Post Covid-19 Complications Detection Using ML Review Paper

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

MD. Nadim Hasan
Student of American International University-Bangladesh (AIUB)
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 COVID-19 outbreak has brought about major health crises globally and the persistent nature of long COVID has put a significant strain on hospitals as a result of scarce healthcare resources. Heart problems are discovered to be the most prevalent among persons who have recovered from COVID-19 among several long-lasting effects (Gupta et al., 2021). There are so many cases of what is generally occurring with post-COVID patients. This work tries to find out the proper machine learning methods to predict the post-COVID complexities which can provide the highest accuracy. The true objective of this endeavor is to conduct additional machine learning-based research in this area of medical science and to develop an accurate machine-learning model. In the case of COVID patients, there are a lot of outcomes in a post-COVID condition. Proper problem definition, data acquisition, Model selection, evaluation of training and testing models, etc., are also significant obstacles to constructing a decent prediction model. There were proposed machine learning models according to previous research which model is combined with multiple models. This research might contribute to making further approaches to predicting the complexities of other diseases like COVID-19.

Keywords
machine learning, learning models, ensemble-based learning, COVID-19, post-COVID complexity.

1. Introduction
The SARS-COV-2 is the virus which makes the illness (Covid-19) condition of coronavirus. The majority of infected individuals will only develop mild to moderate respiratory infections and recover on their own, with the exception of a small number of people who will require medical care (WHO 2021). So many positive cases were coming since 2019 and people are suffering from this virus. However, A large number of vaccination led to a decline in several COVID-19 cases. Post-COVID issues, which are sporadically encountered by individuals infected with COVID, refer to the virus's long-lasting effects. While the majority of COVID-19 patients develop symptoms between a few days and a few weeks after contracting the virus, post-COVID complications can manifest at least four weeks after the initial infection (CDC 2021).

Machine learning is a branch of computing algorithms that is constantly evolving and aims to simulate human intelligence by acquiring knowledge from its environment. They are regarded as the workhorses of the new big data era

(El Naqa et al. 2015). It is possible to determine any kind of long-term cases of the diseases like covid-19 and make predictions from the data set from the Machine learning algorithm.

1.1 Objectives

There are so many data-scientists, computer scientists, doctors, and others who work together to predict diseases and long-term diseases by existing data and machine learning models. In this situation there must be a question that is it possible to predict the post-COVID complications with the existing data of a patient, and what will be the data processing criteria? The primary objective of this study is to develop a suitable machine learning model and plan for long-COVID or post-COVID complexity prediction using a machine-learning learning strategy.

2. General Overview

Decision-making, natural language processing, and pattern recognition are among the many duties that machine-learning algorithms perform. They work in a wide range of industries, including marketing, banking, and healthcare. Additionally, machine learning is utilized to enhance goods and services like voice assistants, self-driving cars, and recommendation systems (Sarker 2021). Models of machine learning exist in a diversity of forms.
Supervised learning is a technique in which a system employs labelled training data to discover patterns and predict outcomes for new data. It relies on the guidance provided by the labeled examples to make accurate predictions. Unsupervised learning, in which system is given raw data and left to discover patterns and connections on its own. In semi-supervised learning, both labelled and unlabeled inputs are provided to the system (IBM n.d.).

Nowadays machine learning is quite easier to make predictions by using ML model and various algorithm. Researchers are making prediction using ML and those are nearly perfect in most of the time. In these cases, there need the right tools is decision making as well as proper data interruption and statistical analysis. The quality of the values in the dataset, the choice of the best policy, and careful parameter tuning all have an impact on how effectively a machine learning model performs.

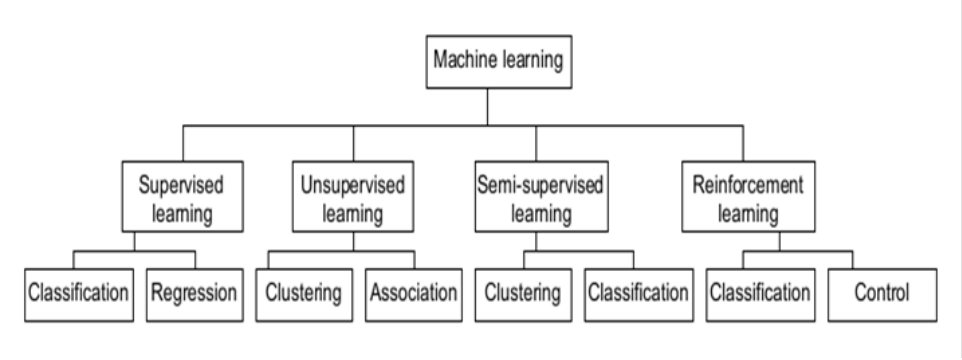

Figure 1. Various machine learning types

The process of using machine learning for prediction typically involves five steps. First, a dataset is collected and preprocessed, which may include cleaning and transforming the data, and splitting it into training and test sets. Next, a suitable machine learning model is chosen for the problem and data characteristics. The model is then trained using the training data, and its precision is evaluated using the test data. After the model is trained, it becomes capable of making predictions on new data based on the knowledge it has acquired during the training process (Géron 2019).

3. Literature Review and Features

Various studies and research publications have used machine learning models to analyses COVID-19 data, including disease progression prediction, risk factor identification, and CT scan analysis for diagnosis (Kwekha-Rashid et al. 2021).

Also, lots of paper found which research actually based on the physical complexity of a patient after have covid. Some examples of post-COVID complications include residual lung damage, cardiac issues, neurological complications, and certain mental health challenges. In some cases, more than a month after being discharged from the hospital, the

patient's physical and mental condition deteriorated (Weerahandi et al. 2021). By using that data some studies try to figure out the machine learning model as well as measure the accuracy. On the other hand, there are some researches whom proposed model uses stacking ensemble, which combines multiple models to improve the accuracy of predictions. The authors used information from patients diagnosed with post-COVID-19 to train as well as test the model. They found that the proposed model had a higher accuracy than a single model, and that it could be useful in predicting post-COVID-19 complications and identifying high-risk patients (Gupta et al. 2021). By doing systematic literature review the data are found is Google scholar, IEEExplore, Springer. Here, this work makes some forward and backward search. After conducting extensive research on the topic of "COVID complexity and machine learning prediction" and "machine learning algorithms and predictions," we identified ten high-quality papers that specifically delve into the study of post-COVID complexities. papers were selected mainly based on the relatable work and proper machine learning process.

Table 1. Working outcomes of COVID-19 and machine learning.

Paper Title	Main work and applied algorithm
1. Stacking Ensemble-Based Intelligent Machine Learning Model for Predicting Post-COVID-19 Complications	The proposed model, Main Outcomes, outperforms popular techniques such as random forests, decision trees, support vector machines, and artificial neural networks and obtains a remarkable 93.23 percent accuracy. The results clearly demonstrate the superior performance of Main Outcomes over existing baseline techniques. In addition, the proposed strategy exhibits a high precision of 95.74% and specificity, showcasing its effectiveness in diagnosing cardiac problems compared to other strategies. The strategy also demonstrates strong recall (92.05%) and accuracy (95.24%), further supporting its robust performance (Gupta et al., 2021).
2. Machine learning-based prediction of COVID-19 diagnosis based on symptoms	A machine learning technique was developed and trained using data from 51,831 test subjects (4,769 of whom were certain to have COVID-19). The test set included information from 47,401 individuals from the subsequent week, of which 3,624 were confirmed cases of COVID-19 (Zoabi et al., 2021).
3. Machine-Learning-Based COVID-19 Detection with Enhanced cGAN Technique Using X-ray Images.	This paper introduces a machine-learning technique that efficiently detects the presence of the SARS-CoV-2 virus in an RT-PCR test using only eight simple queries. The model was trained using data from Israeli citizens who underwent SARS-CoV-2 testing in the early phases of the pandemic. This methodology has global applicability and can priorities and screen testing for viruses in the general population (Hamdi et al., 2022).
4. Review on COVID ‐ 19 diagnosis models based on machine learning and deep learning approaches	This study examines over 200 papers and divides the research paths into two categories: ML and DL. In addition, it presents COVID-19 public datasets that were created and compiled by multiple nations (Alyasseri et al., 2021),
5. Predicting the Disease Outcome in COVID-19 Positive Patients Through Machine Learning: A Retrospective Cohort Study with Brazilian Data	The study database contains information on 8,443 patients whose cases have been concluded, indicating either recovery or death. Through rigorous experimentation, the results showcase the exceptional performance of the optimal prediction model. With a Receiver Operating Characteristic (ROC) AUC value of 0.92, a Sensitivity of 0.88, and a Specificity of 0.82, the model predicts the progression of the disease consistently and accurately. This demonstrates its efficacy in assisting healthcare professionals in making accurate disease outcome predictions (De Souza et al., 2021)
6. Prediction of COVID-19 using long short-term memory by integrating principal component analysis and clustering techniques	This study employs the long short-term memory (LSTM) deep learning technique to estimate the total number of COVID-19-positive cases across all 36 states in Nigeria. This approach is employed to capture and analyze patterns in the data for accurate predictions of COVID-19 cases. Principle component analysis (PCA) is used in the suggested method to choose significant features from the dataset and K-means clustering to find outliers (Ilu et al., 2022).

7. A Review on the Detection of the Post COVID-19 Symptoms for Long Term Diseased Patients using Machine Learning Algorithms	The paper investigates the use of machine learning techniques to identify and classify the effects of COVID-19 on individuals with pre-existing chronic conditions. It examines the use of these methods to analyses and categories the pre- and post-COVID-19 effects on patients with underlying health conditions. (Patibandla, 2022)
8. Prognosis patients with COVID-19 using deep learning	The study yielded the following results: maximum probability of correct choice (MPCD) of 0.93, area under the curve (AUC) of 0.93, F2 score of 0.93, recall of 1.00, accuracy of 0.95, precision of 0.91, and specificity of 0.9279. These metrics indicate the model's high performance in terms of correctly identifying and classifying the target variable (Guadiana-Alvarez et al., 2022).
9. Future forecasting prediction of Covid-19 using hybrid deep learning algorithm	On four distinct datasets, a performance analysis revealed that the proposed method outperforms Linear Regression (LR), Multinomial Naive Bayesian (MNB), Random Forest (RF), Stochastic gradient boosting (SGB), and Decision Tree (DT) by 1.40 percentage points, 3.39 percentage points, and 5.32 percentage points, respectively. These results indicate that the suggested method outperforms other techniques across multiple datasets, as demonstrated in the study (Yenurkar & Mal, 2022).
10. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)	In this study, a novel approach was developed by modifying the Inception transfer-learning model, which was then validated internally and externally. During internal validation, the method had an accuracy of 89.5%, a specificity of 0.88, and a sensitivity of 0.87. External testing determined that the dataset's overall accuracy was 79.3%, with a specificity of 0.83 and a sensitivity of 0.67. The algorithm correctly predicted 85.2 percent of COVID-19-positive images, including 46 of 54 that tested negative in the first two nucleic acid tests. These results demonstrate the effectiveness and dependability of the international, publicly accessible, and current Inception transfer datasets. model for learning to recognize instances of COVID-19Evaluate the model's efficacy using multiple metrics, including its accuracy, precision, recall, and F1-score.

Going forward, dealing with health equity issues and interacting with those who have post-COVID-19 conditions will probably be essential. Researchers and medical professionals have learned a great deal about the post-COVID-19 syndrome through collaboration with patient advocates, from its basic existence to its variety of intricate symptoms. The creation of care that addresses their health and well-being will be made possible by continuing to incorporate a diverse variety of people with lived experiences. This will aid in better understanding the post-COVID-19 state (Vu & McGill 2021).

For making proper prediction, there need to consider few things of the patients. Some of them had some health issues like Diabetics, Asthma, high blood pressure, Heart problem, Hyperlipidemia etc. (Weerahandi et al. 2021). So, it is very complex to make focus on the valuable feature's predictions. This study drew its data from a study that examined the health status and symptoms of patients with severe COVID-19 following hospital discharge. The survey involved the participation of 161 individuals, who were selected for the purpose of conducting a comprehensive analysis of post-COVID outcomes.

Prior covid	During covid	Post-Covid
1. Diabetes	1. lungs infection (highly infected)	1. Tiredness
2. Hyper tension (high blood pressure)	2. lungs infection (hardly infected)	2. Mental trauma
3. Hyper tension (low blood pressure)	3. high fever.	3. Difficult breathing
4. Thyroid	4. shortage of oxygen.	4. Chest pain
5. Irregular menstrual cycle.	5. Mental breakdown	5. Heart palpitation
6. lungs issues		6. Hair fall

7. kidney issues 8. heart issues 9. asthma 10. None		7. Headache 8. Sleeping problem 9. Memory lost issue 10. Rash 11. Stomach pain 12. lungs issues 13. change menstrual cycle 14. others 15. None

4. Data Collection

Here's the proposed methodology for training a machine learning model for multiple prior COVID data and multiple predicting multiple post-COVID complications based on those prior and during COVID health issues.

- Problem Definition: we will train our model based on prior and during COVID health issues and based on those data our model will predict post-COVID complications.
- Data Collection: Collect and pre-process the data, making sure it is cleaned, formatted, and ready to be used.
- Feature Engineering: Extract relevant features from the data to create new ones that can help improve the model's performance.
- Model Selection: Choose a model that is suitable for the problem at hand, taking into consideration the complexity of the data, the number of features, and the desired outcome. (Neural network, Decision tree, SVM, Random Forest, Gradient Boosting, KNN).
- Model Training: Using techniques such as cross-validation to prevent overfitting and determine the optimal hyperparameters, train the model on the preprocessed data.
- Model Evaluation: Use metrics such as accuracy, precision, recall, and F1-score to assess the model's performance. If,

  True Positive (TP): The number of positive cases that the model correctly predicted,
  True Negative (TN): The number of negative cases that the model correctly predicted,
  False Positive (FP): The number of negative cases incorrectly predicted as positive by the model,
  False Negative (FN): The number of positive cases incorrectly predicted as negative by the model.,

  Then,

  $$\text{accuracy} = \frac{TN+TP}{TN+FP+TP+FN}$$ (Sammut, 2011)

- Model Refinement: If necessary, refine the model by making changes to the hyperparameters, adding or removing features, or choosing a different model altogether.
- Model Deployment: Deploy the model in a production environment and monitor its performance.

This methodology is a good starting point, but it may need to be adjusted based on the specific requirements of your project (De Souza et al., 2021) (Aljameel et al. 2021)

5. Results and Discussion

From all of the research, there are surely several outcomes form 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 patient's pre-existing condition. To address this complexity and improve prediction accuracy, a model is required that can integrate multiple machine

learning models. This proposed methodology aims to combine various ML models effectively to make the best predictions, considering the diverse factors associated with post-COVID outcomes. After all the research, this research proposed an ensemble model to get maximum accuracy of after-COVID complexity. In order to improve classification, prediction, function approximation, and other aspects of sentiment classification, ensemble-based learning approaches the performance of a model or lessen the possibility of making a wrong move (Narayan et al. 2016)

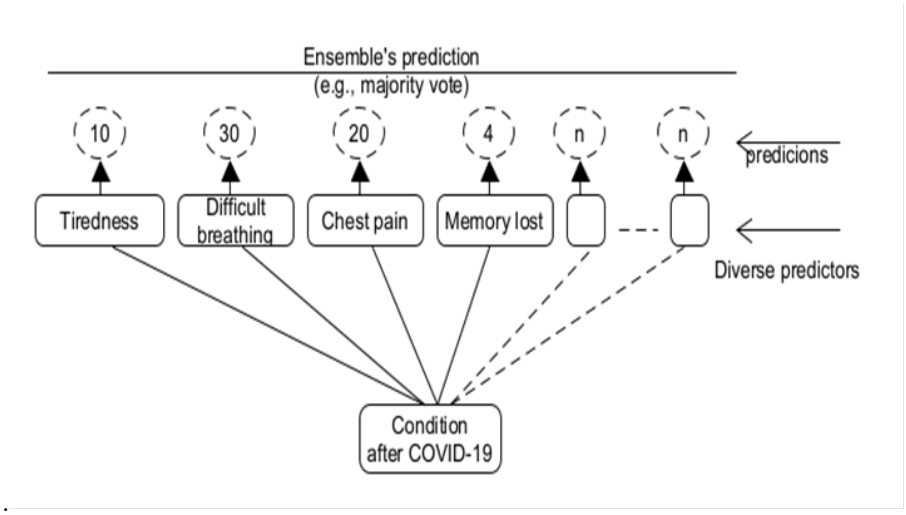

Figure 2. Ensemble's predictions

Using the model, it can separately predict multiple outcomes like tiredness, memory lost, chest pain etc. as well as it can diverse the predictions. The suggested model is tested against different benchmark methods, including decision trees, support vector machines, random forest as well as artificial neural networks. network etc.

6. Conclusion

In conclusion, this conference paper has delved into the potential of machine learning to predict post-COVID-19 complications. The COVID-19 pandemic has created a global health emergency, as millions of people have been infected and a significant proportion of them have developed long-term complications. By using machine learning techniques, this study has shown that it is possible to identify individuals who may be at a higher risk of developing these complications. This research demonstrates the effectiveness of machine learning in analyzing and predicting complex medical data.

Using machine learning in healthcare has the ability to completely change how medical professionals approach patient care. By accurately identifying individuals who may be at risk of post-COVID-19 complications, healthcare providers can prioritize their care, allocate resources more efficiently, and ensure that they receive the treatments they need as soon as possible. Additionally, this research has important implications for the development of preventative measures and treatment protocols for post-COVID-19 patients.

This study also emphasizes the significance of ongoing monitoring and evaluation of post-COVID-19 patients, as well as the need for additional research in this area. The application of machine learning in this field is still in its infancy, and there is much work to be done to refine and enhance these techniques. However, the outcomes of the research show that the ML has tremendous potential for improving patient outcomes and ensuring that post-COVID-19 patients receive the care they need.

In conclusion, this conference paper has demonstrated the power of machine learning in predicting post-COVID-19 complications, and it provides a glimpse into the future of medical care. The outcomes of this research will be of interest to healthcare providers, researchers, and policymakers alike, and they highlight the need for continued investment in this area of research

References

Aljameel, S. S., Khan, I. U., Aslam, N., Aljabri, M., & Alsulmi, E. S., Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients. Scientific Programming, 2021, 1–10, 2021. https://doi.org/10.1155/2021/5587188.

Alyasseri, Z. A. A., Al-Betar, M. A., Doush, I. A., Awadallah, M. A., Abasi, A. K., Makhadmeh, S. N., Alomari, O. A., Abdulkareem, K. H., Adam, A., Damasevicius, R., Mohammed, M. A., & Zitar, R. A. (2021). Review on COVID-19 diagnosis models based on machine learning and deep learning approaches. Expert Systems, 2021. https://doi.org/10.1111/exsy.12759.

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Inc.

Centers for Disease Control and Prevention. (2021, September 16). Long COVID or Post-COVID Conditions. https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html.

De Souza, F. S. H., Hojo-Souza, N. S., Dos Santos, E. B., Da Silva, C. M., & Guidoni, D. L., Predicting the Disease Outcome in COVID-19 Positive Patients Through Machine Learning: A Retrospective Cohort Study with Brazilian Data. Frontiers in Artificial Intelligence, 4, 2021. https://doi.org/10.3389/frai.2021.579931.

Guadiana-Alvarez, J. L., Hussain, F., Morales-Menendez, R., Rojas-Flores, E., García-Zendejas, A., Escobar, C. A., Ramírez-Mendoza, R. A., & Wang, J., Prognosis patients with COVID-19 using deep learning. BMC Medical Informatics and Decision Making, 22(1), 2022. https://doi.org/10.1186/s12911-022-01820-x.

Gupta, A., Jain, V., & Singh, A. (2021). Stacking Ensemble-Based Intelligent Machine Learning Model for Predicting Post-COVID-19 Complications. New Generation Computing, 2021. https://doi.org/10.1007/s00354-021-00144-0.

Hamdi, M., Ksibi, A., Ayadi, M., Elmannai, H., & Alzahrani, A. I. A. (2022). Machine-Learning-Based COVID-19 Detection with Enhanced cGAN Technique Using X-ray Images. Electronics, 11(23), 3880, 2022. https://doi.org/10.3390/electronics11233880.

IBM. (n.d.). What is Supervised Learning? | IBM. https://www.ibm.com/topics/supervised-learning.

Ilu, S. Y., Rajesh, P., & Mohammed, H. (2022). Prediction of COVID-19 using long short-term memory by integrating principal component analysis and clustering techniques. Informatics in Medicine Unlocked, 31, 2022, 100990. https://doi.org/10.1016/j.imu.2022.100990.

Issam El Naqa, Li, R., & Murphy, M. J. , Machine Learning in Radiation Oncology: Theory and Applications. Springer International Publishing, 2015.

Kwekha-Rashid, A. S., Abduljabbar, H. N., & Alhayani, B. (2021). Coronavirus disease (COVID-19) cases analysis using machine-learning applications. Applied Nanoscience, 2021. https://doi.org/10.1007/s13204-021-01868-7.

Narayan, R., Roy, M., & Dash, S. , Ensemble-based Hybrid Machine Learning Approach for Sentiment Classification-A Review. International Journal of Computer Applications, 146(6), 31–36, 2016. https://doi.org/10.5120/ijca2016910813.

Patibandla, A. , A Review on the Detection of the Post COVID-19 Symptoms for Long Term Diseased Patients using Machine Learning Algorithms. Journal of Physics: Conference Series, 2327(1), 012073, 2022. https://doi.org/10.1088/1742-6596/2327/1/012073.

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

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science, 2(3), 1–21, 2021. https://doi.org/10.1007/s42979-021-00592-x.

Vu, T., & McGill, S. C., An Overview of Post–COVID-19 Condition (Long COVID). Canadian Journal of Health Technologies, 1(9), 2021. https://doi.org/10.51731/cjht.2021.160.

Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., Cai, M., Yang, J., Li, Y., Meng, X., & Xu, B. , A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19), 2020. https://doi.org/10.1101/2020.02.14.20023028.

Weerahandi, H., Hochman, K. A., Simon, E., Blaum, C., Chodosh, J., Duan, E., Garry, K., Kahan, T., Karmen-Tuohy, S. L., Karpel, H. C., Mendoza, F., Prete, A. M., Quintana, L., Rutishauser, J., Santos Martinez, L., Shah, K., Sharma, S., Simon, E., Stirniman, A. Z., & Horwitz, L. I. , Post-Discharge Health Status and Symptoms in Patients with Severe COVID-19. Journal of General Internal Medicine, 36(3), 738–745, 2021. https://doi.org/10.1007/s11606-020-06338-4.

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

Yenurkar, G., & Mal, S. (2022). Future forecasting prediction of Covid-19 using hybrid deep learning algorithm. Multimedia Tools and Applications, 2022. https://doi.org/10.1007/s11042-022-14219-7.

Zoabi, Y., Deri-Rozov, S., & Shomron, N. (2021). Machine learning-based prediction of COVID-19 diagnosis based on symptoms. Npj Digital Medicine, 4(1), 1–5, 2021. https://doi.org/10.1038/s41746-020-00372-6.

Biographies

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. 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.

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.