Connecting the Dots: Linking Coronary Diseases with Covid-19 Patients through Support Vector Machine Algorithm

K. H. Pranavi
Department of IT, Student III year,
GVP College of Engineering for Women,
India
E-mail: pranavikannepogu@gmail.com

M. Bhanu Sridhar
Department of IT, HOD,
GVP College of Engineering for Women,
India
E-mail: sridharbhanu@gvpcew.ac.in

***Abstract*– The COVID-19 pandemic has left a lasting impact on global health, with a significant portion of survivors experiencing persistent health effects termed as 'LONG COVID' or 'POST COVID-19 SYNDROME'. In this research, we propose a novel approach utilizing Support Vector Machine (SVM) algorithm to analyse patient data and predict the multifaceted nature of post-COVID-19 Syndrome, particularly focusing on the interlinkage of coronary diseases with COVID-19 patients. Our methodology involves collecting and analysing extensive patient data, including pre-conditions and post-conditions of COVID-19, to identify patterns and associations between various health issues. By leveraging the high-dimensional capabilities of SVM, we aim to provide accurate predictions and insights into the long-term health complications of COVID-19 survivors, thereby contributing to a better understanding of this critical area of healthcare. This approach stands out due to its ability to handle nonlinear relationships, noise in data, and large datasets effectively, offering valuable insights for healthcare professionals in managing post-COVID-19 complications.**

Keywords: Multisystem Inflammatory Syndrome, Post-COVID-19 Syndrome, Coronary Diseases, Long-term Health Effects, Vaccine.

I. INTRODUCTION

Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. Most people infected with the virus experienced mild to moderate respiratory illness and recover without requiring special treatment. However, some people were seriously ill and required medical attention. Older people and those with underlying medical conditions like cardiovascular disease, diabetes, chronic respiratory disease, or cancer were more likely to develop serious illness. It has profound impact on global health, economies and daily life. Though there is a decrease in the rapid growth of the virus, some continued to experience a range of persistent health effects by the COVID patients. These complication and symptoms are known as "Post-COVID Effects" or "Post-COVID Syndrome". These effects are vast and affecting various organ systems. Survivors report ongoing fatigue, respiratory problems, chest pain and other issues. Therefore, understanding and addressing post-COVID effects on health is concerned as it not only impacts the quality of life for affected individuals but also holds implications on healthcare.

II. MULTISYSTEM INFLAMMATORY SYNDROME IN CHILDREN(MIS-C)

Some of the earliest evidence that SARS-CoV-2 infection leads to dysregulated immune responses came from paediatric patients who presented with multisystem inflammatory syndrome in children (MIS-C) which involves diffuse organ system involvement and a clinical spectrum that overlaps with other hyperinflammatory syndromes. Multisystem inflammatory syndrome in children is rare but serious condition that has been reported in children adolescents after they have been infected with or exposed to COVID-19. It is characterized by widespread inflammation that affects multiple organ systems in the body. It typically occurred several weeks after initial COVID-19 infection and presented with a variety of symptoms.

MIS-C symptoms can vary widely but often include fever, abdominal pain, diarrhoea, red eyes, etc. These affect multiple organ systems which leads to serious complications such as cardiac dysfunction, respiratory and neurological issues. This is a result of an abnormal immune response triggered by the initial COVID-19 infection leading to excessive inflammation throughout the body. Therefore, understanding how COVID-19 affects the risk of post-COVID-19 complications such as autoimmune disease will help to implement preventive measures and early treatment in individuals who have had COVID-19 to prevent morbidity and mortality.

III. LITERATURE SURVEY

The literature review in this research paper indeed offers a thorough overview of existing research and developments related to COVID-19 and its impact on health. The references cited cover a range of topics, including health disparities, mental health effects, vaccine efficacy and side effects, long-term health effects of COVID-19, and analysis of preventive behaviour. Additionally, the inclusion of diverse sources such as journals, conference proceedings, and online datasets suggests a comprehensive approach to gathering information and staying up-to-date with the latest developments in the field. This indicates the authors' commitment to presenting a well-informed and evidence-based discussion within their research paper.

This work is licensed under a Creative Commons Attribution-Noncommercial 4.0 International License

Asian Journal of Convergence in Technology ***Volume X and Issue I***
ISSN NO: 2350-1146 I.F-5.11

Beatty and Alexis L. in the analysis of vaccine types and adverse effects concluded that many adults who received a COVID-19 vaccination are facing adverse effects basing on some of the factors such as full vaccination dose, brand of vaccine, younger age etc. The most common side effects were found to be injection site pain, fatigue, headache, muscle pain, chills and joint pains are declared by Riad and Abanoub whereas post-Covid syndrome includes some of the possible symptoms like lung(respiratory) symptoms, neurological symptoms, joint and muscle pains, blood clots by Mayo Clinic Staff regarding long term effects of COVID-19.

Overall, the literature review showcases the authors' familiarity with the subject and strengthens the credibility of their own research by situating it within the broader context of existing knowledge and studies in the field of COVID-19 and its health effects.

IV. VACCINATION REACTIONS

After the rapid growth of coronavirus infection, there was a release of vaccinations. But there are also some adverse effects and factors by the people who took the vaccinations. These effects may be long-term. Factors that most strongly associated with adverse effects were full vaccination dose, brand of vaccine, younger age and having had COVID 19 before vaccination. This conclude that some individuals experienced more adverse effects after COVID-19 vaccination, but serious adverse effects are rare.

In randomized clinical trials of COVID-19 vaccines, reported adverse effects included injection site events (such as pain, redness, swelling) and systemic effects (for example, fatigue, headache, muscle or joint pain), with rare serious adverse events. Most adverse effects were mild, but studies reported approximately 50% to 90% of participants experiencing some adverse effects.

V. DATA COLLECTION

Collecting data on the impact of COVID-19 on health is a crucial step in understanding the long-term effects of disease. It is observed that with increase in the variants of COVID-19 waves the number of cases and the deaths have increased. Depending on the nature of the data, we use appropriate methods or algorithms to analyse the collected information.

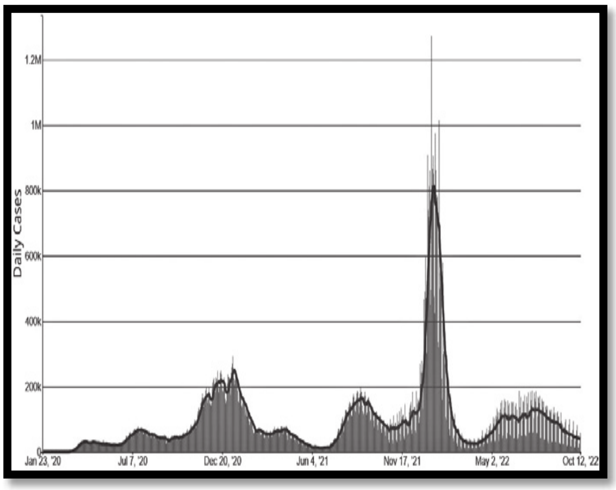

Fig. 1. Epidemic waves in the SARS-CoV-2 pandemic. From left to right: Beta, Delta, early Omicron variants, and Omicron BA.4/BA.5.

On collecting and comparing the past data on health before COVID-19 and the present data post COVID, it is seen that the health rate is rapidly going down for which it is important to feel concerned about the lives of people. And it is also observed that the COVID-19 survivors are more prominently seen in these conditions and its potential long-term consequences.

VI. INTERLINK OF DISEASES

On research, it is observed that the coronavirus results in increased rates of respiratory failure and death among infected individuals. It is also found that the people with diabetes, heart disease and asthma are pretentious to the post COVID effects. These conditions are known to disproportionately affect the ones who already are impacted by the health disparities.

It is also witnessed that unintended consequences of the COVID-19 outbreak; patients with non-COVID-19 medical conditions avoid healthcare to reduce the risk of COVID-19 exposure. Hence it was hard to find the cause of patient whether it was either due to COVID-19 or not.

As the health disparities impact more on the COVID-19 survivors, we can say that there is an interlink of diseases on post COVID effects.

VII. ANALYSIS THROUGH DATASET

1	SEX	P_TYPE	AGE	INTUBED	PNEUMONI	PREGNANT	DIABETES	ASTHMA	HIPERTENSI	CARDIOVAS	OBESITY	TOBACCO	ICU
2	1	1	65	97	1	2	2	2	1	2	2	2	97
3	2	1	72	97	1	97	2	2	1	2	1	2	97
4	2	2	55	1	2	97	1	2	2	2	2	2	2
5	1	1	53	97	2	2	2	2	2	2	2	2	97
6	2	1	68	97	2	97	1	2	1	2	2	2	97
7	1	2	40	2	1	2	2	2	2	2	2	2	2
8	1	1	64	97	2	2	2	2	2	2	2	2	97
9	1	1	64	97	1	2	1	2	1	2	2	2	97
10	1	2	37	2	2	2	1	2	1	2	1	2	2
11	1	2	25	2	2	2	2	2	2	2	2	2	2
12	1	1	38	97	2	2	2	2	2	2	2	2	97
13	2	2	24	2	2	97	2	2	2	2	2	2	2
14	2	2	30	2	2	97	2	2	2	2	2	2	2
15	2	1	55	97	2	97	2	2	2	2	2	2	97
16	1	1	48	97	2	2	1	2	2	2	2	2	97
17	1	1	23	97	2	2	2	2	2	2	2	2	97
18	1	2	80	2	1	2	2	2	1	2	2	2	1
19	2	1	61	97	2	97	2	2	2	2	2	2	97
20	2	1	54	97	2	97	2	2	2	2	2	2	97

Fig. 2. Dataset of the patient health information including pre-conditions and post-conditions of COVID-19

The dataset was provided by the Mexican government. This dataset contains an enormous number of anonymized patient-related information including pre-conditions. The raw dataset consists of 21 unique features and 1,048,576 unique patients. **In the Boolean features, 1 means "yes" and 2 means "no". values as 97 and 99 are missing data**.

- sex: 1 for female and 2 for male.
- age: of the patient.
- p-type (patient type): type of care the patient received in the unit. 1 for returned home and 2 for hospitalization.
- pneumonia: whether the patient already have air sacs inflammation or not.
- pregnancy: whether the patient is pregnant or not.
- diabetes: whether the patient has diabetes or not.
- asthma: whether the patient has asthma or not.

- hypertension: whether the patient has hypertension or not.
- cardiovascular: whether the patient has heart or blood vessels related disease.
- obesity: whether the patient is obese or not.
- tobacco: whether the patient is a tobacco user.
- intubed: whether the patient was connected to the ventilator.
- icu: Indicates whether the patient had been admitted to an Intensive Care Unit.
- date died: If the patient died indicate the date of death, and 9999-99-99 otherwise.

VIII. SUPPORT VECTOR MACHINES

In this study, the proposed approach was implemented using a standard desktop computer with an Intel Core i7 processor and 16GB of RAM. The analysis and modelling tasks were carried out using Python programming language, leveraging popular libraries such as scikit-learn for machine learning algorithms and data preprocessing, and pandas for data manipulation. Support vector machines can handle high dimensional data this is important which means this type of real-world data sets are high dimensional they have a large number of features. SVM can handle the type of data without sacrificing accuracy. It can handle data with non-linear relationships and is important because such type of datasets can have nonlinear relationships. SVM can learn these relationships and make accurate predictions even when relationships are complex as the data has complex relationships for which SVM suits the best. SVM are robust to noise in the data it is important because the real time data sets contain noisy data but can ignore noise and make accurate predictions even when the data is noisy. Overall, SVM is a powerful machine learning algorithm that can be used to solve variety of classification and regression problems where the data is high dimensional non-linear and noisy.

Fig. 3.

The data of 100 patients is used to predict the accurate information avoiding the noisy data.

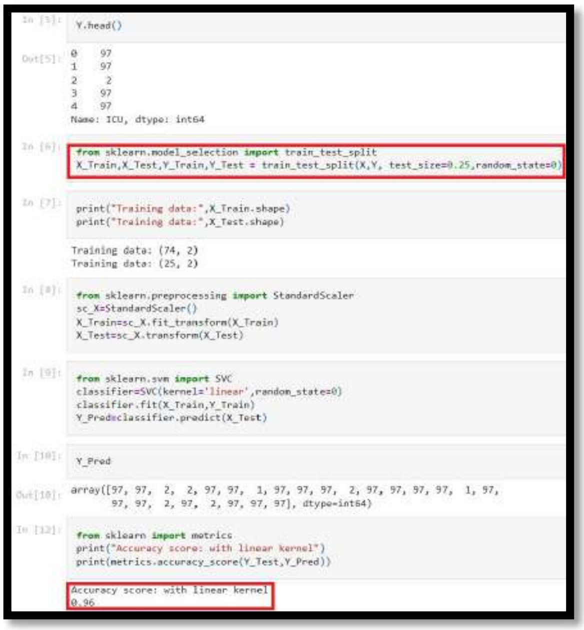

Fig. 4. Performance of the SVM model

Certainly, as the data is performed on SVM it is observed that the accuracy turned out to be 96 percent. It is done in such a way that 75 percent of the data is trained and 25 percent of the data is tested. Likewise, the data is predicted to achieve the accuracy. Hence it is appropriate to use SVM as it gave the accurate values when compared with other algorithms.

IX. COMPARISION OF SVM WITH OTHER MACHINE LEARNING ALGORITHMS

One of the most used algorithms other than Support Vector Machines that gives more accurate data and accurate predictions is Random forest. SVM is a discriminative classifier that finds the hyperplane that best separates classes in the feature space whereas Random Forest are an ensemble learning method that operates by constructing a multitude of decision trees at training time. Similarly, SVM can handle non-linear relationships through the use of kernel functions, which map input data into a higher-dimensional space and Random Forest naturally handle non-linearity without the need for explicit transformation, as they aggregate the predictions of multiple decision trees.

The choice between SVM and Random Forest depends on the nature of the data, the problem at hand, considerations such as interpretability, computational resources, and the size of the dataset. It's often beneficial to experiment with both algorithms and choose the one that performs better for a specific task. But when it comes to the above data, the dataset used is large. To avoid noisy data in such datasets in the process of training data with Random Forest is quite difficult. Due to which Support Vector Machines are used.

X. LIMITATIONS AND POTENTIAL CHALLENGES

The primary challenges encountered during the implementation of the approach:

A. Data Availability and Quality:

While the utilized dataset obtained from the Mexican government, the completeness and accuracy of the data has varied. Missing or incomplete data entries, as denoted by values such as 97 or 99, posed challenges during preprocessing that affected the performance of our models.

B. Model Interpretability:

Although Support Vector Machine (SVM) model yielded high predictive accuracy in this study, it is inherently less interpretable compared to other machine learning algorithms. Understanding the underlying decision boundaries and feature importance can be challenging, especially in complex datasets with high dimensionality.

C. Ethical and Privacy Considerations:

Lastly, ethical and privacy considerations must be carefully addressed when working with sensitive healthcare data. While our study utilized anonymized patient information, ensuring compliance with data protection regulations and maintaining patient confidentiality remains paramount. Future research endeavours should prioritize transparency and accountability in data handling practices to uphold ethical standards and foster trust among stakeholders.

XI. CONTRIBUTIONS AND NOVELTY

The research makes several significant contributions to the field of post-COVID health outcomes prediction and healthcare management:

A. Integration of Support Vector Machine (SVM) Algorithms:

Our study introduces the novel application of SVM algorithms for analysing patient data and predicting post-COVID health outcomes. By leveraging SVM's robustness to high-dimensional data, non-linear relationships, and noise, it can provide accurate predictions of long-term health complications in COVID-19 survivors. This integration of SVM into the healthcare domain extends the repertoire of machine learning techniques for addressing complex health challenges.

B. Interlinkage of Coronary Diseases with COVID-19:

The study sheds light on the interlinkage between coronary diseases and COVID-19, highlighting the importance of understanding the multifaceted nature of post-COVID health complications. By identifying associations between pre-existing conditions, COVID-19 infection status, and long-term health outcomes, it can provide insights into potential risk factors and preventive measures for healthcare professionals.

XII. CONCLUSION

This research offers valuable insights into the complex interplay between COVID-19 infection and post-acute health complications, particularly focusing on the association between coronary diseases and COVID-19 survivors. By utilizing Support Vector Machine (SVM) algorithm, it has demonstrated the efficacy of predictive modeling in identifying individuals at risk of long-term health challenges following COVID-19 infection.

The findings of the study underscore the importance of understanding the multifaceted nature of post-COVID health outcomes and the need for personalized healthcare interventions. By identifying high-risk patient populations and primitively addressing their health needs, healthcare providers can mitigate the long-term impact of the pandemic and improve patient outcomes.

Looking ahead, there are several avenues for further research and application. Longitudinal studies tracking COVID-19 survivors over extended periods can provide valuable insights into the trajectory of post-acute health complications and the effectiveness of interventions over time. Integrating multimodal data sources, such as genomic and clinical imaging data, can enhance predictive modelling accuracy and facilitate the discovery of novel biomarkers for risk assessment.

XIII. FUTURE RESEARCH AND POTENTIAL IMPROVEMENNTS

A. Incorporating Multimodal Data Sources:

Future research endeavours could explore the integration of multimodal data sources, including genomic, proteomic, and clinical imaging data, to enhance the predictive capabilities of our models. By combining diverse data modalities, researchers can capture a more comprehensive understanding of the biological mechanisms underlying post-COVID health complications and identify novel biomarkers for risk assessment.

B. Ensemble Learning Approaches:

Additionally, incorporating ensemble learning approaches such as Random Forest or Gradient Boosting Machines could further enhance the predictive performance of our models. Ensemble methods combine multiple base learners to improve prediction accuracy and generalizability, offering a promising avenue for refining our integration approach and mitigating overfitting.

C. Enhanced Model Interpretability:

Improving the interpretability of predictive models is essential for facilitating actionable insights and decision-making in clinical settings. Future research efforts should focus on developing interpretable machine learning techniques, such as feature importance analysis and model explainability methods, to elucidate the underlying factors driving post-COVID health outcomes and inform personalized interventions.

D. Longitudinal Studies and Real-time Monitoring:

Longitudinal studies tracking patients over extended periods could provide valuable insights into the trajectory of post-COVID health complications and the efficacy of interventions over time. Furthermore, implementing real-time

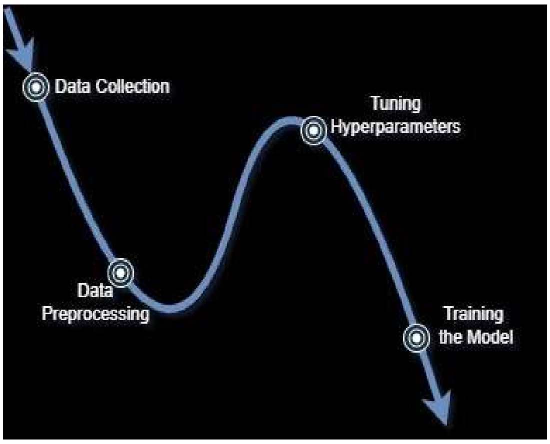

Fig. 5.

monitoring systems leveraging electronic health records and wearable devices could enable proactive healthcare management and early detection of adverse outcomes among COVID-19 survivors.

REFERENCES

[1] Kim, E.J., Marrast, L. & Conigliaro, J. COVID-19: Magnifying the Effect of Health Disparities. J GEN INTERN MED 35, 2441–2442 (2020). https://doi.org/10.1007/s11606-020-05881-4

[2] Fiorillo, Andrea, and Philip Gorwood. "The consequences of the COVID-19 pandemic on mental health and implications for clinical practice." European psychiatry 63.1 (2020): e32.

[3] Boyle, Laura, et al. "Selective sweeps in SARS-CoV-2 variant competition." Proceedings of the National Academy of Sciences 119.47 (2022): e2213879119.

[4] Sharma, Chetan, and Jagadeesh Bayry. "High risk of autoimmune diseases after COVID-19." Nature Reviews Rheumatology (2023): 1-2.

[5] Beatty, Alexis L., et al. "Analysis of COVID-19 vaccine type and adverse effects following vaccination." JAMA network open 4.12 (2021): e2140364-e2140364.

[6] “COVID-19 Dataset” https://www.kaggle.com/datasets/meirnizri/covid19-dataset

[7] Riad, Abanoub, et al. "Prevalence of COVID-19 vaccine side effects among healthcare workers in the Czech Republic." Journal of clinical medicine 10.7 (2021): 1428.

[8] “COVID-19: Long-term effects”, By Mayo Clinic Staff, June 22, 2023 https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/coronavirus-long-term-effects/art-20490351

[9] Kim, Sunhee, and Seoyong Kim. "Analysis of the impact of health beliefs and resource factors on preventive behaviors against the COVID-19 pandemic." International journal of environmental research and public health 17.22 (2020): 8666.

[10] Aymerich, Claudia, et al. "COVID-19 pandemic effects on health worker’s mental health: Systematic review and meta-analysis." European Psychiatry 65.1 (2022): e10.