Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Mental health and long COVID status prediction among recovered COVID-19 patients: A comparison of machine learning methods

Tran Anh Tuan[a,b*], Win Win Myo[c] , Le Thanh Thao Trang[d], Nguyen Thi The Nhan[e], Tran Dai An[f], Dao Thi Thanh Loan[g]

[a]*School of Informatics, Walailak University, Nakhon Si Thammarat, Thailand*
[b]*Informatics Innovation Center of Excellence (IICE), Walailak University, Nakhon Si Thammarat, Thailand*
[c]*University of Information Technology, Yangon, Myanmar*
[d]*Dong Thap Medical College, Dong Thap, Vietnam*
[e]*Pasteur Institute in Ho Chi Minh city, Ho Chi Minh city, Vietnam*
[f]*Dong Thap Provincial Center for Disease Control, Dong Thap, Vietnam*
[g]*Dak Lak College of Pedagogy, Dak Lak, Vietnam*

C H R O N I C L E

Article history:
Received: January 2, 2024
Received in revised format: March 18, 2024
Accepted: May 23, 2024
Available online: May 23, 2024

Keywords:
Predictive Model
Machine Learning
Mental health
Long COVID
COVID-19

A B S T R A C T

The COVID-19 pandemic has led to different health outcomes, including long COVID (LCo) and mental health (MH) disorders, impacting millions globally. To enable early healthcare diagnosis, including the prediction of MH conditions and LCo, various research studies have utilized machine learning (ML) techniques. However, there is still a gap in understanding the mental health of recovered COVID-19 patients with long COVID using ML techniques. This study aims to bridge this gap by developing and evaluating ML models, including support vector machine, multilayer perceptron (MLP), k-nearest neighbor, gradient boosting, voting classifier, and extreme gradient boosting, tailored for mental health and long COVID datasets from recovered COVID-19 patients. Additionally, feature selection methods, e.g., Recursive Feature Elimination (RFE) and Extra Trees (ET), and optimized models with hyper-parameter tuning will be employed. Our experiments utilize the dataset of recovered COVID-19 patients. Among these ML models, the MLP with ET-based features achieved the highest accuracy and AUC scores in this dataset, with 1.00 and 0.97 ± 0.02, respectively. The research reveals the high prevalence and risk factors of mental health disorders and long COVID from the dataset. These findings will contribute to personalized healthcare strategies for individuals navigating the complexities of post-COVID-19 recovery, integrating machine learning insights into mental health and long COVID support.

1. Introduction

The COVID-19 epidemic has had a notable impact on every activity and everyone worldwide (Almajali & Masadeh, 2021; Elnagar et al., 2022; Kanrak & Nonthapot, 2024), incredibly individual health and well-being, resulting in a range of outcomes from asymptomatic cases to severe complications such as multiorgan failure and death (Zhang et al., 2024). As of March 31st, 2024, over 774 million verified infections and seven million fatalities were globally reported (WHO, 2024). Among these cases, it is estimated that at least 65 million COVID-19 patients worldwide suffer from a condition known as long COVID (Davis et al., 2023). Long COVID is characterized by persistent symptoms such as fatigue, shortness of breath, and cognitive issues that can last for months after the initial COVID infection (Engel et al., 2023; Pei et al., 2021; WHO, 2022). Recent studies have also examined the association of mental health disorders with COVID-19 outcomes, revealing that individuals having mental health disorders face a higher risk of COVID-19 admittance and fatality (Linh et al., 2024; Phu et al., 2023; Zhang et al., 2024). As recovered COVID-19 patients navigate post-recovery life, understanding and predicting their mental

* Corresponding author.
E-mail address tran.tu@wu.ac.th (T. A. Tuan)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2024.5.018

health and potential long COVID symptoms become paramount for effective healthcare management (De Oliveira Almeida et al., 2023; Islam et al., 2024; Lopez-Leon et al., 2021; Rass et al., 2021).

Machine learning (ML) methods have recently shown potential in healthcare applications, including mental health prediction and disease prognosis. Numerous studies have investigated the application of ML algorithms to forecast mental health disorders, e.g., depression, anxiety, and stress, for various datasets (Cho et al., 2021; Chung & Teo, 2023; Jha et al., 2022; Juliet & others, 2023; Katiyar et al., 2024; Kim & Chang, 2023; Ku & Min, 2024; Malik & Khan, 2023; Nasir et al., 2024; Nison et al., 2023; Pramodhani et al., 2022; Priya et al., 2020; Qasrawi et al., 2022; Rezapour & Hansen, 2022; Singh et al., 2022; Trivedi et al., 2022; Tuan et al., 2024; Tyshchenko, 2018; Vaishnavi et al., 2022). Additionally, ML techniques have been employed to identify and forecast long COVID symptoms, COVID-19-related outcomes and their impact on recovered patients (Afrash et al., 2022; Arshed et al., 2021; Chadaga et al., 2024; Gupta et al., 2022; Hossen et al., 2024a; Hussein et al., 2023; Islam et al., 2024; Patel et al., 2023; Patterson et al., 2021; Pfaff et al., 2022; Prout et al., 2020; Sarmiento Varón et al., 2023; Yazdani et al., 2022; Zoabi et al., 2021).

However, there is a shortage of research investigating the mental health of individuals with long COVID who have recovered from COVID-19 using ML techniques. Additionally, recent studies have shown that there is a high incidence of mental health issues (depression, anxiety, and stress) and long COVID in recovered COVID-19 patients. Therefore, our research aims to explore and identify an ML model tailored for mental health and long COVID datasets from recovered COVID-19 patients. This research employs Multilayer Perceptron (MLP), K-nearest neighbor (KNN), Support Vector Machine (SVM), Gradient Boosting, Voting Classifier, and Extreme Gradient Boosting (XGBoost) to predict the status of mental health and long COVID from recovered COVID-19 patients. We employ two feature selection methods - Recursive Feature Elimination –(RFE) and Extra Trees (ET) - and use the Decision Tree (DT) algorithm to assess subsets of features yielding high accuracy. Hyperparameter tuning is also conducted to identify the optimal model. In assessments, we use 10-fold cross-validation.

2. Related Work

In this section, previous studies related to mental health and COVID-19 prediction (including long COVID symptoms) using ML will be discussed.

2.1. Mental health prediction with machine learning methods

In this part, numerous studies have delved into ML on various datasets of mental health will be discussed.

For datasets of the general population (including pregnant women and older people), Cho et al. (2021) centered their study on ML-based depression prediction among community dwellers, utilizing data from the Korea National Health and Nutrition Examination Surveys in 2014 and 2016. The authors emphasized employing techniques such as SMOTE and LASSO for feature reduction, which resulted in good accuracy and an area under the receiver operating characteristic curve. Nguyen and Byeon (2022) introduced a deep neural network (DNN) model to predict depression in elderly individuals amid the COVID-19 pandemic. The authors' explainable DNN model incorporated LIME for prediction explainability, showcasing notable precision and recall, facilitating early identification of treatment needs. Qasrawi et al. (2022) developed ML models that accurately predict depression and anxiety in pregnant and postpartum women during the COVID-19 pandemic. The models identified key predictors such as stress during pregnancy, family and social support, financial issues, and income. These findings highlight the crucial role of early detection in maintaining maternal mental health. Singh et al. (2022) investigated the efficacy of ML models in predicting depression, anxiety, and stress levels using survey responses and demographic data. The study found that the SVM model performed better than other models in accurately predicting these mental health outcomes, with a good F1-score. Kumar et al. (2023) aimed to tackle mental health challenges arising from the COVID-19 pandemic by comparing anxiety, stress, and depression levels, along with psychological effects, between Rajyoga meditators and the general population. Kim and Change (2023) utilized data from 5,420 participants in the 2020 Korea National Health and Nutrition Examination Survey (KNHANES), with 4,138 participants not having depression and 1,282 having depression. The authors developed and evaluated three ML algorithms: random forest (RF), logistic regression (LR), and DNN. The LR model had the highest area under the curve (AUC). Dhariwal et al. (2024) employed AI techniques to analyze neurological disorder prevalence using the Cities Health Initiative dataset, achieving high accuracy with models like Convolutional Neural Network (CNN) and XGBoost. The findings supported healthcare professionals in diagnosing disorders promptly and offered insights for policymakers on pollution and addiction regulations. Katiyar et al. (2024) conducted ML to predict serious mental health issues, e.g., anxiety, depression, and postpartum depression (PPD), especially impacting women due to socio-economic stressors. ML algorithms such as gradient boosting, RF, and deep recurrent neural network (DRNN) offer accurate predictions, aiding in effective clinical strategies and reducing self-harm.

For datasets of patients, Trivedi et al. (2022) investigated depression as a significant concern in mental health, impacting individuals across ages and genders equally, attributed to factors, e.g., workplace dynamics, everyday stressors, and challenges in interpersonal relationships. Their study evaluated and contrasted the performance of various ML algorithms in predicting depressive episodes, highlighting neural network algorithms that achieved notable classification accuracy. Tuan et al. (2024)

presented three ML approaches with optimized hyperparameters to predict mental health among recovered COVID-19 individuals. Their findings highlight SVM as the most fitting model for precise mental health predictions in this group.

For datasets of college students, Malik et al. (2023) used ML to predict anxiety, stress, and depression severity from 400 students. The authors conducted various algorithms, with KNN emerging as the most effective, followed by LR. Nison et al. (2023) presented a ML strategy to screen college students for depression using non-clinical data, achieving an average predictive accuracy of up. By leveraging general information such as demographics, physical health, and relationships without direct mental health queries, the study aimed to enhance classification accuracy, overcoming the drawbacks of conventional depression screening methods. Ku and Min (2024) applied ML algorithms to predict depression and anxiety using data from 4,184 UNSA students, evaluating CNN along with four other ML models for accuracy under varied response levels. CNN showed superior resilience and accuracy, especially in handling biased or inaccurate responses, suggesting its efficacy in mental health prediction with self-reported data.

For datasets of workers in technology and health fields, Prout et al. (2020) sought to pinpoint factors predicting psychological distress amid COVID-19, revealing elevated distress levels in younger individuals, women, and non-binary individuals. Employing ML methods, the authors emphasized somatization and reduced reliance on adaptive defense mechanisms as robust predictors of distress during a global health crisis. Priya et al. (2020) used ML to assess anxiety, depression, and stress severity using data from a standard questionnaire (DASS-21). The authors applied DT, RF, Naïve Bayes (NB), SVM, and KNN, with NB showing the highest accuracy. However, RF was identified as the best model due to class imbalance, and it was evaluated using the F1-score. Jha et al. (2022) utilized ML algorithms to predict depression and anxiety prevalence using DASS42 and WESAD datasets. These algorithms, including probabilistic, nearest neighbor, neural network, and tree-based methods, were used to identify and improve patient care early. The experimental results were in prediction accuracy and aiming for even better outcomes in patient recovery. Juliet et al. (2023) explored ML's efficacy in mental health analysis and prediction, aiming to identify the most accurate model. The authors investigated KNN, SVM, RF, LR, and DT models, comparing their accuracy for mental health prediction. Chung and Teo (2023) assessed the effectiveness of ML algorithms in forecasting mental health concerns using an OSMI dataset. Among these algorithms, Gradient Boosting stood out with an impressive accuracy rate, showcasing the potential for automated mental health diagnostics.

2.2. *COVID-19 prediction with machine learning methods*

In this part, studies have employed ML on various COVID-19 prediction (including long COVID symptoms) will be discussed. For long COVID symptoms prediction, Patterson et al. (2021) classified long COVID patients using the RF method from Immunologic profiles with 224 individuals. The model achieved high accuracy and F1-score, indicating strong performance. Sudre et al. (2021) applied the RF method to classify short and long COVID symptoms from a dataset of users' self-reports collected from mobile health applications. The model was evaluated with five-fold cross-validation and AUC. Jiang et al. (2022) used three ML models to classify long COVID from the N3C dataset. In terms of AUC, the best model among these was the XGBoost model. Pfaff et al. (2022) used the XGBoost method, and the results showed that the model was highly accurate in classifying long COVID-19 patients. Gupta et al. (2022) developed a technique for early cardiac problem diagnosis in COVID-19 survivors to predict long COVID. The authors employed an ensemble approach using a stacked model trained on data from 180 COVID-19 patients (heart-related survey). The result of heart disease prediction was high in terms of accuracy. Patel et al. (2023) introduced the RF method to identify the most relevant blood proteins for long COVID cases. The 3-fold cross-validation was used to assess the model, which outperformed with the highest accuracy, F1-score. Isalam et al. (2024) investigated the impact of COVID-19 on COVID-recovered patients in Bangladesh through statistical analysis and ML techniques. The authors identified vital post-COVID-19 health factors and their correlations, highlighting the DT model's effectiveness in predicting post-COVID-19 outcomes. For COVID-19 prediction, Yazdani et al. (2022) used ML algorithms for early detection of COVID-19 in patients. The study results showed that the two methods, SVM and C4.5, achieved the highest performance. Afrash et al. (2022) used a model combining genetic algorithm (GA) with ML methods to predict hospital mortality due to COVID-19. The GA-SVM model predicted severe disease and high outcome mortality and optimized medical resources to improve the treatment of COVID-19 patients. In the study of Sarmiento et al. (2023), the authors explored ML methods, and the results showed that ML played a vital role in developing diagnostic and behavioral prediction strategies, epidemiology and supported the development and monitoring policies of public health during a global epidemic. Hussein et al. (2023) proposed an enhanced KNN algorithm, eKNN, which showed improved performance in detecting COVID-19 compared to the traditional KNN algorithm. Nasir et al. (2024) used different ML models (LR, RF, and XGBoost) to predict COVID-19 from one million patients in the European Commission (EC) dataset. The experimental results showed that the best accuracy was achieved using the RF and XGBoost methods. Hossen et al. (2024b) proposed a system that combined ensemble feature selection methods with ML classifiers for effective COVID-19 infection identification. The authors evaluated classifiers such as DT, NB, KNN, MLP, and SVM using two COVID-19 datasets to analyze their performance with ensemble feature selection methods, which yielded similar an accuracy score for classification across the symptoms and COVID-19 presence dataset.

In summary, there is a gap in research regarding mental health and long COVID prediction in recovered COVID-19 patients, lacking integrated ML models that can forecast both mental health disorders and long COVID symptoms. Existing studies have focused on mental health or COVID-19 prediction but have not bridged these domains. Integrating these predictions is crucial for a comprehensive understanding of post-COVID-19 health challenges and can inform personalized healthcare strategies. Our research has the following contributions: 1) through statistical analysis, we understand the prevalence and association between mental health and long COVID, providing valuable insights; 2) we have the outstanding ML model from KNN, MLP, SVM, Gradient Boosting, Voting Classifier, and XGBoost to predict mental health conditions and long COVID status accurately.

3. Method

3.1. Data Collection

The dataset used in our study comprises information collected from 549 participants who had experienced and recovered from COVID-19 (Trang et al., 2023). These participants had been discharged from the hospital six months earlier in Dong Thap, Vietnam. Prior to its implementation, the dataset procedure received approval from the Research Ethics Committee for Human, at Walailak University (Ref: WU-EC-PU-0-214-65).

The dataset consisted of a structured questionnaire with three sections. The first section collected sociodemographic information including gender, age, occupation, marital status, monthly income, education level along with additional data on participants' height, weight, underlying diseases, and COVID-related information. Body mass index (BMI) calculation followed WHO guidelines using the weight-to-height squared ratio (WHO, 2000). In the second section, the 21-item of DAS scale (DASS-21) was employed to assess participants' emotional states related to depression, anxiety, and stress, with ratings on a 4-point scale leading to 5 categorizations ("normal", "mild", "moderate", "severe", "extremely severe") for each symptom. For depression, the levels are defined as follows: normal (0-9), mild (10-13), moderate (14-20), severe (21-27), and extremely severe (≥ 28). Similarly, for anxiety, the levels are delineated as normal (0-7), mild (8-9), moderate (10-14), severe (15-19), and extremely severe (≥ 20). In terms of stress, the levels are categorized as normal (0-14), mild (15-18), moderate (19-25), severe (26-33), and extremely severe (≥ 34) (Lovibond, 1995). Participants are then grouped into "normal", "mild", "moderate", "severe", or "extremely severe" categories for each symptom based on their scores. Participants who scored within the "mild" to "extremely severe" range were classified as experiencing mental health symptoms. The third section addressed 13 long COVID-19 symptoms identified through literature reviews and surveys. Participants indicated yes/no for symptoms (e.g., cough, fatigue, diarrhea) experienced six months after discharge from the hospital following COVID-19 infection. Those reporting at least one symptom were categorized as having long COVID symptoms (Aiyegbusi et al., 2021; Saltzman et al., 2023).

3.2. The Proposed Method

The framework based on ML for detecting MH issues and LCo symptoms of recovered COVID-19 patients is depicted in Fig. 1. Three significant phases in the framework are (1) "*data pre-processing*", (2) "*feature selection*", and (3) "*classification and evaluation*". The input of the framework is a dataset of mental health and long COVID symptoms of recovered COVID-19 patients, and the output is various classes.

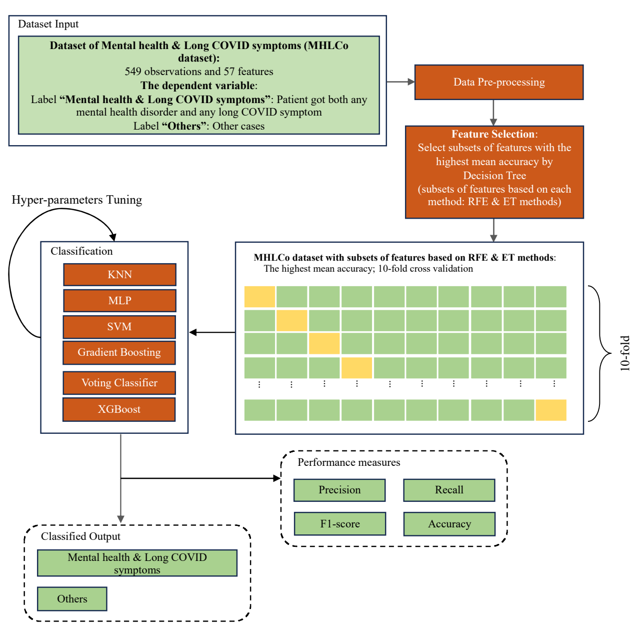


Fig. 1. Workflow of ML method for mental health and long COVID symptoms detection

3.2.1. Data pre-processing

The dataset, named the Mental Health and Long COVID Symptoms (MHLCo), includes responses to 21 questions (DASS-21), information on 13 long COVID symptoms, and sociodemographic details of individuals recovered from COVID-19. The MHLCo dataset consists of 549 rows and 57 columns (Trang et al., 2023). It encompasses diverse variable types, including ordinal and categorical variables. These variables are encoded and normalized in this phase using the Scikit-learn library within Google Colab (Scikit-learn, 2024).

3.2.2. Feature Selection

Selecting features plays an essential role in enhancing the performance of ML models (Guyon & Elisseeff, 2003). The feature selection procedure in the framework involves the following steps: 1) ranking features based on importance using the RFE method (Guyon et al., 2002) and the ET method (Geurts et al., 2006), 2) selecting subsets of features in each ranking method where features with higher rankings take priority, 3) evaluating each subset of features in each ranking method (RFE and ET) with 10-fold cross-validation using the DT method, and 4) choosing the subset with the highest mean accuracy in each ranking method. Note that each subset of features contains at least two features. In cases where more than two subsets exhibit identical accuracy, the subset with fewer features is selected.

3.2.3. Classification

There are various ML algorithms, e.g., KNN, MLP, SVM, Gradient Boosting, Voting Classifier, and XGBoost, which have demonstrated strong performance in classifying health-related data (Aljrees, 2024; Batool & Byun, 2024; Chadaga et al., 2024; Chen & Guestrin, 2016; Chung & Teo, 2023; Saha et al., 2024; Tuan et al., 2024). In this research, these ML algorithms are employed to predict the status of mental health and long COVID-19 symptoms.

Achieving optimal results in ML models requires careful optimization of various hyperparameters to avoid overfitting (Bergstra et al., 2011). Therefore, this study employed a grid search approach to explore values of potential hyperparameters for each model. The objective was to identify the best possible combinations to achieve the highest accuracy. Table 1 presents a comprehensive list of the hyperparameters utilized in this optimization process.

Table 1
Hyper-parameters for ML models

ML model	Hyperparameters
KNN	• Option 1: metric: [‘minkowski’] • Option 2: metric: [‘minkowski’], n_neighbors: [5, 10, 20, 50] • Option 3: n_neighbors: [5, 10, 20, 50], algorithm: [‘auto’, ‘ball_tree’, ‘kd_tree’, ‘brute’]
MLP	• Option 1: activation: [‘identity’, ‘logistic’, ‘tanh’, ‘relu’], hidden_layer_sizes: [10, 20, 50, 100] • Option 2: activation: [‘identity’, ‘logistic’, ‘tanh’, ‘relu’], hidden_layer_sizes: [10, 20, 50, 100], solver: [‘lbfgs’, ‘sgd’,‘adam’]
SVM	• Option 1: kernel: [‘linear’], C: [1, 10, 20] • Option 2: kernel: [‘rbf’], C: [1, 10, 20], gamma: [0.1, 0.01, 0.02] • Option 3: kernel: [‘poly’], gamma: [0.1, 0.01, 0.02], degree: [1, 10, 20]
Gradient Boosting	• Option 1: n_estimators: [50, 100, 150], max_depth: [3, 5, 7], min_samples_split: [2, 5, 10] • Option 2: n_estimators: [50, 100, 150], max_depth: [3, 5, 7], min_samples_split: [2, 5, 10], min_samples_leaf: [1, 2, 4], criterion: [‘friedman_mse, ‘mse’, ‘mae’]
Voting Classifier	• Option 1: rf__n_estimators: [50, 100], gb__n_estimators: [50, 100], lr__C: [0.1, 1.0] • Option 2: rf__n_estimators: [50, 100], gb__n_estimators: [50, 100], lr__C: [0.1, 1.0], voting : [‘hard’, ‘soft’]
XGBoost	• Option 1: max_depth: [3, 4, 5], learning_rate: [0.1, 0.01, 0.001] • Option 2: max_depth: [3, 4, 5], learning_rate: [0.1, 0.01, 0.001], n_estimators: [100, 200, 300], subsample: [0.8, 0.9, 1.0], colsample_bytree: [0.8, 0.9, 1.0]

The hyper-parameters for the six ML models are presented in Table 1, with values assigned to each option derived from existing studies. Each ML model comprises various options, each with multiple parameters that are automatically selected and generated in different combinations. For KNN, the ‘*metric*’ parameter calculates distance, ‘*n_neighbors*’ determines the number of neighbors, and the ‘*algorithm*’ parameter specifies the nearest neighbor algorithm. Similarly, in MLP, ‘*hidden_layer_sizes*’ denotes the neurons in each hidden layer, ‘*activation*’ represents the activation function in these layers, and ‘*solver*’ decides the optimization algorithm for weight adjustment during training. In SVM, the ‘*C*’ parameter serves as a regularization factor affecting decision boundary smoothness and accurate classification. Additionally, ‘kernel’ defines the kernel function type, ‘*gamma*’ controls single training example influence, and ‘*degree*’ applies to polynomial kernel functions.

For the Gradient Boosting, '*n_estimators*' sets the boosting stages, '*max_depth*' limits tree nodes, '*min_samples_split*' determines node split samples, and '*criterion*' measures split quality. In the Voting Classifier, '*voting*' predicts class labels, '*rf__n_estimators*' and '*gb__n_estimators*' are the '*n_estimators*' for RF and Gradient Boosting classifiers, '*lr__C*' is LR's regularization strength. In XGBoost, '*max_depth*' limits tree nodes, '*learning_rate*' shrinks each tree's contribution, '*n_estimators*' sets boosting stages, '*subsample*' is the training instance subsample ratio, and '*colsample_bytree*' is the column subsample ratio when building each tree. For example, a KNN option could be expressed as metric: ['minkowski'], n_neighbors: [5, 10, 20, 50], resulting in combinations {metric: 'minkowski', n_neighbors: 5}, {metric: 'minkowski', n_neighbors: 10}, {metric: 'minkowski', n_neighbors: 20}, {metric: 'minkowski', n_neighbors: 50}. The hyperparameters with the highest mean accuracy are automatically selected for each model.

3.2.4. Performance measures

To evaluate and select the best-performing model in classification tasks, we utilize well-known measurement metrics, including precision (P), recall (R), F1-score, and accuracy with 10-fold (Aljrees, 2024; Batool & Byun, 2024; Saha et al., 2024; Singh et al., 2022; Vaishnavi et al., 2022; Zoabi et al., 2021). These metrics are calculated as follows:

$$\mathrm{P} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}) \tag{1}$$

$$\mathrm{R} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \tag{2}$$

$$\mathrm{F1} - \mathrm{score} = (2 \times \mathrm{P} \times \mathrm{R})/(\mathrm{P} + \mathrm{R}) \tag{3}$$

$$\mathrm{Accuracy} = \mathrm{TP} + \mathrm{TN}/(\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}) \tag{4}$$

where TP (True Positive) indicates instances that are predicted as positive and are actually positive, FP (False Positive) refers to instances that are predicted as positive but are actually negative, TN (True Negative) denotes instances that are predicted as negative and are actually negative, and FN (False Negative) represents instances that are predicted as negative but are actually positive.

3.2.5. Classified output of mental health status with long COVID symptoms among recovered COVID-19 patients

The classification output represents the status of recovered COVID-19 patients with mental health and long-term COVID symptoms. Two distinct statuses were identified: observations with experiences of both mental health and long COVID symptoms were categorized as "Mental health & long COVID symptoms" status, while the remaining cases were considered "Others".

4. Result

4.1. Characteristics of dataset

The attributes divided into three sections were (1) socio-demographic information, underlying diseases, and COVID-19 details include *gender, age, areas, education, marital status, monthly income, employment status, family infected, hypertension, diabetes, heart disease, cancer, respiratory disease, kidney disease, other diseases, non-communicable diseases} (No_NCDs), COVID classification, COVID treatment, number of days in hospital, number of infected times, sleeping hours, and body mass index (BMI);* (2) mental health includes 21 questions (DASS-21); and (3) long COVID symptoms are *cough, chest pain, headache, shortness of breath, dizziness, fatigue, joint pain, diarrhea, change in taste, change in smell, amnesia, insomnia, other symptoms, no symptoms*.

We summarized 21 questions of DASS-21 (Q1 to Q21) with keywords as depicted in the following*: difficulty winding down (Q1), mouth dryness awareness (Q2), lack of positive feelings (Q3), breathing difficulty (Q4), difficulty initiating tasks (Q5), tendency to overreact (Q6), trembling (Q7), feeling of using nervous energy (Q8), worry about social panic (Q9), lack of anticipation (Q10), agitation meaningless (Q11), difficulty relaxing (Q12), feeling down-hearted (Q13), intolerance to interruptions (Q14), proximity to panic (Q15), lack of enthusiasm (Q16), low self-worth (Q17), sensitivity or touchiness (Q18), heart awareness energy (Q19), unexplained fear (Q20), sense of life being (Q21).*

4.2. Statistical analysis

4.2.1. Depression, anxiety, and stress among recovered COVID-19 patients

Out of 549 participants, 136 (24.8%) were diagnosed with depression, ranging in severity from mild to extremely severe (n=60 and n=11, respectively). Anxiety was identified in 228 (41.5%) participants, who experienced varying levels of severity, including mild (n=81), moderate (n=88), severe (n=33), and extremely severe (n=26). Stress was reported by 139 participants (25.3%), primarily at mild levels (n=67), followed by moderate (n=45), severe (n=23), and extremely severe (n=4). Among those who had recovered from COVID-19, 292 (53.2%) were diagnosed with at least one mental health symptom, including depression, anxiety, and stress. Table 2 presents the measurements of depression, anxiety, and stress experienced by COVID-19 patients who have recovered.

Table 2
The levels of depression, anxiety and stress among recovered COVID-19 patients

Levels	Number of participant (n = 549) (%)			
	Depression	**Anxiety**	**Stress**	**Any mental health**
Normal	413 (75.23)	321 (58.47)	410 (74.68)	257 (46.81)
Rate from mild to Extremely Severe	136 (24.77)	228 (41.53)	139 (25.32)	292 (53.19)
Mild	60 (10.93)	81 (14.75)	67 (12.20)	-
Moderate	49 (8.93)	88 (16.03)	45 (8.20)	-
Severe	16 (2.91)	33 (6.01)	23 (4.19)	-
Extremely Severe	11 (2.00)	26 (4.74)	4 (0.73)	-

4.2.2. The characteristics of long COVID symptoms among recovered COVID-19 patients

A total of 549 participants who experienced long COVID symptoms reported all thirteen symptoms, comprising twelve significant symptoms and minority symptoms represented under "Other symptoms". Most participants reported two symptoms, representing the median value, with an interquartile range (IQR) of one to three. Among recovered COVID-19 patients experiencing long COVID, more than one-third of participants reported amnesia (49.2%, SE ± 0.02) and fatigue (34.8%, SE ± 0.02), with median standardized scores of 0.0 [IQR: 0–33.3] and 0.0 [IQR: 0–20.0], respectively (see Fig. 2). The remaining symptoms of long COVID were reported less frequently, with median standardized scores of zero. The reported percentages of participants experiencing these symptoms were as follows: cough (26.2%, SE ± 0.02), joint pain (20.0%, SE ± 0.02), headache (18.8%, SE ± 0.02), insomnia (18.6%, SE ± 0.02), dizziness (13.8%, SE ± 0.01), shortness of breath (13.1%, SE ± 0.01), chest pain (8.4%, SE ± 0.01), diarrhea (2.0%, SE ± 0.01), change in smell (2.0%, SE ± 0.01), change in taste (2.0%, SE ± 0.01), and other symptoms (3.0%, SE ± 0.01).

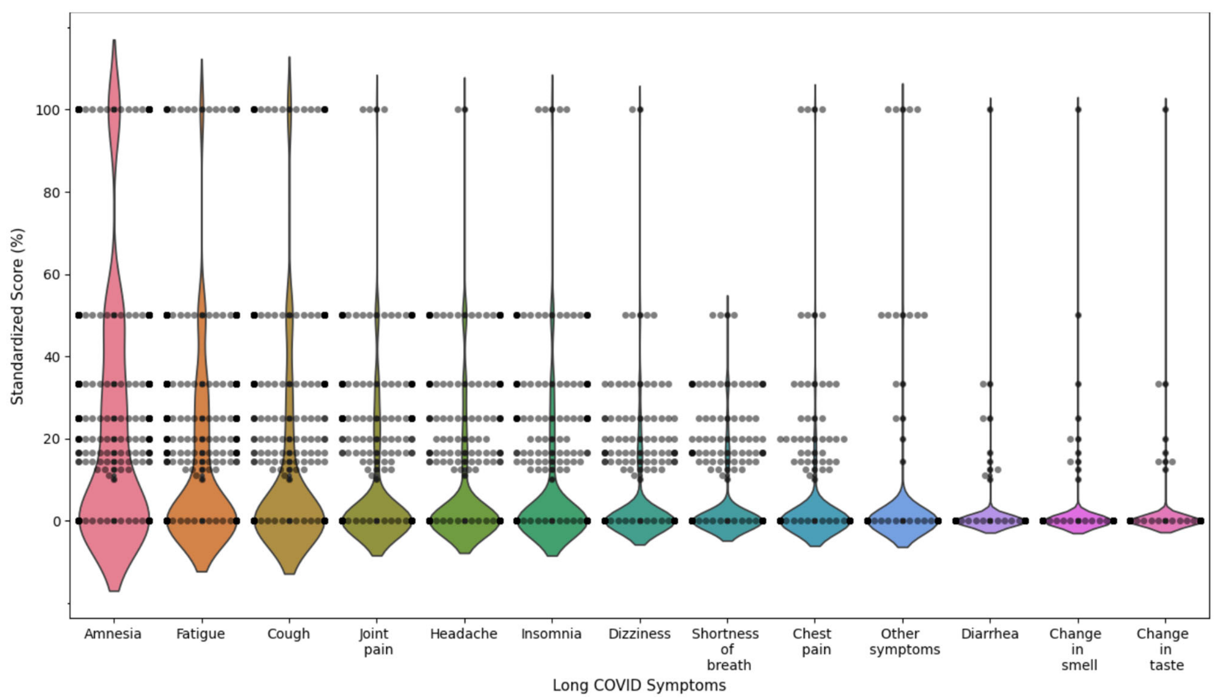


Fig. 2. Standardized scores for each long COVID symptom listed by participants

4.2.3. Association between mental health and long COVID symptoms among recovered COVID-19 patients

In our research, we analysed the association between mental health issues and the set of 13 long COVID symptoms, specifically focusing on depression, anxiety, and stress among recovered COVID-19 patients. The outcomes of our investigation revealed a notable correlation between mental health issues and all long COVID symptoms in those who had recovered from COVID-19. Individuals who encountered any of the long COVID symptoms faced an infinite increase in the risk of depression and stress, and a 319-fold rise in the risk of anxiety (all $p < 0.001$). Moreover, participants who reported any of the long COVID symptoms faced a 255-fold escalation in the risk of experiencing any mental health disorder ($p < 0.001$). The symptoms posing the highest risk for depression were amnesia (OR = 2.89) and fatigue (OR = 1.71) (all $p < 0.001$). Leading risk factors for anxiety included amnesia (OR = 1.35) and fatigue (OR = 0.71) (all $p < 0.05$). Similarly, for stress, the most significant risk factors were amnesia (OR = 2.98) and fatigue (OR = 1.65) (all $p < 0.001$). Additional details regarding the relationship between mental health disorder (depression, anxiety, and stress) status and all long COVID symptoms can be found in Table 3.

Table 3
The association between depression, anxiety and stress and long COVID symptoms among recovered COVID-19 patients (n = 549)

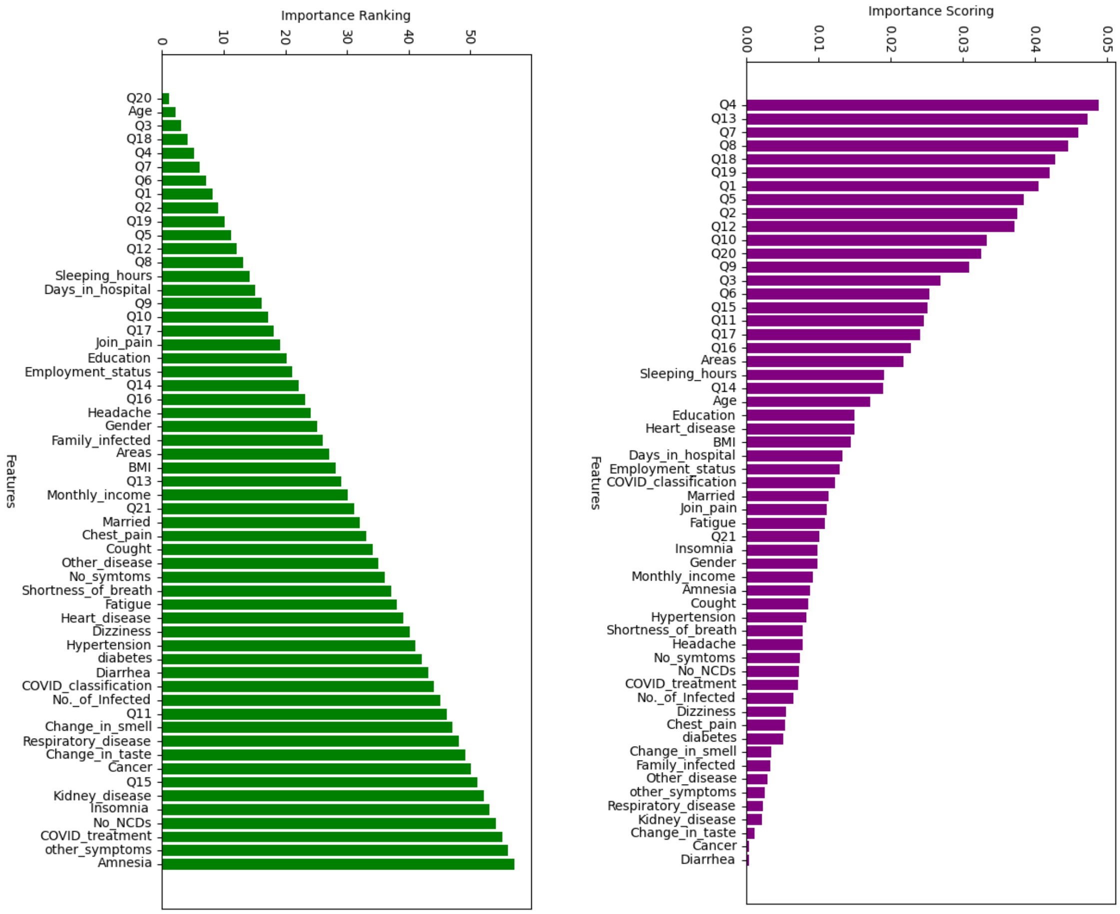
Long COVID symptoms		Depression (n=136)			Anxiety (n=228)			Stress (n=139)			Any mental health disorder (n=292)		
		Yes	No	ORs	Yes	No	ORs	Yes	No	ORs	Yes	No	ORs
Amnesia (n=270)	Yes	65	205	2.89***	108	162	1.35*	73	197	2.98***	141	129	0.85
	No	71	208		120	159		66	213		151	128	
Fatigue (n=191)	Yes	58	133	1.71***	101	90	0.71*	59	132	1.65***	122	69	0.41***
	No	78	280		127	231		80	278		170	188	
Cough (n=144)	Yes	36	108	1.08	66	78	0.48***	38	106	1.05	78	66	0.31***
	No	100	305		162	243		101	304		214	191	
Joint pain (n=110)	Yes	32	78	0.75	54	56	0.32***	30	80	0.73*	68	42	0.19***
	No	104	335		174	265		109	330		224	215	
Headache (n=103)	Yes	28	75	0.69*	48	55	0.31***	20	83	0.70*	59	44	0.19***
	No	108	338		180	266		119	327		233	213	
Insomnia (n=102)	Yes	23	79	0.70*	49	53	0.30***	26	76	0.67**	55	47	0.20***
	No	113	334		179	268		113	334		237	210	
Dizziness (n=76)	Yes	20	56	0.48***	35	41	0.21***	18	58	0.48***	45	31	0.13***
	No	116	357		193	280		121	352		247	226	
Shortness of breath (n=72)	Yes	21	51	0.44***	40	32	0.17***	23	49	0.42***	48	24	0.1***
	No	115	362		188	289		116	361		244	233	
Chest pain (n=46)	Yes	14	32	0.26***	29	17	0.09***	16	30	0.24***	33	13	0.05***
	No	122	381		199	304		123	380		259	244	
Other symptoms (n=18)	Yes	4	14	0.11***	6	12	0.05***	1	17	0.12***	8	10	0.04***
	No	132	399		222	309		138	393		284	247	
Diarrhea (n=11)	Yes	6	5	0.04***	11	0	0***	7	4	0.03***	11	0	0
	No	130	408		217	321		132	406		281	257	
Change in smell (n=11)	Yes	5	6	0.05***	8	3	0.01***	5	6	0.04***	10	1	0.004***
	No	131	407		220	318		134	404		282	256	
Change in taste (n=9)	Yes	4	5	0.04***	6	3	0.01***	3	6	0.04***	6	3	0.01***
	No	132	408		222	318		136	404		286	254	
Any symptom (n=546)	Yes	136	410	-	227	319	319***	139	407	-	291	255	255***
	No	0	3		1	2		0	3		1	2	

* $p<0.05$; ** $p < 0.01$; *** $p < 0.001$

4.3. Prediction of mental health and long COVID status among recovered COVID-19 patients

4.3.1. Feature selection for dataset of mental health status and long COVID status

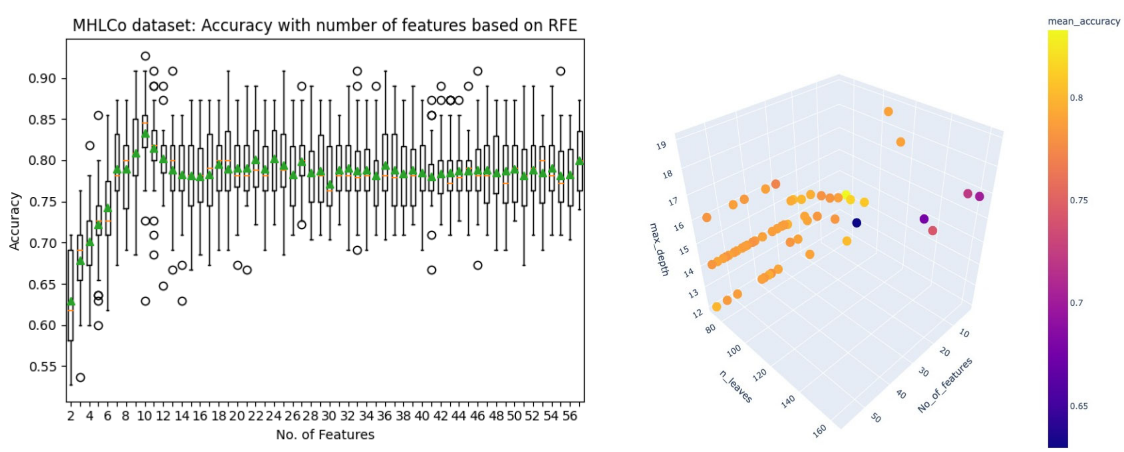
Features of the MHLCo dataset were arranged in order of importance by the RFE and ET methods, as shown in Fig.3. In the case of the RFE method, the top ten features, in order of importance, were Q20, age, Q3, Q18, Q4, Q7, Q6, Q1, Q2, and Q19. The group comprising these top ten features exhibited the highest mean accuracy of 0.83, as illustrated in Fig.4(a). The accuracy was determined utilizing the decision tree technique with specified hyperparameters, where the number of leaves (n_leaves) was set to 90, and the maximum depth of the tree (max_depth) was 15, as denoted by the yellow point in Fig.4(b). Similarly, for the ET method, the top ten features in terms of importance were Q4, Q13, Q7, Q8, Q18, Q19, Q1, Q5, Q2, and Q12. In this case, the group encompassing these top ten features, along with eight additional features (Q10, Q20, Q9, Q3, Q6, Q15, Q11, Q17), exhibited the highest mean accuracy of 0.82, as depicted in Fig.5(a). The accuracy was also determined using the decision tree technique, with the number of leaves (n_leaves) set to 82 and the maximum depth of the tree (max_depth) set to 15, as indicated by the yellow point in Fig.5(b).



(a) Importance Ranking by RFE method

(b) Importance Scoring by ET method

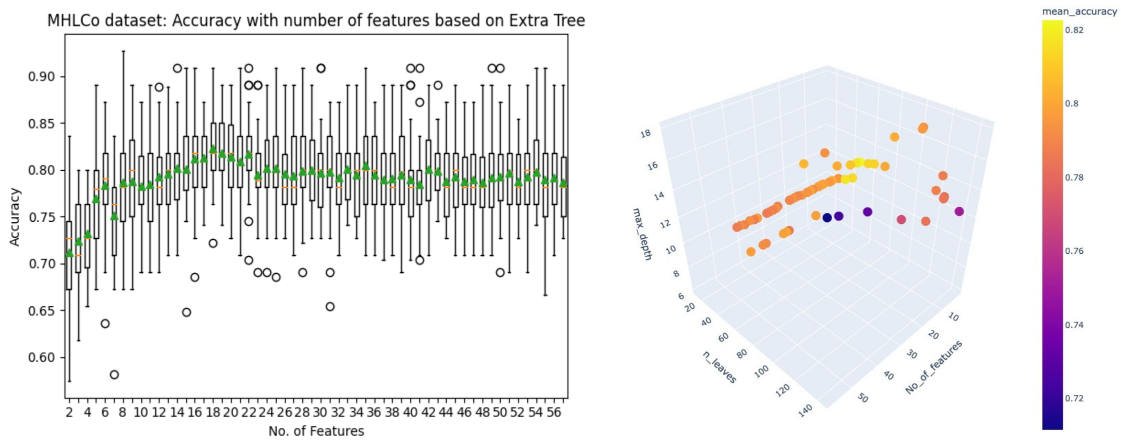
Fig. 3. Importance features based on RFE and ET methods for the MHLCo dataset



(a) Accuracy with a number of features based on RFE method

(b) Accuracy with hyper-parameters of Decision Tree

Fig. 4. Accuracy with hyper-parameters of decision tree method based on RFE feature selection for the MHLCo dataset



(a) Accuracy with a number of features based on ET method

(b) Accuracy with hyper-parameters of Decision Tree

Fig. 5. Accuracy with hyper-parameters of decision tree method based on ET feature selection for the MHLCo dataset

4.3.2. Performance of machine learning algorithms

Six ML models were used in this study: KNN, MLP, SVM, Gradient Boosting, Voting Classifier, and XGBoost. The results of experiments obtained from these methods with the optimized hyper-parameters and 10-fold cross-validation are depicted in Fig.6 and Table 4.

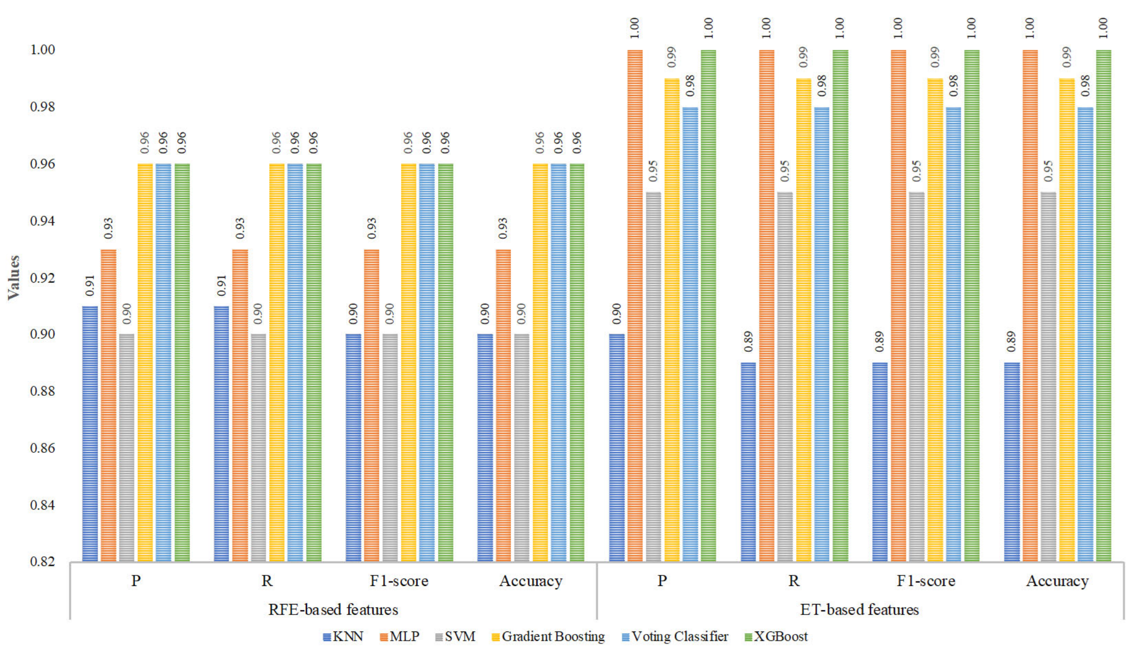
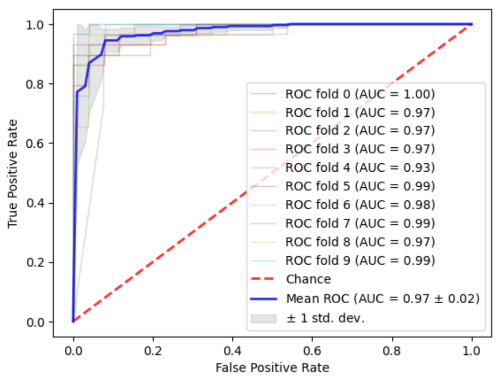


Fig. 6. Results of the ML models with feature selection methods (RFE and ET)

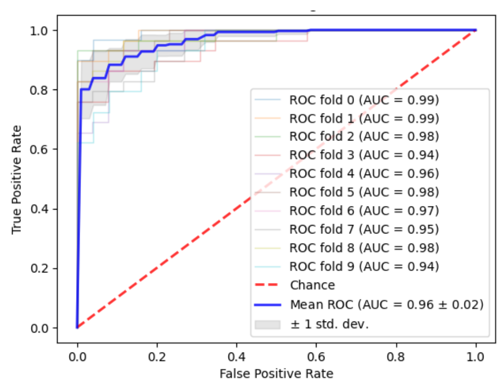
In terms of features selected using the RFE method, the MLP model achieved an accuracy, precision, recall, and F1-score of 0.93, outperforming KNN and SVM among single classifiers. However, when considering ensemble methods (Gradient Boosting, Voting Classifier, and XGBoost), these models displayed an even higher accuracy of 0.96, along with consistently high accuracy, precision, recall, F1 scores, all at 0.96. Additionally, the mean AUC values provide deeper insights into the models' ability to distinguish between classes. Among the models with the highest mean accuracy, Voting Classifier emerged as the standout performer, boasting a mean AUC of 0.95 ± 0.01, surpassing both Gradient Boosting (0.94 ± 0.01) and XGBoost (0.94 ± 0.02) (See Fig. 6 and Fig. 8 for more details). Switching to ET method-based features, the MLP model showed remarkable performance with accuracy, precision, recall, and F1 Score all equal to 1.0, surpassing KNN and SVM among single classifiers. However, among ensemble methods such as Gradient Boosting, Voting Classifier, and XGBoost, the XGBoost method demonstrated the highest accuracy of 1.0 and consistently high precision, recall, and F1 scores, all at 1.0. Further analysis using mean AUC values sheds light on the models' discriminatory capabilities. Among models with the highest mean accuracy, MLP stood out with a mean AUC of 0.97 ± 0.02, surpassing XGBoost (0.96 ± 0.02) (See Fig. 6 and Fig. 7 for more details).

Table 4
Results of the ML models with feature selection methods (RFE and ET)

	RFE-based features				ET-based features			
	P	**R**	**F1-score**	**Accuracy**	**P**	**R**	**F1-score**	**Accuracy**
KNN	0.91	0.91	0.90	0.90	0.90	0.89	0.89	0.89
MLP	0.93	0.93	0.93	0.93	1.00	1.00	1.00	1.00
SVM	0.90	0.90	0.90	0.90	0.95	0.95	0.95	0.95
Gradient Boosting	0.96	0.96	0.96	0.96	0.99	0.99	0.99	0.99
Voting Classifier	0.96	0.96	0.96	0.96	0.98	0.98	0.98	0.98
XGBoost	0.96	0.96	0.96	0.96	1.00	1.00	1.00	1.00

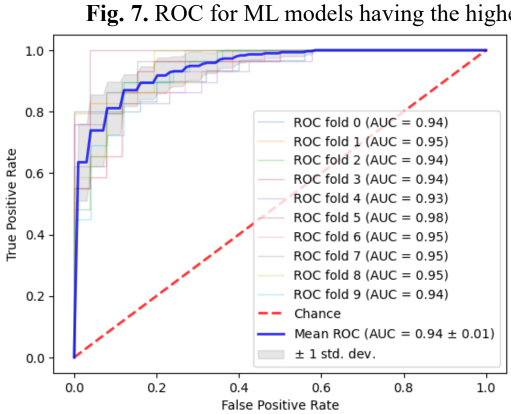


(a) ROC of the MLP

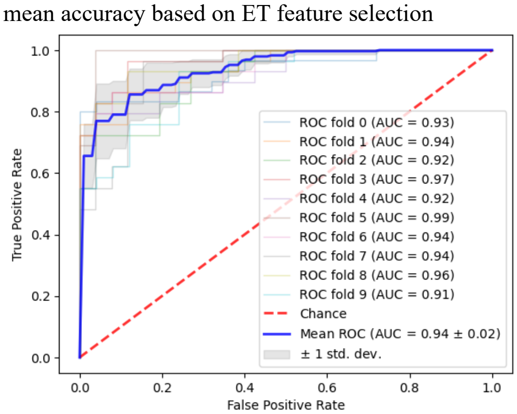


(b) ROC of the XGBoost

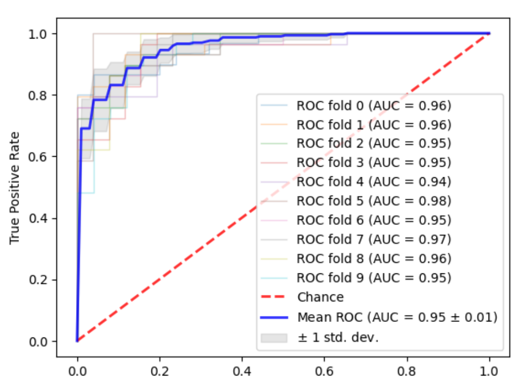
Fig. 7. ROC for ML models having the highest mean accuracy based on ET feature selection



(a) ROC of the Gradient Boosting



(b) ROC of the XGBoost



(c) ROC of the Voting Classifier

Fig. 8. ROC for ML models having the highest mean accuracy based on RFE feature selection

5. Discussion

This research delves deeply into the mental health and long COVID status experienced by individuals who have recovered from COVID-19. Our findings highlighted a substantial prevalence of any mental health issue (at least one symptom among depression, anxiety, or stress) in recovered COVID-19 patients, accounting for 53.19%. Additionally, we observed a significant occurrence of long COVID symptoms, with amnesia affecting 270 patients (49.18%), fatigue impacting 191 patients (34.79%), and cough affecting 144 patients (26.23%) being the most prevalent. Furthermore, our investigation pinpointed amnesia, fatigue, and cough as the primary risk factors influencing the mental health disorders of recovered COVID-19 patients experiencing these symptoms. This aligns with other research emphasizing that mental health and long COVID symptoms were high prevalence (Do Duy et al., 2020; Duong et al., 2020; Phu et al., 2023; Suwanbamrung et al., 2023).

Out of 57 features, the RFE method selected 10 features, while the ET method selected 18 features, both yielding the highest mean accuracy exceeding 0.80. Additionally, the study showcased the predictive prowess of various ML models (KNN, MLP, SVM, Gradient Boosting, Voting Classifier, and XGBoost), all achieving accuracy scores exceeding 0.90 with 10-fold cross-validation in predicting mental health conditions and long COVID status among recovered COVID-19 patients. Specifically, Voting Classifier exhibited the best performance in both accuracy and AUC among the models selected through RFE feature selection, 0.96 and 0.95± 0.01, respectively. On the other hand, among the ensemble methods chosen through ET feature selection, MLP excelled in accuracy and AUC among the models, 1.00 and 0.97 ± 0.02, respectively. Most ML models performed relatively well.

The optimal ML models for predicting mental health and long COVID from recovered COVID-19 patients did not require features related to sociodemographic factors, underlying diseases, or long COVID specifically. However, including these features resulted in consistently high mean accuracies, all exceeding 0.75. This observation, evident in the MHLCo datasets, indicates a potential influence of sociodemographic factors, underlying diseases, and long COVID-related details on the mental health and long COVID status of recovered COVID-19 patients, which aligns with findings from previous research (Linh et al., 2024; Phu et al., 2023).

However, our research has some limitations. Our data primarily relied on structured questionnaires to assess psychiatric symptoms, which were not conducting clinical diagnoses. Moreover, the exclusive emphasis on mental health disorders and long COVID symptoms overlooks other essential aspects of mental well-being, potentially constraining the model's broader applicability. These limitations should be taken into account when interpreting findings from this field in the future.

6. Conclusion

This research proposed a framework based on ML designed to predict mental health disorders and long COVID from a dataset of recovered COVID-19 patients. By leveraging a dataset comprising socio-demographic information, underlying diseases, mental health (DASS-21), and long COVID symptoms, various ML models (KNN, MLP, SVM, Gradient Boosting, Voting Classifier, and XGBoost) with hyper-parameters tuning and feature selection methods (RFE and ET) were employed to achieve high accuracy in prediction. The MLP with ET-based features demonstrated and performed well in our experimentation with the highest accuracy and AUC scores of 1.00 and 0.97 ± 0.02, respectively. Furthermore, we recognized the potential of incorporating markers, e.g., biochemical and physiological indicators, to better understand mental health conditions and long COVID. Future research endeavors will focus on integrating the markers into survey data to enhance the support provided for mental health and long COVID issues. This integration could be potential strategies for personalized intervention tailored to individuals based on insights derived from ML predictions, thereby improving overall healthcare outcomes in the post-COVID-19 recovery phase.

Acknowledgments

The authors would like to express their gratitude to the School of Informatics and the Informatics Innovation Center of Excellence (IICE) at Walailak University, Thailand, and Dong Thap Medical College, Vietnam, for their support during this research.

Institutional Review Board Statement

The Dong Thap Medical College Research Ethics Committee approved this research using the reused data without individual information of participants (secondary data) (Approval No.: 02/180/QD-CDYT).

References

Afrash, M. R., Shanbehzadeh, M., & Kazemi-Arpanahi, H. (2022). Predicting risk of mortality in COVID-19 hospitalized patients using hybrid machine learning algorithms. *Journal of Biomedical Physics & Engineering*, *12*(6), 611.

Aiyegbusi, O. L., Hughes, S. E., Turner, G., Rivera, S. C., McMullan, C., Chandan, J. S., Haroon, S., Price, G., Davies, E. H., Nirantharakumar, K., & others. (2021). Symptoms, complications and management of long COVID: a review. *Journal of the Royal Society of Medicine*, *114*(9), 428–442.

Aljrees, T. (2024). Improving prediction of cervical cancer using KNN imputer and multi-model ensemble learning. *Plos One*, *19*(1), e0295632.

Almajali, D., & Masadeh, R. (2021). Antecedents of students' perceptions of online learning through covid-19 pandemic in Jordan. *International Journal of Data and Network Science*, *5*(4), 587–592.

Arshed, M. A., Qureshi, W., Khan, M. U. G., & Jabbar, M. A. (2021). Symptoms based Covid-19 disease diagnosis using machine learning approach. *2021 International Conference on Innovative Computing (ICIC)*, 1–7.

Batool, A., & Byun, Y.-C. (2024). Towards Improving Breast Cancer Classification using an Adaptive Voting Ensemble Learning Algorithm. *IEEE Access*.

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems*, *24*.

Chadaga, K., Prabhu, S., Sampathila, N., Chadaga, R., Umakanth, S., Bhat, D., & GS, S. K. (2024). Explainable artificial intelligence approaches for COVID-19 prognosis prediction using clinical markers. *Scientific Reports*, *14*(1), 1783.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794.

Cho, S.-E., Geem, Z. W., & Na, K.-S. (2021). Predicting depression in community dwellers using a machine learning algorithm. *Diagnostics*, *11*(8), 1429.

Chung, J., & Teo, J. (2023). Single classifier vs. ensemble machine learning approaches for mental health prediction. *Brain Informatics*, *10*(1), 1–10.

Davis, H. E., McCorkell, L., Vogel, J. M., & Topol, E. J. (2023). Long COVID: major findings, mechanisms and recommendations. *Nature Reviews Microbiology*, *21*(3), 133–146.

De Oliveira Almeida, K., Nogueira Alves, I. G., de Queiroz, R. S., de Castro, M. R., Gomes, V. A., Santos Fontoura, F. C., Brites, C., & Neto, M. G. (2023). A systematic review on physical function, activities of daily living and health-related quality of life in COVID-19 survivors. *Chronic Illness*, *19*(2), 279–303.

Dhariwal, N., Sengupta, N., Madiajagan, M., Patro, K. K., Kumari, P. L., Abdel Samee, N., Tadeusiewicz, R., Pławiak, P., & Prakash, A. J. (2024). A pilot study on AI-driven approaches for classification of mental health disorders. *Frontiers in Human Neuroscience*, *18*, 1376338.

Do Duy, C., Nong, V. M., Van, A. N., Thu, T. D., Do Thu, N., & Quang, T. N. (2020). COVID-19-related stigma and its association with mental health of health-care workers after quarantine in Vietnam. *Psychiatry and Clinical Neurosciences*, *74*(10), 566.

Duong, K. N. C., Le Bao, T. N., Nguyen, P. T. L., Van, T. V., Lam, T. P., Gia, A. P., Anuratpanich, L., Van, B. V., & others. (2020). Psychological impacts of COVID-19 during the first nationwide lockdown in Vietnam: web-based, cross-sectional survey study. *JMIR Formative Research*, *4*(12), e24776.

Elnagar, A., Alnazzawi, N., Afyouni, I., Shahin, I., Nassif, A., & Salloum, S. (2022). An empirical study of e-learning post-acceptance after the spread of COVID-19. *International Journal of Data and Network Science*, *6*(3), 669–682.

Engel, F. D., da Fonseca, G. G. P., Cechinel-Peiter, C., Backman, C., da Costa, D. G., & de Mello, A. L. S. F. (2023). Impact of the COVID-19 Pandemic on the Experiences of Hospitalized Patients: A Scoping Review. *Journal of Patient Safety*, *19*(2), e46.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*, 3–42.

Gupta, A., Jain, V., & Singh, A. (2022). Stacking ensemble-based intelligent machine learning model for predicting post-COVID-19 complications. *New Generation Computing*, *40*(4), 987–1007.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*(Mar), 1157–1182.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*, 389–422.

Hossen, M. J., Ramanathan, T. T., Al Mamun, A., & others. (2024a). An Ensemble Feature Selection Approach-Based Machine Learning Classifiers for Prediction of COVID-19 Disease. *International Journal of Telemedicine and Applications*, *2024*.

Hossen, M. J., Ramanathan, T. T., Al Mamun, A., & others. (2024b). An Ensemble Feature Selection Approach-Based Machine Learning Classifiers for Prediction of COVID-19 Disease. *International Journal of Telemedicine and Applications*, *2024*.

Hussein, A. N., Makki, S. V. A.-D., & Al-Sabbagh, A. (2023). Comprehensive study: machine learning approaches for COVID-19 diagnosis. *International Journal of Electrical and Computer Engineering (IJECE)*, *13*(5), 5681–5695.

Islam, M. N., Islam, M. S., Shourav, N. H., Rahman, I., Faisal, F. Al, Islam, M. M., & Sarker, I. H. (2024). Exploring post-COVID-19 health effects and features with advanced machine learning techniques. *Scientific Reports*, *14*(1), 9884.

Jha, A., Abirami, M. S., & Kumar, V. (2022). Predictive Model for Depression and Anxiety Using Machine Learning Algorithms. *International Conference on Deep Sciences for Computing and Communications*, 133–147.

Jiang, S., Loomba, J., Sharma, S., & Brown, D. (2022). Vital measurements of hospitalized COVID-19 patients as a predictor of long COVID: An EHR-based cohort study from the RECOVER program in N3C. *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 3023–3030.

Juliet, S., & others. (2023). Investigations on Machine Learning Models for Mental Health Analysis and Prediction. *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, 1–7.

Kanrak, M., & Nonthapot, S. (2024). Analysis of the tourism network post the COVID-19 pandemic: Implications for revitalization. *International Journal of Data and Network Science*, *8*(3), 1781–1792.

Katiyar, K., Fatma, H., & Singh, S. (2024). Predicting Anxiety, Depression and Stress in Women Using Machine Learning Algorithms. In *Combating Women's Health Issues with Machine Learning* (pp. 22–40). CRC Press.

Kim, S. W., & Chang, M. C. (2023). The usefulness of machine learning analysis for predicting the presence of depression with the results of the Korea National Health and Nutrition Examination Survey. *Annals of Palliative Medicine*, *12*(4), 74756–74856.

Ku, W. L., & Min, H. (2024). Evaluating Machine Learning Stability in Predicting Depression and Anxiety Amidst Subjective Response Errors. *Healthcare*, *12*(6), 625.

Kumar, P., Chandra, S., & others. (2023). Prediction and comparison of psychological health during COVID-19 among Indian population and Rajyoga meditators using machine learning algorithms. *Procedia Computer Science*, *218*, 697–705.

Linh, H. N., Loan, N. T., Uyen, N. T. T., Nam, T. T., Phu, D. H., & others. (2024). Prevalence and risk factors associated with long COVID symptoms in children and adolescents in a southern province of Vietnam. *Asian Pacific Journal of Tropical Medicine*, *17*(3), 119–128.

Lopez-Leon, S., Wegman-Ostrosky, T., Perelman, C., Sepulveda, R., Rebolledo, P. A., Cuapio, A., & Villapol, S. (2021). More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. *Scientific Reports*, *11*(1), 16144.

Lovibond, S. H. (1995). Manual for the depression anxiety stress scales. *Sydney Psychology Foundation*.

Malik, S. S., & Khan, A. (2023). Anxiety, Depression and Stress prediction among College Students using Machine Learning Algorithms. *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, 1–5.

Nasir, A., Makki, S. V. A.-D., & Al-Sabbagh, A. (2024). Pandemia Prediction Using Machine Learning. *PRZEGLĄD ELEKTROTECHNICZNY*, *5*, 211–214.

Nguyen, H. V., & Byeon, H. (2022). Explainable Deep-Learning-Based Depression Modeling of Elderly Community after COVID-19 Pandemic. *Mathematics*, *10*(23), 4408.

Nison, P., Vuttipittayamongkol, P., Boonyapuk, P., & Kemavuthanon, K. (2023). A Machine Learning Approach for Depression Screening in College Students Based on Non-Clinical Information. *2023 International Conference On Cyber Management And Engineering (CyMaEn)*, 413–417.

Patel, M. A., Knauer, M. J., Nicholson, M., Daley, M., Van Nynatten, L. R., Cepinskas, G., & Fraser, D. D. (2023). Organ and cell-specific biomarkers of Long-COVID identified with targeted proteomics and machine learning. *Molecular Medicine*, *29*(1), 26.

Patterson, B. K., Guevara-Coto, J., Yogendra, R., Francisco, E. B., Long, E., Pise, A., Mora, J., & Mora-Rodr\'\iguez, R. A. (2021). Immune-based prediction of COVID-19 severity and chronicity decoded using machine learning. *Frontiers in Immunology*, *12*, 700782.

Pei, H., Wu, Q., Xie, Y., Deng, J., Jiang, L., & Gan, X. (2021). A qualitative investigation of the psychological experiences of COVID-19 patients receiving inpatient care in isolation. *Clinical Nursing Research*, *30*(7), 1113–1120.

Pfaff, E. R., Girvin, A. T., Bennett, T. D., Bhatia, A., Brooks, I. M., Deer, R. R., Dekermanjian, J. P., Jolley, S. E., Kahn, M. G., Kostka, K., & others. (2022). Identifying who has long COVID in the USA: a machine learning approach using N3C data. *The Lancet Digital Health*, *4*(7), e532–e541.

Phu, D. H., Maneerattanasak, S., Shohaimi, S., Trang, L. T. T., Nam, T. T., Kuning, M., Like, A., Torpor, H., & Suwanbamrung, C. (2023). Prevalence and factors associated with long COVID and mental health status among recovered COVID-19 patients in southern Thailand. *PloS One*, *18*(7), e0289382.

Pramodhani, R. J., Vineela, P. S. S., Aseesh, V. S., Kumar, K., & Devi, B. S. K. (2022). Stress Prediction and Detection in Internet of Things using Learning Methods. *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)*, 303–309.

Priya, A., Garg, S., & Tigga, N. P. (2020). Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Computer Science*, *167*, 1258–1267.

Prout, T. A., Zilcha-Mano, S., Aafjes-van Doorn, K., Békés, V., Christman-Cohen, I., Whistler, K., Kui, T., & Di Giuseppe, M. (2020). Identifying predictors of psychological distress during COVID-19: a machine learning approach. *Frontiers in Psychology*, *11*, 586202.

Qasrawi, R., Amro, M., VicunaPolo, S., Al-Halawa, D. A., Agha, H., Seir, R. A., Hoteit, M., Hoteit, R., Allehdan, S., Behzad, N., & others. (2022). Machine learning techniques for predicting depression and anxiety in pregnant and postpartum women during the COVID-19 pandemic: A cross-sectional regional study. *F1000Research*, *11*.

Rass, V., Beer, R., Schiefecker, A. J., Kofler, M., Lindner, A., Mahlknecht, P., Heim, B., Limmert, V., Sahanic, S., Pizzini, A., & others. (2021). Neurological outcome and quality of life 3 months after COVID-19: A prospective observational cohort study. *European Journal of Neurology*, *28*(10), 3348–3359.

Rezapour, M., & Hansen, L. (2022). A machine learning analysis of COVID-19 mental health data. *Scientific Reports*, *12*(1), 14965.

Saha, R., Malviya, L., Jadhav, A., & Dangi, R. (2024). Early stage HIV diagnosis using optimized ensemble learning technique. *Biomedical Signal Processing and Control*, *89*, 105787.

Saltzman, L. Y., Longo, M., & Hansel, T. C. (2023). Long-COVID stress symptoms: Mental health, anxiety, depression, or posttraumatic stress. *Psychological Trauma: Theory, Research, Practice, and Policy*.

Sarmiento Varón, L., González-Puelma, J., Medina-Ortiz, D., Aldridge, J., Alvarez-Saravia, D., Uribe-Paredes, R., & Navarrete, M. A. (2023). The role of machine learning in health policies during the COVID-19 pandemic and in long COVID management. *Frontiers in Public Health*, *11*, 1140353.

Scikit-learn. (2024). *Cross-validation: evaluating estimator performance*. https://scikit-learn.org/stable/modules/cross_validation.html

Singh, S., Gupta, H., Singh, P., & Agrawal, A. P. (2022). Comparative Analysis of Machine Learning Models to Predict Depression, Anxiety and Stress. *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*, 1199–1203.

Sudre, C. H., Murray, B., Varsavsky, T., Graham, M. S., Penfold, R. S., Bowyer, R. C., Pujol, J. C., Klaser, K., Antonelli, M., Canas, L. S., & others. (2021). Attributes and predictors of long COVID. *Nature Medicine*, *27*(4), 626–631.

Suwanbamrung, C., Pongtalung, P., Trang, L. T. T., Phu, D. H., & Nam, T. T. (2023). Levels and risk factors associated with depression, anxiety, and stress among COVID-19 infected adults after hospital discharge in a Southern Province of Thailand. *Journal of Public Health and Development*, 72–89.

Trang, L. T. T., Le, C. N., Chutipatana, N., Shohaimi, S., & Suwanbamrung, C. (2023). Prevalence and predictors of depression, anxiety, and stress among recovered COVID-19 patients in Vietnam. *Roczniki Państwowego Zakładu Higieny*, *74*(2).

Trivedi, N. K., Tiwari, R. G., Witarsyah, D., Gautam, V., Misra, A., & Nugraha, R. A. (2022). Machine Learning Based Evaluations of Stress, Depression, and Anxiety. *2022 International Conference Advancement in Data Science, E-Learning and Information Systems (ICADEIS)*, 1–5.

Tuan, T. A., Trang, L. T. T., An, T. D., Nghia, N. H., & Loan, D. T. T. (2024). An Analysis of Machine Learning for Detecting Depression, Anxiety, and Stress of Recovered COVID-19 Patients. *Journal of Human, Earth, and Future*, *5*(1), 1–18.

Tyshchenko, Y. (2018). Depression and anxiety detection from blog posts data. *Nature Precis. Sci., Inst. Comput. Sci., Univ. Tartu, Tartu, Estonia*.

Vaishnavi, K., Kamath, U. N., Rao, B. A., & Reddy, N. V. S. (2022). Predicting mental health illness using machine learning algorithms. *Journal of Physics: Conference Series*, *2161*(1), 12021.

WHO. (2000). *The Asia-Pacific perspective: redefining obesity and its treatment*.

WHO. (2022). *Post COVID-19 condition (Long COVID)*. https://www.who.int/europe/news-room/fact-sheets/item/post-covid-19-condition

WHO. (2024). *COVID-19 epidemiological update – 12 April 2024*. https://www.who.int/publications/m/item/covid-19-epidemiological-update-edition-166

Yazdani, A., Zahmatkeshan, M., Ravangard, R., Sharifian, R., & Shirdeli, M. (2022). Supervised machine learning approach to COVID-19 detection based on clinical data. *Medical Journal of the Islamic Republic of Iran*, *36*.

Zhang, Y., Chinchilli, V. M., Ssentongo, P., & Ba, D. M. (2024). Association of Long COVID with mental health disorders: a retrospective cohort study using real-world data from the USA. *BMJ Open*, *14*(2), e079267.

Zoabi, Y., Deri-Rozov, S., & Shomron, N. (2021). Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *Npj Digital Medicine*, *4*(1), 3.