Predicting Post-COVID Complications of Bangladeshi People Using Machine Learning: Insights, Models, and Future Directions

MD Nadim Hasan
Student of American International University-Bangladesh (AIUB)
Dhaka, Bangladesh

Syed Mominin Islam Tamim
Student of American International University-Bangladesh (AIUB) and
Jr. Software Engineer at Venus IT Ltd.,
Dhaka, Bangladesh

Md. Tafsimul Islam Tanzid
Student of American International University-Bangladesh (AIUB) and
Jr. Software Engineer at Divergent Technologies Ltd.,
Dhaka, Bangladesh

Tanvir Ahmed
Lecturer, American International University-Bangladesh (AIUB)
Dhaka, Bangladesh

Abstract

The SARS-CoV-2 virus, which causes COVID-19 pandemic has had a profound influence on world health. While most infected individuals experience mild to moderate respiratory infections, some develop post-COVID complications with long-lasting effects. Detecting these complications early is crucial for better patient outcomes. This research explores the possibility of predicting post-COVID complications using patient data and machine learning algorithms. The study reviews existing literature on post-COVID complications and previous research efforts that have utilized machine learning in healthcare. It proposes model selection and training to improve prediction accuracy. The dataset is collected from COVID patients through a survey conducted with random Bangladeshi participants, and data preprocessing techniques are applied. Linear Regression, Decision Trees, Random Forest, and K-Nearest Neighbors are the four machine learning models selected and trained on the dataset. The models' performance is evaluated, and their accuracy and effectiveness in predicting post-COVID complications are compared. The findings show that Decision Trees and Linear Regression have the best accuracy. Overall, this research highlights the potential of machine learning in post-COVID complexity detection and contributes to advancing strategies for managing the pandemic and its long-term effects. The study also proposes some future directions for related work, which could be helpful in long-term research.

Keywords

Machine Learning, Learning Outcomes, Numeric Data, Post-COVID Complexity, Accuracy

1. Introduction

The SARS-CoV-2 virus is the cause of the coronavirus illness (Covid-19). Except for a few people who are extremely ill and require medical attention, the majority of infected persons will only experience mild to moderate respiratory infections and recover on their own (World Health Organization 2021). Many positive cases have been reported since 2019, and people are suffering from this virus. However, the large number of vaccinations has led to a decline in several COVID-19 cases. Post-COVID issues, occasionally encountered by individuals who have contracted COVID, refer to the long-lasting effects associated with the virus. The development of post-COVID problems can occur at least four weeks after the original infection, even though the majority of COVID-19 patients show symptoms within a span of a few days to a few weeks after getting the virus (CDC 2021).

A discipline of computing techniques called machine learning, which is constantly developing, aims to imitate human intelligence by learning from its environment. It is regarded as the workhorse of the new big data era. It is possible to determine any kind of long-term cases of diseases like COVID-19 and make predictions from the dataset using machine learning algorithms. The main intention of this research is to ensure an applicable machine-learning model for detecting post-COVID physical complexity.

This research aims to address a fundamental question: Can post-COVID complications be predicted using patient data, and how can machine learning algorithms assist in this endeavor? Working collaboratively, data scientists, computer scientists, physicians, and other professionals use machine learning models and existing data to anticipate illnesses and long-term health concerns. However, the specific data processing criteria and challenges for predicting post-COVID complications remain unexplored. By examining the viability of leveraging patient data and machine learning approaches for identifying post-COVID physical difficulties, this study aims to close this gap. We will gain important insights regarding the potential of machine learning in anticipating and preventing post-COVID problems through thorough study.

2. Related Work

Few works have been done on the topic of post-complexity detection of a COVID patient. However, a paper titled 'Studying the post-COVID-19 condition: research challenges, strategies, and priorities' was published in BMC Medicine (Munblit et al. 2022). The paper discusses the challenges and strategies for studying the post-COVID-19 condition, also known as 'long COVID' or 'post-acute sequelae of SARS-CoV-2 infection (PASC).' The study also emphasizes the application of machine learning to identify individuals at risk for long-term COVID-19 and forecast the severity of symptoms. Another paper titled 'Identification of high-risk COVID-19 patients using machine learning' in PLOS ONE (Quiroz-Juárez et al. 2021) introduces a machine-learning system that can determine if a patient—whether they are truly sick or only believe they are—has a higher chance of surviving than dying, or vice versa. Historical information is used to train the algorithm, including COVID-19-related data, demographic data, and medical history. The study shows that the suggested strategy may accurately identify high-risk patients at each of the four recognized clinical phases, facilitating better hospital capacity planning and prompt treatment. In order to predict cardiac illness following COVID-19 infection, A. Gupta et al. present a deep neural network-modeled stacking ensemble-based binary classifier. You can find the paper on Europe PMC (Gupta, Jain, & Singh 2021).

Another paper titled 'COVID-19 detection using federated machine learning' reported its findings in PLOS ONE. The research analyzes and contrasts machine learning and soft computing methods in forecasting the COVID-19 epidemic (Al-Maqtari et al. 2022). A paper titled 'COVID-19 Infection Detection Using Machine Learning' was published in IEEE. The study describes a machine learning-based predictor for COVID-19 infections and evaluates the predictive power of five machine learning models. The research on post-COVID complexity detection and risk prediction for COVID-19 patients emphasizes the valuable role of machine learning in addressing pandemic challenges. Studies have focused on understanding long COVID and demonstrated machine learning's ability to identify at-risk patients and predict symptom severity. Additionally, algorithms have been developed to detect high-risk COVID-19 patients, aiding in timely treatment and hospital planning (Wang et al. 2021). These efforts contribute to advancing AI-driven approaches for improved COVID-19 management. Further research is needed to fully harness machine learning's potential in addressing post-COVID complexities and enhancing pandemic response strategies.

3. Methodology

The basic goal of this study is to apply certain machine learning models to primary data collected from more than 500 people living in Bangladesh. First of all, this research needs to identify the real outcomes and what kind of complexity may happen after COVID-19. Then make the survey question for random COVID patients. It might be easy to collect data from hospitals. But the main obstacle to this is that general people in not so friendly to share their problems with a disease like COVID. A paper was discussed. It talks about the necessity of creating a Core Outcome Set (COS) for post-COVID-19 situations as quickly as is practical to increase data quality, harmonization, and comparability between various geographical regions. The article urges the creation of a worldwide project that includes all pertinent parties, such as healthcare workers, researchers, methodologists, patients, and carers (Munblit et al. 2022). Another paper named "Burden of post-COVID-19 Syndrome and Implications for Healthcare." It talks about the effects of post-COVID-19 syndrome and how they affect healthcare. The study assesses the use of healthcare services particularly because of COVID-19 and limits its analysis to healthcare interactions that were reportedly connected to symptoms that persisted or worsened, complications, or new medical diagnoses associated with COVID-19, as well as normal follow-up following COVID-19 (Menges et al. 2021). An article's writers discovered that post-COVID-19 syndrome is a widespread and diverse condition that has an impact on a variety of elements of health and wellbeing. Additionally, they discovered that a greater probability of developing a post-COVID-19 syndrome was linked to older age, female sex, pre-existing comorbidities, and the severity of acute sickness. They recommended more studies to comprehend the processes, risk factors, and most effective treatment of post-COVID-19 disorders (Greenhalgh et al. 2020).

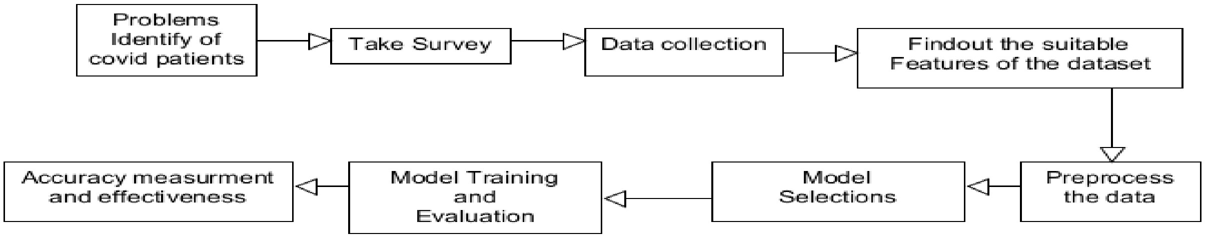

Problems Identify of covid patients
Take Survey
Data collection
Findout the suitable Features of the dataset
Preprocess the data
Model Selections
Model Training and Evaluation
Accuracy measurment and effectiveness

Figure 1. Working methodology of post-covid complexity detection

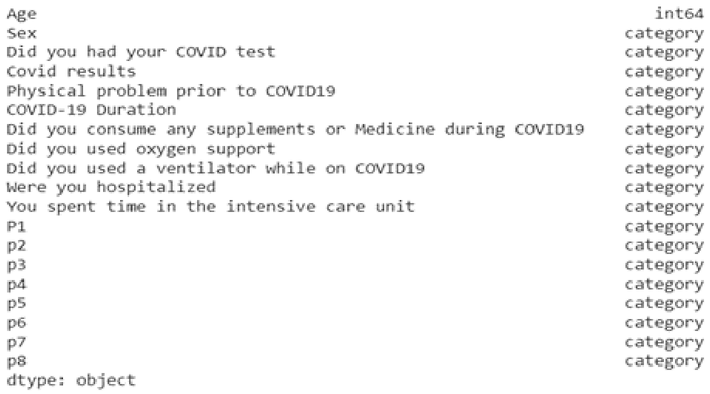
4. Data Collection and Preprocessing

For making a proper prediction, there needs to be a few things about the patients. Some of them had some health issues like Diabetes, Asthma, high blood pressure, Heart problems, Hyperlipidemia, etc. (Weerahandi et al. 2021). So, it is very complex to make focus on the valuable feature's predictions. The information utilized in this study came from a study that looked at individuals with severe COVID-19's health and symptoms after being released from the hospital. The survey involved the participation of five hundred individuals, who were selected for the purpose of conducting a comprehensive analysis of post-COVID outcomes.

Table 1. Covid Patients Conditions

Prior covid	Diabetes, Hypertension (high blood pressure), Hypertension (low blood pressure), Thyroid, Irregular menstrual cycle, lung issues, kidney issues, heart issues, asthma, None
During covid	lungs infection (highly infected), lungs infection (hardly infected), high fever, shortage of oxygen, Mental breakdown
Post-Covid	Tiredness, Mental trauma, Difficulty breathing, Chest pain, Heart palpitation, Hair fall, Headache, sleeping problems, Memory loss issues, Rash, Stomach pain, lung issues, changing menstrual cycle, others, None

After identifying the conditions of the COVID patients, there were taken a survey and put the result in the dataset. There were used Python language and libraries for the preprocessing of the data from the survey we found that a maximum of eight problems can occur in a patient. Then replace that with P1, P2, etc. means problem one, two, and so on. As well as we delete some unnecessary data like patients name which is really not needed things and very private things for a patient. There were taken actual data which was really needed for this research.



```
Age                                                      int64
Sex                                                   category
Did you had your COVID test                           category
Covid results                                         category
Physical problem prior to COVID19                     category
COVID-19 Duration                                     category
Did you consume any supplements or Medicine during COVID19   category
Did you used oxygen support                           category
Did you used a ventilator while on COVID19            category
Were you hospitalized                                 category
You spent time in the intensive care unit             category
P1                                                    category
p2                                                    category
p3                                                    category
p4                                                    category
p5                                                    category
p6                                                    category
p7                                                    category
p8                                                    category
dtype: object
```

Figure 2. Dataset of the survey

After finishing the data collection, we go for normalization. When referring to data transformation onto a standardized scale in the context of machine learning, normalization refers to the process of retaining the underlying differences in the value ranges while converting the data. The primary objective of normalization is to bring features onto a comparable scale, leading to enhanced model performance and increased training stability. Scaling within a predetermined range, applying clipping, log scaling, and using the z-score approach are all methods used for normalization. These normalization techniques are essential for maximizing the performance of machine learning models (Deepchecks, n.d.). Another part of the data was to find out the correlation between the prior COVID complexity and the post-COVID complexity of the COVID-19 patients.

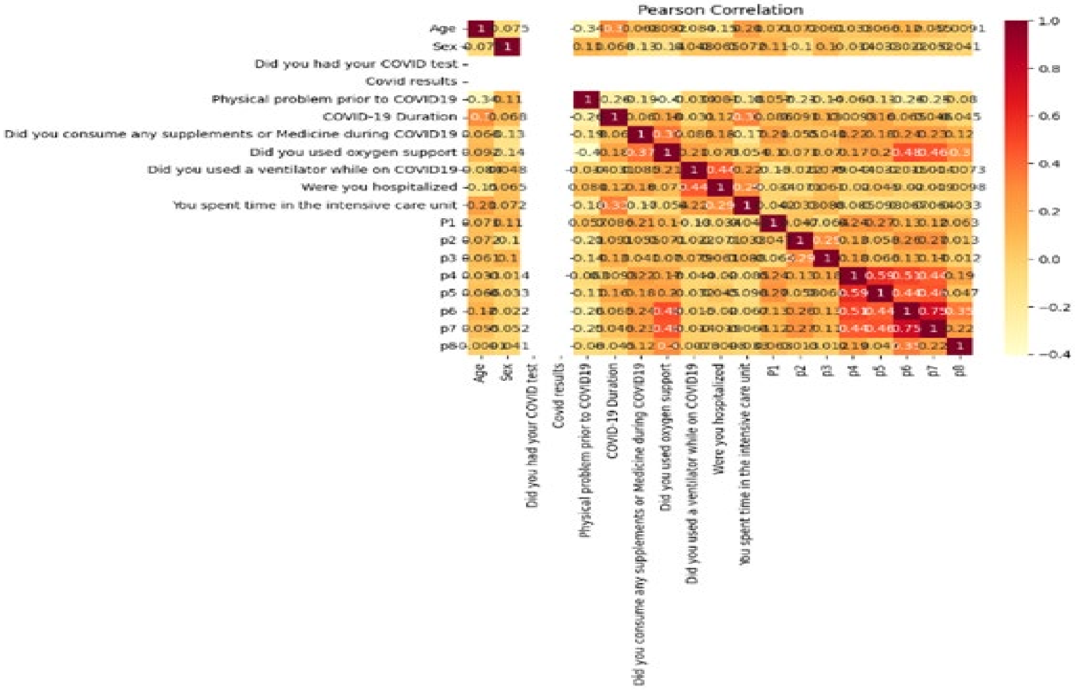


Figure 3. Co-relation of the columns

5. Model Selection and Training

Many different businesses, including marketing, finance, and healthcare, use machine learning algorithms to carry out operations like decision-making, natural language processing, and pattern detection. Additionally, they play a role in enhancing products and services, like voice assistants, self-driving cars, and recommendation systems (Sarker 2021). Supervised learning uses labeled training data to learn patterns and make predictions. Unsupervised learning works with raw data to discover patterns independently. Semi-supervised learning involves a mix of labeled and unlabeled input (IBM, n.d.).

When using machine learning in any situation, including the forecasting of post-COVID problems, model selection is crucial. It entails the identification of the most fitting machine learning algorithm or a combination of algorithms that can achieve precise predictions, considering the provided data and specific problem specifications. In this research all data are numeric. The complexities are replaced by random numeric numbers. So, it's easy to apply some machine learning models which is appropriate for training the numerical dataset.

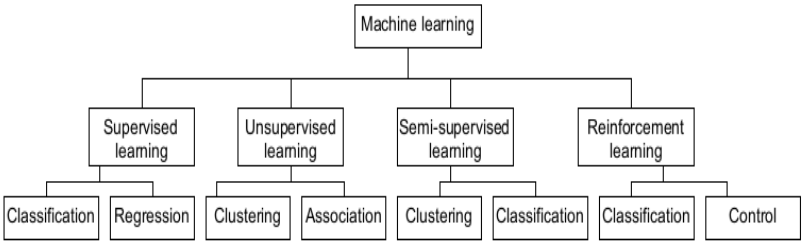


Figure 4. Various machine learning types.

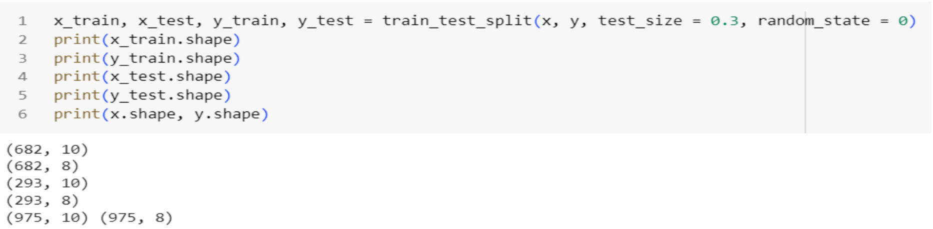
From some research, there were several good machine-learning models for the numerical data for our data set. For this research purpose, there was selected four suitable types of machine learning models. It may be increased in further deep kind if research. Linear Regression: This statistical technique looks at the connection between continuous variables. It predicts the dependent variable using one or more independent variables, fitting a line represented by the equation y = mx + b. This widely used technique finds applications in finance, economics, social sciences, and engineering (Fahrmeir et al. 2022).

Decision Trees: Decision Trees are efficient classification models organized in a tree structure, widely used for learning discriminatory patterns from data. The algorithm selects the best attribute for the root and partitions the examples into sets, creating nodes and branches accordingly through a top-down induction approach (Webb et al., 2011).

Random Forest: An ensemble learning technique called Random Forest is used for problems like classification, regression, and others. During training, it builds a lot of decision trees and produces the mean prediction (regression) or the mode of classes (classification). Random decision forests are used to address the issue of decision trees' propensity to overfit their training data. Despite frequently outperforming decision trees, random forests are less accurate than gradient-boosted trees (Attanasi & Coburn 2021).

K-Nearest Neighbors (KNN): K-Nearest Neighbors (KNN) is a simple approach that categorizes new instances based on a similarity metric, such as distance functions, and keeps all examples that are currently accessible. As a non-parametric technique, KNN has been utilized in statistical estimation and pattern recognition since the early 1970s. The algorithm selects K's nearest neighbors to the new data point and assigns the class label through majority voting. The classification distance metric may employ Euclidean, Manhattan, or Minkowski distance (Sarang 2023).
For training the dataset first of all need to understate the data properly. In this research, eight different columns need to be predicted which are detonated by Y and actual data is denoted by X.

y = df[['P1','p2','p3','p4','p5','p6','p7','p8']]
x = df [['Sex', 'Did you have your COVID test', 'Covid results', 'Physical problem prior to COVID-19', 'COVID-19 Duration', 'Did you consume any supplements or Medicine during COVID-19',' Did you used oxygen support', 'Did you used a ventilator while on COVID19',' Were you hospitalized', 'You spent time in the intensive care unit']]
After defining these tried to find out the actual shape of the dataset as well as choose the training and the test portion of the dataset randomly chose 30% for the test and 70 for the test.

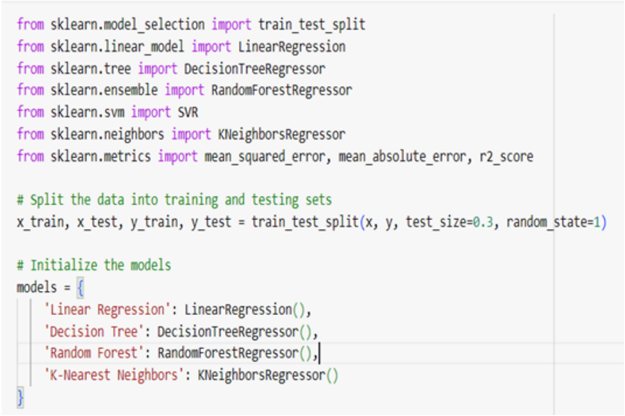


```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.3, random_state = 0)
print(x_train.shape)
print(y_train.shape)
print(x_test.shape)
print(y_test.shape)
print(x.shape, y.shape)
```

```
(682, 10)
(682, 8)
(293, 10)
(293, 8)
(975, 10) (975, 8)
```

Figure 5. Shope of the dataset

After taking the proper training and test part now this time to take the necessary library for training the dataset. There are so many built-in libraries in Python language. For general coding purposes, there were imported libraries such as NumPy, Pandas, Matplotlib, and Sklearn for the machine learning models. After all of the work tried to initialize the data for the algorithms



```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
from sklearn.neighbors import KNeighborsRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

# Split the data into training and testing sets
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=1)

# Initialize the models
models = {
    'Linear Regression': LinearRegression(),
    'Decision Tree': DecisionTreeRegressor(),
    'Random Forest': RandomForestRegressor(),
    'K-Nearest Neighbors': KNeighborsRegressor()
}
```

Figure 6. Initialize the dataset

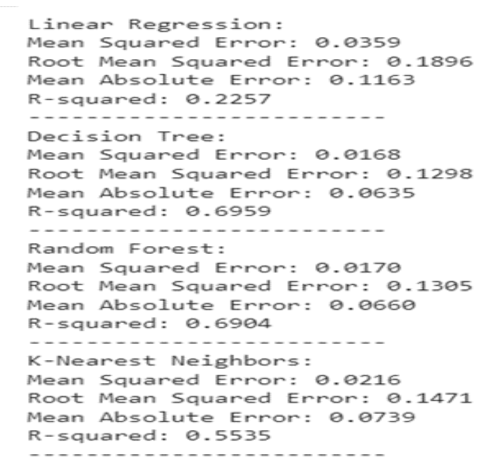
6. Result and Analysis

From all of the research, there are surely several outcomes from the COVID patients in their post-covid conditions. Due to the complex nature of post-COVID cases, applying a single general machine-learning algorithm can be challenging. The analysis reveals the presence of multiple features and entities linked to a COVID-19 patient's pre-existing condition. To address this complexity and improve prediction accuracy, a model is required that can integrate multiple machine learning models. There were already mentioned that this dataset has eight different outcomes. That means these algorithms have to predict an eight-dimensional (8D) array for the predictions. Here is the accuracy result for the applied models.

Table 2. Models Accuracy

Models	Accuracy
Linear Regression	98.50%
Decision Trees	84.34%
Random Forest	54.50%
K-Nearest Neighbors (KNN)	50.50%

for checking the biasness of the models and the perfectness of the trainnig dataset there were measured measn Squared error(MSE), root mean squarid Error(RMSE), Mean Absolute Error(MAE), R-squard error for the several machine learning models and findout some results. Mean Squared Error (MSE) measures the average squared difference between the predicted and actual values. It is used to evaluate how well a regression model performs on continuous data. Root Mean Squared Error (RMSE) is the square root of the mean squared error. It is used to evaluate how well a regression model performs on continuous data. Mean Absolute Error (MAE) measures the average absolute difference between the predicted and actual values. It is used to evaluate how well a regression model performs on continuous data. R-squared error measures how well the regression line fits the data points. It is used to evaluate how well a regression model performs on continuous data (Fahrmeir et al., 2021).



```
Linear Regression:
Mean Squared Error: 0.0359
Root Mean Squared Error: 0.1896
Mean Absolute Error: 0.1163
R-squared: 0.2257
-------------------------
Decision Tree:
Mean Squared Error: 0.0168
Root Mean Squared Error: 0.1298
Mean Absolute Error: 0.0635
R-squared: 0.6959
-------------------------
Random Forest:
Mean Squared Error: 0.0170
Root Mean Squared Error: 0.1305
Mean Absolute Error: 0.0660
R-squared: 0.6904
-------------------------
K-Nearest Neighbors:
Mean Squared Error: 0.0216
Root Mean Squared Error: 0.1471
Mean Absolute Error: 0.0739
R-squared: 0.5535
-------------------------
```

Figure 7. Different regression error results based on the machine learning model

The result shows that the two models have height accuracy linear regressions with 98.50% and the decision tree with 84.34%.

To evaluate the performance of the models there must need to make the confusion matrix Utilize metrics such as accuracy, precision, recall, and F1-score to evaluate the performance of the model. If,
True Positive (TP): The quantity of affirmative cases correctly predicted by the model,
True Negative (TN): The percentage of accurately anticipated negative situations by the model.,
False Positive (FP): The number of falsely projected negative instances as positive by the model,
False Negative (FN): The number of positive occurrences that the model misinterpreted as negative.,

Then, $\text{accuracy} = \frac{TN + TP}{TN + FP + TP + FN}$ (Sammut, 2011)
For the Liener Regressions performance, there were evaluate a graph where true value and predicted values were compared

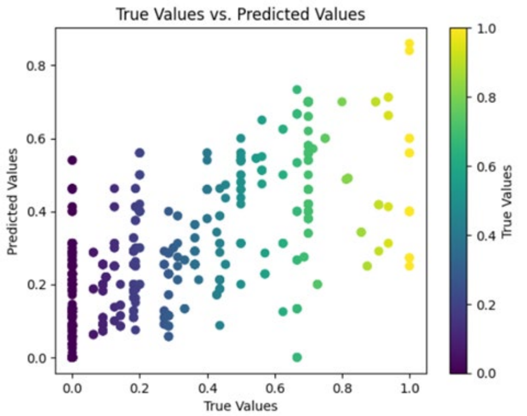
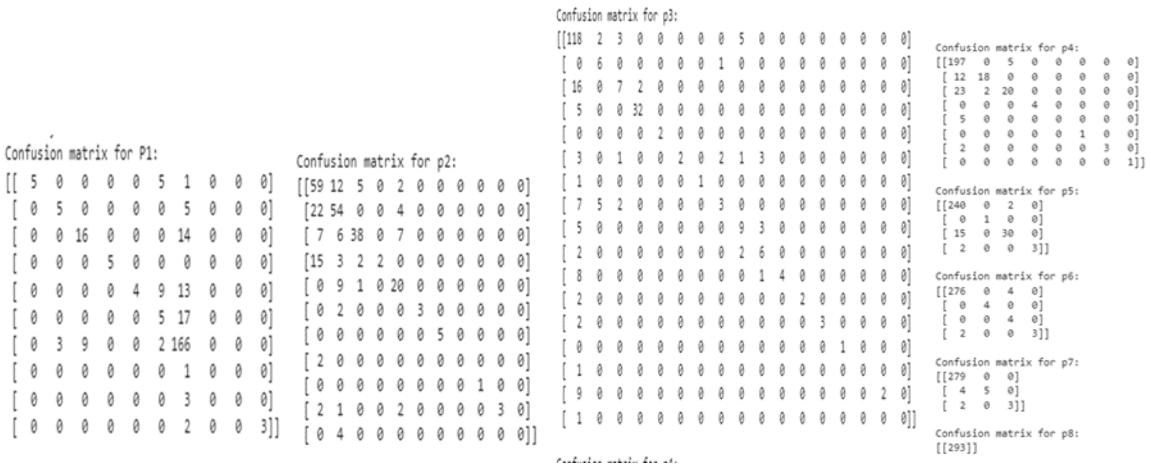


Figure 8. True Values vs Predicted values

On the other hand for the decession tree algorithm there were made confusion matrix for several outcomes.



```
Confusion matrix for P1:
[[  5   0   0   0   0   5   1   0   0   0]
 [  0   5   0   0   0   0   5   0   0   0]
 [  0   0  16   0   0   0  14   0   0   0]
 [  0   0   0   5   0   0   0   0   0   0]
 [  0   0   0   0   4   9  13   0   0   0]
 [  0   0   0   0   0   5  17   0   0   0]
 [  0   3   9   0   0   2 166   0   0   0]
 [  0   0   0   0   0   0   1   0   0   0]
 [  0   0   0   0   0   0   3   0   0   0]
 [  0   0   0   0   0   0   2   0   0   3]]

Confusion matrix for p2:
[[59 12  5  0  2  0  0  0  0  0  0]
 [22 54  0  0  4  0  0  0  0  0  0]
 [ 7  6 38  0  7  0  0  0  0  0  0]
 [15  3  2  2  0  0  0  0  0  0  0]
 [ 0  9  1  0 20  0  0  0  0  0  0]
 [ 0  2  0  0  0  3  0  0  0  0  0]
 [ 0  0  0  0  0  0  5  0  0  0  0]
 [ 2  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  1  0  0]
 [ 2  1  0  0  2  0  0  0  0  3  0]
 [ 0  4  0  0  0  0  0  0  0  0  0]]

Confusion matrix for p3:
[[118   2   3   0   0   0   0   0   5   0   0   0   0   0   0   0   0]
 [  0   6   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0]
 [ 16   0   7   2   0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  5   0   0  32   0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   2   0   0   0   0   0   0   0   0   0   0   0   0]
 [  3   0   1   0   0   2   0   2   1   3   0   0   0   0   0   0   0]
 [  1   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0]
 [  7   5   2   0   0   0   0   3   0   0   0   0   0   0   0   0   0]
 [  5   0   0   0   0   0   0   0   9   3   0   0   0   0   0   0   0]
 [  2   0   0   0   0   0   0   0   2   6   0   0   0   0   0   0   0]
 [  8   0   0   0   0   0   0   0   0   1   4   0   0   0   0   0   0]
 [  2   0   0   0   0   0   0   0   0   0   0   2   0   0   0   0   0]
 [  2   0   0   0   0   0   0   0   0   0   0   0   3   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0]
 [  1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  9   0   0   0   0   0   0   0   0   0   0   0   0   0   0   2   0]
 [  1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]]

Confusion matrix for p4:
[[197   0   5   0   0   0   0   0]
 [ 12  18   0   0   0   0   0   0]
 [ 23   2  20   0   0   0   0   0]
 [  0   0   0   4   0   0   0   0]
 [  5   0   0   0   0   0   0   0]
 [  0   0   0   0   0   1   0   0]
 [  2   0   0   0   0   0   3   0]
 [  0   0   0   0   0   0   0   1]]

Confusion matrix for p5:
[[240   0   2   0]
 [  0   1   0   0]
 [ 15   0  30   0]
 [  2   0   0   3]]

Confusion matrix for p6:
[[276   0   4   0]
 [  0   4   0   0]
 [  0   0   4   0]
 [  2   0   0   3]]

Confusion matrix for p7:
[[279   0   0]
 [  4   5   0]
 [  2   0   3]]

Confusion matrix for p8:
[[293]]
```

Figure 9. Confusion matrixes for Decision Trees

7. Conclusion

In this study, we investigated the possibility of predicting post-COVID complications using patient data and machine learning algorithms. Collaborating data scientists, computer scientists, doctors, and experts from various fields worked together to harness existing data and machine learning models for disease prediction and long-term health conditions. Our research focused on detecting post-COVID physical complexities, which can have long-lasting effects on individuals who have contracted COVID-19. Through rigorous analysis and the application of various machine learning models, we found that linear regression and decision trees exhibited the highest accuracy in predicting post-COVID complications, achieving 98.50% and 84.34% accuracy, respectively. These findings demonstrate the potential of machine learning in identifying patients at risk of developing post-COVID conditions and predicting the severity of symptoms. However, it is important to note that the complexity of post-COVID cases makes it challenging to rely solely on a single machine-learning algorithm. Therefore, integrating multiple models may lead to improved prediction accuracy.

References

A. Gupta, V. Jain, and A. Singh, "Stacking Ensemble-Based Intelligent Machine Learning Model for Predicting Post-COVID-19 Complications," New Generation Computing, Dec. 2021. [Online]. Available: https://doi.org/10.1007/s00354-021-00144-0.

CDC, "Long COVID or Post-COVID Conditions," Sep. 16, 2021. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html.

C. Sammut, Encyclopedia of machine learning : with 78 tables. New York, Ny Springer, 2011.

D. Munblit et al., "Studying the post-COVID-19 condition: research challenges, strategies, and importance of Core Outcome Set development," BMC Medicine, vol. 20, no. 1, Feb. 2022. [Online]. Available: https://doi.org/10.1186/s12916-021-02222-y.

D. Munblit et al., "Studying the post-COVID-19 condition: research challenges, strategies, and importance of Core Outcome Set development," BMC Medicine, vol. 20, no. 1, Feb. 2022. [Online]. Available: https://doi.org/10.1186/s12916-021-02222-y.

D. Menges et al., "Burden of post-COVID-19 syndrome and implications for healthcare service planning: A population-based cohort study," PLOS ONE, vol. 16, no. 7, p. e0254523, Jul. 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0254523.

E. D. Attanasi and T. C. Coburn, "Random Forest," pp. 1–4, Jan. 2021. [Online]. Available: https://doi.org/10.1007/978-3-030-26050-7_265-1.

G. I. Webb et al., "Decision Tree," Encyclopedia of Machine Learning, pp. 263–267, 2011. [Online]. Available: https://doi.org/10.1007/978-0-387-30164-8_204.

H. Weerahandi et al., "Post-Discharge Health Status and Symptoms in Patients with Severe COVID-19," Journal of General Internal Medicine, vol. 36, no. 3, pp. 738–745, Jan. 2021. [Online]. Available: https://doi.org/10.1007/s11606-020-06338-4.

I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," SN Computer Science, vol. 2, no. 3, pp. 1–21, Mar. 2021. [Online]. Available: https://doi.org/10.1007/s42979-021-00592-x

L. Fahrmeir, T. Kneib, S. Lang, and B. D. Marx, Regression. Berlin, Heidelberg: Springer Berlin Heidelberg, 2021. [Online]. Available: https://doi.org/10.1007/978-3-662-63882-8.

L. Wang, H. Shen, K. Enfield, and K. Rheuban, "COVID-19 Infection Detection Using Machine Learning," IEEE Xplore, Dec. 01, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9671700.

Ludwig Fahrmeir, T. Kneib, S. Lang, and B. D. Marx, Regression. Springer Nature, 2022.

M. A. Quiroz-Juárez et al., "Identification of high-risk COVID-19 patients using machine learning," PLOS ONE, vol. 16, no. 9, p. e0257234, Sep. 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0257234.

P. Sarang, "K-Nearest Neighbors," pp. 131–141, Jan. 2023. [Online]. Available: https://doi.org/10.1007/978-3-031-02363-7_6.

World Health Organization, "Coronavirus disease (COVID-19)," 2021. [Online]. Available: https://www.who.int/health-topics/coronavirus#tab=tab_1.

T. Greenhalgh et al., "Management of post-acute covid-19 in primary care," BMJ, vol. 370, no. 3026, p. m3026, Aug. 2020. [Online]. Available: https://doi.org/10.1136/bmj.m3026.

"What is Normalization in Machine Learning," Deepchecks. [Online]. Available: https://deepchecks.com/glossary/normalization-in-machine-learning/.

"What is Supervised Learning? | IBM," www.ibm.com. [Online]. Available: https://www.ibm.com/topics/supervised-learning .

Biographies

MD Nadim Hasan is a motivated and dedicated Computer Science and Engineering student at American International University-Bangladesh. Currently pursuing a BSc degree, Nadim has a strong academic background in Information Technology, specializing in areas such as database management, Management Information Systems, and basic programming languages. With a passion for exploring the world of computer science, Nadim actively conducts research in machine learning and seeks opportunities to apply theoretical knowledge to practical projects. Possessing a basic understanding of Data Science, Machine Learning, and Artificial Intelligence, Nadim is eager to further develop skills in these areas while demonstrating a strong aptitude for critical thinking and problem-solving.

Syed Mominin Islam Tamim is a passionate Junior Software Engineer at Venus IT Ltd and a student of American International University – Bangladesh (AIUB), currently engaged in a notable project with the Bangladesh Army. With a deep fascination for the applications of machine learning and artificial intelligence (AI), Tamim's academic journey was marked by exceptional achievements and a commitment to expanding his knowledge and skill set. At Venus IT Ltd, he has proven to be an invaluable asset, contributing his expertise to various software projects. Alongside his professional responsibilities, Tamim actively engages with the research community, attending conferences and workshops to broaden his knowledge and collaborate with peers. His enthusiasm extends to personal endeavors, where he explores cutting-edge technologies, works with diverse datasets, and implements machine learning models to address real-world problems. Looking ahead, Tamim aspires to make significant contributions to the field of machine learning and AI, collaborating with experts and driving technological advancements on a global scale. This conference marks his first step towards sharing his expertise and making a lasting impact in the academic realm, positioning him as an emerging professional in the field.

Md. Tafsimul Islam Tanzid is a dedicated and enthusiastic student pursuing a BSc in Computer Science and Engineering at AIUB. With a passion for technology and a keen interest in machine learning and artificial intelligence, Tanzid has embarked on a remarkable journey in the field of web development. Having worked as a Web Developer at Divergent Technologies Ltd, he has gained valuable industry experience and honed his skills in creating innovative and user-friendly web applications. Tanzid's notable accomplishments include successfully launching the live website naghmatune.com. Currently, he is actively involved in the HR-Connect project, showcasing his commitment to leveraging technology to enhance human resources management. With a thirst for knowledge and a drive for research, Tanzid is poised to make significant contributions to the world of machine learning and artificial intelligence