed under the title of this organ or system as follows:

- **Gastrointestinal system:** constipation, diarrhea and acid reflux.
- **Respiratory system:** persistent cough, shortness of breath.
- **Nervous system:** headaches, memory problems and problems with senses of taste and smell.
- **Mental health:** anxiety, depression, sleep problems and substance abuse.
- **Metabolism:** new onset of diabetes, obesity and high cholesterol.
- **Cardiovascular system & Coagulation regulation:** heart failure, blood clots in the legs and lungs.
- **Kidney:** acute kidney infection and chronic kidney disease that can, in severe cases, require dialysis.
- **General health:** anemia.
- **Skin:** rash and hair loss.
- **Musculoskeletal system**: joint pain and muscle weakness.
- **Physical Activity:** Exercises, Walking, running, doing daily chores, Up the stairs, Carrying heavy things.

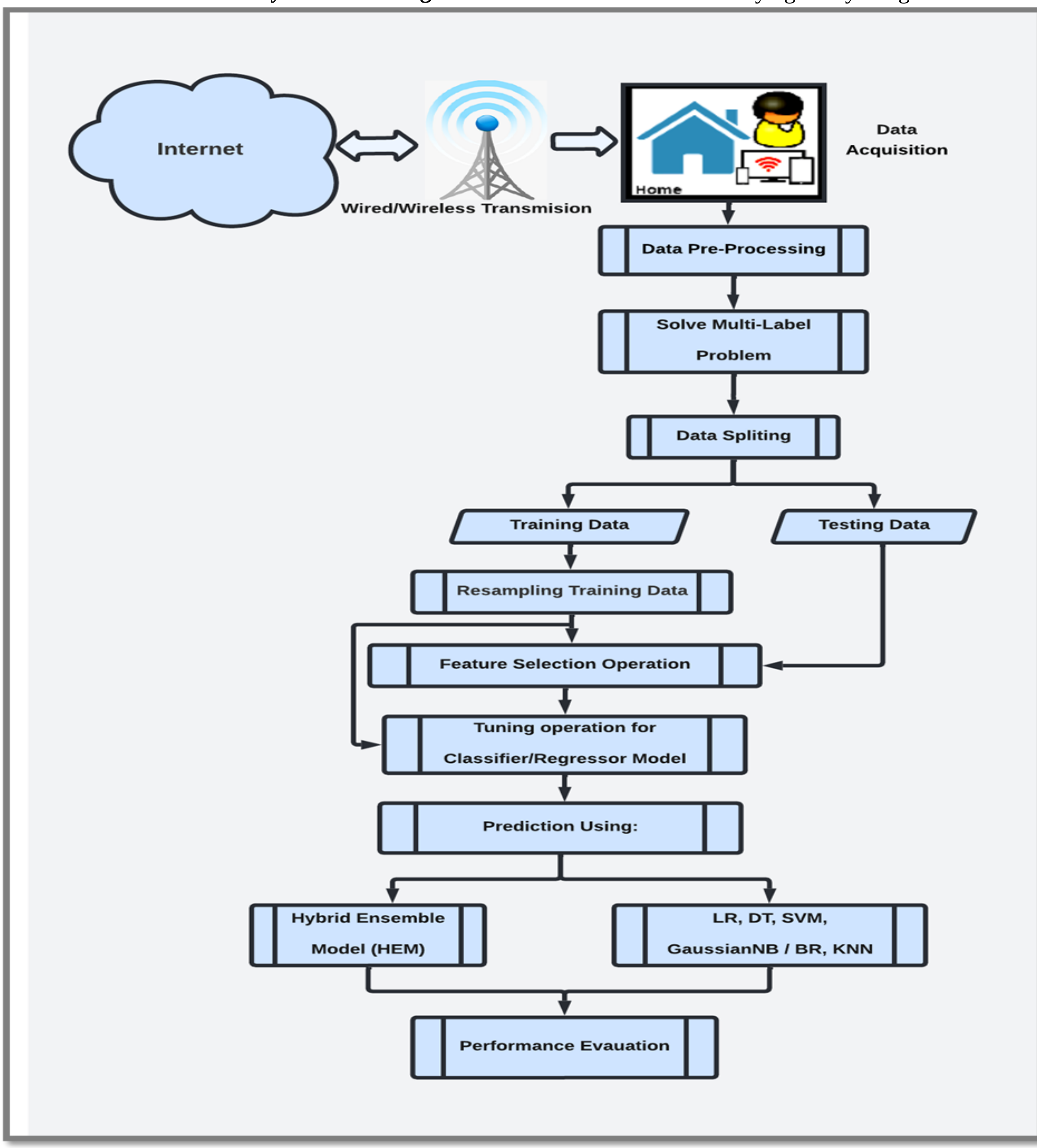


Figure 1. Proposed model framework.

C. Data Clustering

The data was put into 7 groups based on age: 15–25, 25–35, 35–45, 45–55, 55–65, 65–75, and 75–85. People between 15 - 25 were most likely to be in these groups because they use social media the most. The 37 obtained features with 11 outputs. The feature is: Age(in ranges:15-25, 25-35, 35-45, 45-55, 55-65, 65-75, 75-85), Sex, Smoking, follow up months(number of follow up months), infection days(number of infection days: 3 day, 1 week, 2 week, 3 week, Month, More than month), SpO2(an estimate of the oxygen saturation level that is often evaluated using a pulse oximeter: more than 90, between 80 and 90, less than 80), needed SpO2, ICU(entered Intensive Care Unit), chronic disease(high blood pressure, diabetes, kidney disease, Liver disease, heart disease, Asthma or chest allergy, cancer disease, viral liver disease, bronchitis disease, Hypothyroidism, rheumatic inflammation, arthritis disease), symptoms(High temperature, cold, cough, tonsillitis, Runny nose, Sputum, shortness of breath, headache, dizzy, Anorexia, diarrhea, vomiting, Tired, muscle pain, high blood pressure sym, Pneumonia, stomachache).

D. Data Transformation (discretization)

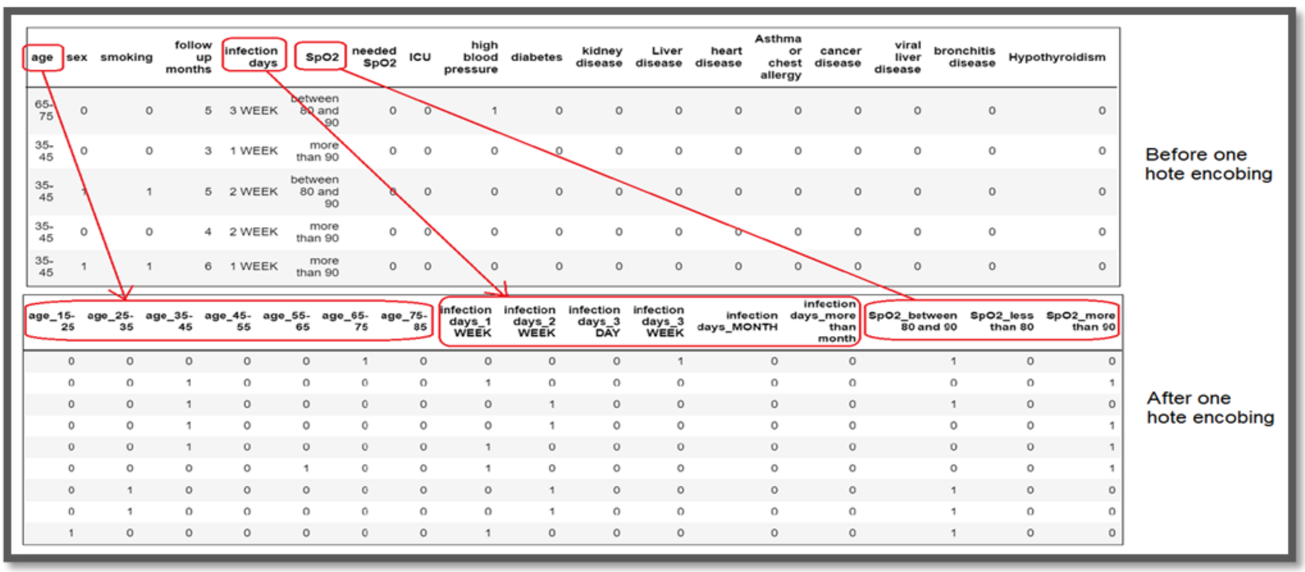
Four fields in the dataset (infection days, SPO2, age, and sex) are of the categorical kind and the rest of which are numeric. These fields were transformed to numeric using one-hot encoding technique, as shown in figure (2), because most ML algorithms require all data to be numeric.

E. Data Normalization

Normalization is the process of putting all features on the same distribution or scale, which is the final stage in data pre-processing. Regardless of the unit of measurement, higher values are given more weight and lower values are given less weight. In this scenario, normalization was utilized to keep the relative disparities between feature values within a specified range. To normalize feature values, the [0:1] range is typically used. Equation (1) yields the normalization of the feature on a scale of 0 to 1. [19]:

$$f_{scaled} = \frac{f - f_{min}}{f_{max} - f_{min}} \qquad \text{…..(1)}$$

Where $f \quad ax$: denotes the feature's greatest value and $fmin$: denotes its smallest value. To do this, the MinMaxScaler function was used.



age	sex	smoking	follow up months	infection days	SpO2	needed SpO2	ICU	high blood pressure	diabetes	kidney disease	Liver disease	heart disease	Asthma or chest allergy	cancer disease	viral liver disease	bronchitis disease	Hypothyroidism
65-75	0	0	5	3 WEEK	between 80 and 90	0	0	1	0	0	0	0	0	0	0	0	0
35-45	0	0	3	1 WEEK	more than 90	0	0	0	0	0	0	0	0	0	0	0	0
35-45	1	1	5	2 WEEK	between 80 and 90	0	0	0	0	0	0	0	0	0	0	0	0
35-45	0	0	4	2 WEEK	more than 90	0	0	0	0	0	0	0	0	0	0	0	0
35-45	1	1	6	1 WEEK	more than 90	0	0	0	0	0	0	0	0	0	0	0	0

age_15-25	age_25-35	age_35-45	age_45-55	age_55-65	age_65-75	age_75-85	infection days_1 WEEK	infection days_2 WEEK	infection days_3 DAY	infection days_3 WEEK	infection days_MONTH	infection days_more than month	SpO2_between 80 and 90	SpO2_less than 80	SpO2_more than 90
0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0
0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1
0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0
0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1
0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1
0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1
0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0
0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0
1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0

Figure 2. Transformation on dataset.

3.3. Solve Multi-label Classification Problem

After the pre-processing worked on the data that represents the data of people who recovered from COVID-19 within 6 months after recovery, there will be 11 effects (outputs) for each person (record), which means that this data is of a multi-label classification type. To solve this problem, we transform this data into a single label using the classifier chains method, because it will keep the correlation [20]. The process will be repeated 11 times for the model, which means that every time we produce a target output, it will be added as an additional input to the feature being entered into the model and so on until the process is complete. This last process of the model will use the existing features with the 10 outputs resulting from the previous operations (whole output + features) to get the final prediction.

3.4 Data Splitting

The data generated by the previous processes is randomly separated into two groups, the training group (which contains

80% of the data) and the testing group (20% of the data).

3.5 Resampling Training Data

Unbalanced data sets have about equal categories, but one has more samples than the other. Classifiers may perform well on the majority class but badly on the minority due to its increased effect. Unbalanced data sets are often resampled to produce a more balanced distribution of class states. Random undersampling and oversampling are resampling methods. Undersampling removes majority class samples to balance the collection. Random oversampling balances datasets by duplicating minority class examples [21].

Because the data is unbalanced, we changed it to balance in order to achieve accuracy and unbiasedness in the model, either by undersampling or oversampling[22]. Figure (3) depicts the data after resampling techniques were used to obtain the Gastrointestinal system output.

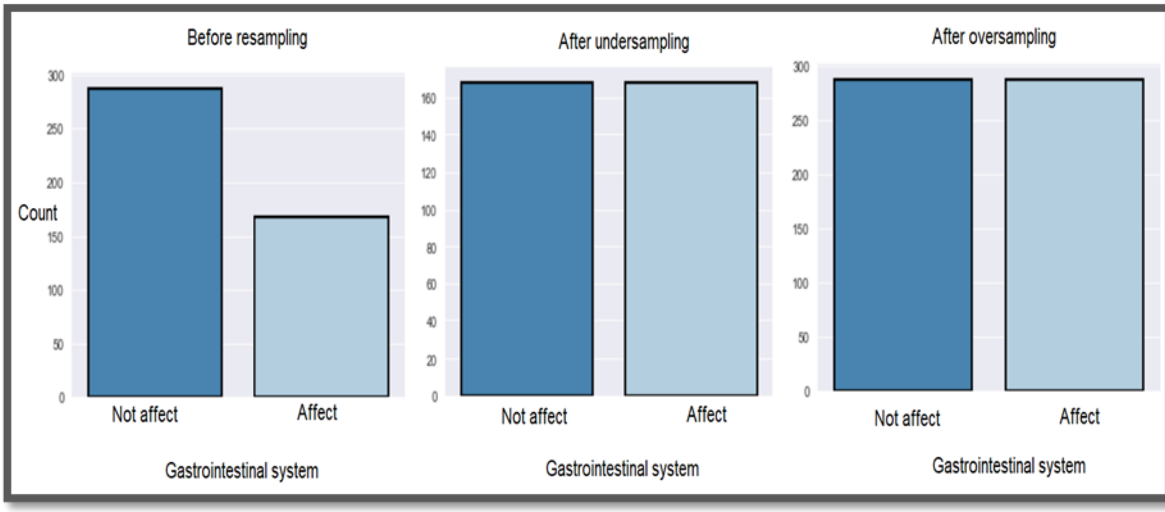


Figure 3. Resampling data from first group.

3.6. Features Selection (FS) Operation

Algorithms based on swarm intelligence are a clear solution for enhancing feature subset selection in the wrapper methodology. Wrapper models evaluate the quality of the features using a predetermined ML approach, and the FS process avoids the algorithm's representational biases [23]. The Glowworm Swarm Optimization (GSO) Approach, a swarm optimization method based upon this Ant-Colony Optimization (ACO) suite of algorithms, was used, which was originally created by Krishnanand et al [24].

The GSO Algorithm was updated by transforming the fitness function to a classification function to be utilized in feature selection, which will speed up training and improve accuracy. Her GSO chooses the feature subset. The fitness value of a glowworm is then calculated using K-fold cross-validation. There are K subsets of the training data. K-1 training subsets are utilized as inputs for GSO, and one test subset is used to determine the fitness of each glowworm. The glowworm's fitness value will be the mean of the K classification accuracies calculated. While the test dataset is not used in this GSO feature selection procedure, it is used in the final evaluation, when the classification accuracy of the best feature subsets is determined. The RF algorithm determines the classification accuracy of each created glowworm model. Figure (4) depicts every aspect of the suggested method. Algorithm 1 includes feature selection operation steps.

Algorithm-1: Feature Selection Operation using GSO Algorithm

Input:

1. Training set.
2. Hyper-parameter of GSO.
3. α=0.99, max_iteration=100.

Output:

Best feature subset with best fitness.

Method:

Step1: Generate vector of float element between 0 and 1 randomly as following:

$X = \{x1, x2, \ldots xd\}, xi \in [0,1]$

Where d is the total number of features in the dataset.

Step2: Convert X to binary, using the threshold value 0.5 to given selected feature.

If no feature were selected:

Return 1.0 as fitness.

Else

Compute the mean of running 2-fold cross validation on the training set using evaluation metric of the RF classifier/regressor.

Compute the value of optimization function to get best fitness of the best feature that selected using:

F(x)=α * (1- value of evaluation metric) + (1-α) * No. of selected feature / No. of all features.

If the value of optimization function condition is satisfying:

Save the best feature and best fitness.

Step3: Run GSO to get vector of float element between 0 and 1 as following.

Step4: Repeat from step 1 until max_iteration=0.

Step5: Return Best feature subset with best fitness.
Step6: End algorithm

When the feature selection technique was used to predict some effects, the results varied because we received better results for some outputs (effects), but when it was applied to other effects, it did not produce satisfactory results. Because of this, not all effects used this algorithm.

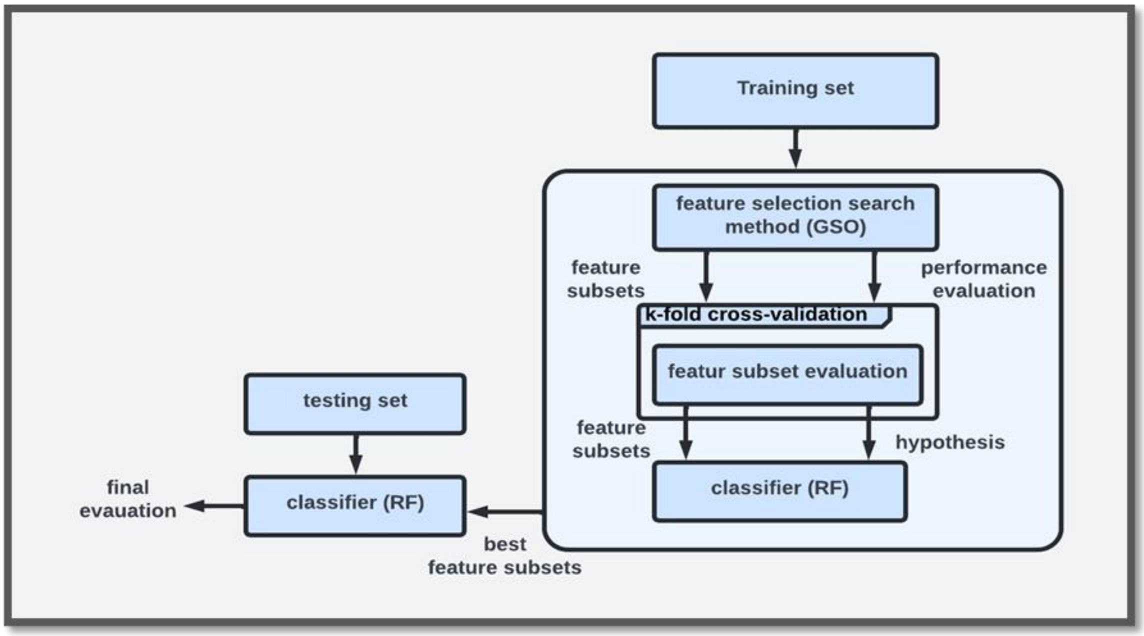


Figure 4. Framework of the proposed features selection model

3.7. Tuning operation for classifier model

To choose the best hyper-parameters for classification algorithms, hyper-parameter determination is typically a challenging problem because choosing the right hyper-parameters can significantly affect how well a prediction model performs, allowing for a more optimal solution and a higher degree of model accuracy [25]. In order to change the hyper-parameter, swarm intelligence optimization utilizing the GSO algorithm has been suggested. This approach will be used to adjust the hyper-parameters for RF algorithm. The tuning procedures for the classifier model are given by algorithm 2.

Algorithm 2: Tuning operation for classifier model using GSO Algorithm

Input:

1. Training set.
2. Default hyper-parameter of GSO except population_size=20 and seed=5.
3. max_iteration=100.
4. The classifier model's hyperparameters and their range.

Output:

Best hyper-parameter with best fitness.

Method:

Step1: Generate the value of hyper-parameter randomly from their range.

Step2: Compute the mean of running 5-fold cross validation on the training set using hyper-parameter of classifier/regressor, that denoted as optimization function to get best fitness.

If the value of optimization function condition is satisfying:

Save the best hyper-parameter and best fitness.

Step3: Run GSO to get the value of hyper-parameter from their range.

Step4: Repeat from step1 until max_iteration.

Step5: Return best hyper-parameter with best fitness.
Step6: End algorithm

3.8. Prediction using RF algorithm as classifier model

After the feature selection process was done with GSO, these selected features will be fed along with the outcome into the RF algorithm to predict the effects on people recovering from COVID-19.

4. Result performance of algorithms

Using training and testing data, the suggested model's performance was evaluated. The model was trained using the training data. The evaluation was done on to testing data using the following metrics: AUC, F1-score, accuracy, hamming loss, as illustrated in Table 1.

Table 1. Performance results of algorithms

Output from proposed model	AUC	F1 Score	Accuracy	Loss
Gastrointestinal system	0.910	0. 927	0.890	0.110
Respiratory system	0.833	0. 744	0.825	0. 175
Nervous system	0. 787	0.859	0.816	0. 184
Mental health	0. 843	0.862	0. 842	0.157
Metabolism	0. 804	0.735	0.889	0. 114
Cardiovascular system & Coagulation regulation	0. 839	0.753	0.85	0. 149
Anemia	0. 830	0.771	0.833	0. 167
Skin	0. 723	0.661	0.710	0. 289
joint pain	0. 868	0.812	0. 886	0.114
acute kidney infection	0.898	0.667	0. 956	0.044
Physical Activity	0.873	0.897	0. 859	0.140

5. Discussion

Many individuals recovered from Covid 19 once the illness spread among people, but there remained a crucial question: Do patients who recover from Covid 19 experience any side effects? Whereas, through the questionnaire that was collected for 457 people during the first 6 months after infection, which includes age, gender, medical history, symptoms associated with infection, and information about the effects that they had after recovery. There were different effects on the recovered persons, as Physical Activity, Nervous system, Mental health, Anemia, Gastrointestinal system, Skin, Respiratory system, Cardiovascular & Coagulation regulation, Metabolism, Joint pain, and Acute kidney were affected, and their effect percentages were: 72.4%, 61.8%, 59%, 53.7%, 36.8%, 34.2%, 30.3%, 28.3%, 24.3%, 9.8%, and 5.0%, respectively. Since the goal of this study is to determine the impact of Covid 19 virus on persons after six months from recovery from it, the second purpose of this study is to build an intelligent model that can predict these impacts. Third, determining the needs of those who have contracted the virus as soon as possible, which ensures early detection of potential effects on those who are recovering, as well as determining the necessity of conducting pulmonary rehabilitation for those who have chronic respiratory conditions, which is intended to enhance their capacity for exercise and breathing.

A RF algorithm was used to predict the impact on the recovery people and evaluate this model using various metrics after performing multiple processing operations on the data and using the GSO algorithm to apply tuning hyper-parameter on RF algorithm, and used to perform the feature selection process in order to select the useful and influential features to obtain better accuracy at a faster time, where obtained good predictors result for all effects and in different proportions as shown in the preceding Table 1.

6.Conclusion

Using data from recovered individuals, including age, sex, medical history, symptoms of infection, and various details about the effects that occurred on people after recovery, an intelligent model using the RF classifier is proposed in this paper to predict the impacts within six months after recovery from COVID-19. This will then reveal the effects that Covid-19 had on persons after they had recovered. Different degrees of effects on the body's organs are observed in recovering patients. This model will help proactively determine how much care and follow-up patients need while they are injured. To identify and treat any residual or newly emerging long-term sequelae in affected and recovered individuals where follow-up and comprehensive assessment and early rehabilitation activities are required for these patients.

Reference

[1] P. Zhou et al., "A pneumonia outbreak associated with a new coronavirus of probable bat origin," Nature, vol. 579, no. 7798, pp. 270–273, Mar. 2020, doi: 10.1038/s41586-020-2012-7.

[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 coronavirus disease-2019 (COVID-19): The epidemic and the challenges," International Journal of Antimicrobial Agents, vol. 55, no. 3. Elsevier B.V., Mar. 01, 2020. doi: 10.1016/j.ijantimicag.2020.105924.

[3] M. H. N. Tayarani, X. Yao, and H. Xu, "Meta-Heuristic Algorithms in Car Engine Design: A Literature Survey," IEEE Transactions on Evolutionary Computation, vol. 19, no. 5, pp. 609–629, Oct. 2015, doi: 10.1109/TEVC.2014.2355174.

[4] D. C. Greenwood, R. O'connor, and M. Sivan, "LONG-TERM CLINICAL OUTCOMES IN SURVIVORS OF CORONAVIRUS OUTBREAKS AFTER HOSPITALISATION OR ICU ADMISSION: A SYSTEMATIC REVIEW AND META-ANALYSIS OF FOLLOW-UP STUDIES," 2020, doi: 10.1101/2020.04.16.20067975.

[5] C. Carvalho-Schneider et al., "Follow-up of adults with noncritical COVID-19 two months after symptom onset," Clinical Microbiology and Infection, vol. 27, no. 2, pp. 258–263, Feb. 2021, doi: 10.1016/j.cmi.2020.09.052.

[6] Y. miao Zhao et al., "Follow-up study of the pulmonary function and related physiological characteristics of COVID-19 survivors three months after recovery," EClinicalMedicine, vol. 25, Aug. 2020, doi: 10.1016/j.eclinm.2020.100463.

[7] C. Huang et al., "6-month consequences of COVID-19 in patients discharged from hospital: a cohort study," The Lancet, vol. 397, no. 10270, pp. 220–232, Jan. 2021, doi: 10.1016/S0140-6736(20)32656-8.

[8] A. K. Das, S. Mishra, and S. S. Gopalan, "Predicting CoVID-19 community mortality risk using machine learning and development of an online prognostic tool," PeerJ, vol. 8, Sep. 2020, doi: 10.7717/peerj.10083.

[9] L. A. Amar, A. A. Taha, and M. Y. Mohamed, "Prediction of the final size for COVID-19 epidemic using machine learning: A case study of Egypt," Infect Dis Model, vol. 5, pp. 622–634, Jan. 2020, doi: 10.1016/j.idm.2020.08.008.

[10] G. Pinter, I. Felde, A. Mosavi, P. Ghamisi, and R. Gloaguen, "COVID-19 pandemic prediction for Hungary; A hybrid machine learning approach," Mathematics, vol. 8, no. 6, 2020, doi: 10.3390/math8060890.

[11] R. Tamhane and S. Mulge, "Prediction of COVID-19 Outbreak using Machine Learning," International Research Journal of Engineering and Technology, 2020, [Online]. Available: www.irjet.net

[12] H. Ko et al., "An Artificial Intelligence Model to Predict the Mortality of COVID-19 Patients at Hospital Admission Time Using Routine Blood Samples: Development and Validation of an Ensemble Model," Journal of Medical Internet Research, vol. 22, no. 12, p. e25442, Dec. 2020, doi: 10.2196/25442.

[13] J. Sarkar and P. Chakrabarti, "A Machine Learning Model Reveals Older Age and Delayed Hospitalization as Predictors of Mortality in Patients with COVID-19," medRxiv, 2020, doi: 10.1101/2020.03.25.20043331.

[14] L. Yan et al., "An interpretable mortality prediction model for COVID-19 patients," Nature Machine Intelligence, vol. 2, no. 5, pp. 283–288, May 2020, doi: 10.1038/s42256-020-0180-7.

[15] R. O. Ogundokun and J. B. Awotunde, "MACHINE LEARNING PREDICTION FOR COVID 19 PANDEMIC IN INDIA 1, *," medRxiv, 2020, doi: 10.1101/2020.05.20.20107847.

[16] C. Iwendi et al., "COVID-19 patient health prediction using boosted random forest algorithm," Frontiers in Public Health, vol. 8, Jul. 2020, doi: 10.3389/fpubh.2020.00357.

[17] World Health organization, "Coronavirus disease (COVID-19): Staying active," World Health organization, Mar. 27, 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19-staying-active (accessed Jun. 11, 2022).

[18] Julia Evangelou Strait, "Among COVID-19 survivors, an increased risk of death, serious illness," Washington University school of medicine in St. Louis., Apr. 22, 2021. https://medicine.wustl.edu/news/among-covid-19-survivors-an-increased-risk-of-death-serious-illness/ (accessed Jun. 11, 2022).

[19] G. Aksu, C. O. Güzeller, and M. T. Eser, "The Effect of the Normalization Method Used in Different Sample Sizes on the Success of Artificial Neural Network Model," International Journal of Assessment Tools in Education, pp. 170–192, Apr. 2019, doi: 10.21449/ijate.479404.

[20] E. Gibaja and S. Ventura, "A tutorial on multilabel learning," ACM Computing Surveys, vol. 47, no. 3, Apr. 2015, doi: 10.1145/2716262.

[21] M.S. Kraiem, F. Sánchez-Hernández, and M.N. Moreno-García, "Selecting the Suitable Resampling Strategy for Imbalanced Data Classification Regarding Dataset Properties," An Approach Based on Association Models, Applied Sciences, 11(18): p. 8546, 2021.

[22] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results," in 2020 11th International Conference on Information and Communication Systems, ICICS 2020, Apr. 2020, pp. 243–248. doi: 10.1109/ICICS49469.2020.239556.

[23] T. Saw and H. Myint, "Swarm Intelligence Based Feature Selection for High Dimensional Classification: A Literature Survey," International Journal of Computer, 2019, [Online]. Available: http://ijcjournal.org/

[24] D. Kaipa Krishnanand N. and Ghose, "Glowworm Swarm Optimization: Algorithm Development," in Glowworm Swarm Optimization : Theory, Algorithms, and Applications, Cham: Springer International Publishing, 2017, pp. 21–56. doi: 10.1007/978-3-319-51595-3_2.

[25] R. K. Tan and S. Bora, "Parameter tuning in modeling and simulations by using swarm intelligence optimization algorithms," in Proceedings - 9th International Conference on Computational Intelligence and Communication Networks, CICN 2017, Mar. 2018, vol. 2018-January, pp. 148–152. doi: 10.1109/CICN.2017.8319375.

التنبؤ بتأثير COVID-19 على المرضى خلال ستة أشهر بعد التعافي، باستخدام خوارزميات الذكاء الاصطناعي

حنان انس قاسم
المديرية العامة للتربية في محافظة نينوى، الموصل، العراق
hanan.csp9@student.uomosul.edu.iq1

غيداء عبدالعزيز الطالب
قسم علوم الحاسوب، كلية علوم الحاسوب والرياضيات، جامعة الموصل، الموصل، العراق
ghaydabdulaziz@uomosul.edu.iq2

تاريخ الاستلام :14/9/2022 **تاريخ القبول:17/1/2023**

الملخص

قد عانى الناس في جميع أنحاء العالم من ذعر شديد نتيجة لفيروس كورونا الجديد (COVID-19). حيث انه من الضروري إجراء تقييمات شاملة للناس في مجالاتهم النفسية والجسدية والاجتماعية وذلك لتأمين مستقبلهم وفهم النتائج المحتملة للمرضى الذين يتعافون من هذا المرض والتأكد مما إذا كان لديهم أي أمراض أخرى. تم التنبؤ بهذه النتيجة المحتملة للأشخاص الذين تعافوا من فيروس Covid 19 من خلال عملية جمع البيانات من الأشخاص الذين أصيبوا سابقًا بهذا الفيروس لتحديد الآثار التي تعرضوا لها، باستخدام تقنيات ذكائية. حيث تم استخدام خوارزمية GSO لاختيار الميزات، ولضبط المعلمات المفرطة لخوارزمية SVM المستخدمة في عملية التنبؤ وذلك لتحديد الآثار التي قد تظهر على الأشخاص المتعافون. سيساعد هذا في التنبؤ بالنتائج المحتملة لتقديم المشورة والدعم النفسي، بالإضافة إلى بعض الإرشادات الموصي بها لشفاء المرضى والمجتمع للعودة إلى الحياة الطبيعية.

الكلمات المفتاحية : التنبؤ ، كوفيد19 ، النشاط البدني ، الذكاء الاصطناعي ، التعلم الآلي ، تحسين سرب الدودة المضيئة.