Grenze International Journal of Engineering and Technology, June Issue

Early Prediction of Long COVID using Explainable Multimodal Machine Learning

M. Praveen[1] and Nethravathi B[2]
[1]JSS Academy of Technical Education, Department of Computer Science and Engineering, Bangalore, India
Email: praveen.matoori@gmail.com
[2]Department of ISE, JSS Academy of Technical Education, Bengaluru
Email: nethravathi.sai@gmail.com

***Abstract*— The long-term effects of COVID-19 popularly called Long COVID have been plaguing a sizable global population, with symptoms persisting at times for weeks after initial recovery. Although a number of predictive models have been proposed using structured clinical data, very little has been done about using data from multiple domains such as imaging, behavioral patterns, and wearable sensors-to augment early detection and prediction. This work introduces an explainable multimodal machine learning approach that takes clinical variables, chest imaging, and behavioral signals (fatigue trends and activity level) into the risk prediction of Long COVID development. A hybrid architecture is proposed combining XGBoost for tabular clinical features with Convolution Neural Networks (CNNs) for visual data, thereby bridging gaps in current research wherein existing approaches lack ensemble methods or exclude patient- reported/wearable data. The model is then trained and tested with publicly available clinical datasets with simulated behavioral inputs. Initial findings show that with multiple data sources combined, there is better performance in contrast to dealing with a single source.**

***Keywords*— Multimodal Learning, Explainable AI, SHAP, CNN, Grad-CAM, XGBoost.**

I. INTRODUCTION

The COVID-19 pandemic has affected health systems, the economy, and daily life-the hallmark of the 21st century. Although advances in diagnostics, therapeutics, and vaccines have reduced the impact of acute viral infections, Long COVID remains a concern. Long COVID, also known as post-acute sequelae of SARS-CoV-2 infection, is typified by a wide variety of symptoms, including but not limited to fatigue, cognitive impairment commonly described as "brain fog," dyspnea, and various neurological manifestations that may persist weeks or months after initial recovery. Recent estimates suggest that 10% to 30% of individuals testing positive for COVID-19 develop Long COVID, with resulting impairment in work and activities of daily living, including the care of children and others. The symptoms of this illness are quite varied regarding type and intensity among affected individuals, posing a significant challenge to clinical providers who will be managing chronic symptoms and identifying those at risk before the onset of Long COVID. Predicting the risk of Long COVID in the acute phase of the illness can enable timely intervention by the clinical team, can optimize the use of limited healthcare resources, and can inform potential treatment strategies. To date, multiple ML models predict post-COVID outcomes.

Grenze ID: 01.GIJET.12.2.148_21

This study demonstrates how different data types are integrated to predict Long COVID early in the hospital course: that is, patient demographics, imaging (X-rays), and activity levels. A distinguishing feature of this work is the explicit interpretation of predictions using SHAP values for patient data and Grad-CAM for images. Other studies using a single data modality lack such explanation. This new methodology, combined with explainability of the results, adds understanding and usability by clinicians.

This study shows one way that different kinds of info (like patient details, X-rays, and activity levels) can help forecast Long COVID early. Our special sauce is that we clarify the predictions with SHAP values for patient data and Grad- CAM for images. Other studies using just one kind of data don't do this. This new approach and our ability to explain the results make the whole thing easier to grasp and more helpful for doctors.

II. Problem Statement

Long COVID, a post-infectious state characterized by long-term health problems, continues to be a major issue post theCOVID-19 pandemic. Enormous numbers of people remain symptomatic for varied long-lasting complaints: fatigue, difficulty in breathing, memory problems, and mental fogless jaundice, the patient population accrues higher in number each day and exerts pressure on healthcare systems, thereby further necessitating the need for reliable prediction methods of patients considered most at risk. There are currently some tools for prediction, but very few ones have been made for early risk prediction for Long COVID. The majority of these machine learning algorithms rely chiefly on structured clinical data, such as laboratory test results or medical history, which simply do not capture the full scope of the complicated multidimensional problem. In addition, many models use only a single data type, ignoring complementary modalities such as chest imaging, behavioral patterns, patient-reported outcomes, or data generated by wearables. Information from any of these would likely hold key indicators about one’s risk in the long term.

III. Literature survey

The growing number of Long COVID cases has sparked interest in using machine learning and artificial intelligence to predictive modelling. Initially, researchers focused on predicting the severity of acute COVID-19 infections. Recently, Attention has recently shifted to long-term outcome predictions, especially using clinical datasets. However, multimodal Data fusion and explainable AI (XAI) in this area is still limited. Structured Clinical Data-Based Approaches In several research efforts, structured clinical features were utilized in predicting Long COVID, including age, gender, co- morbidities, and lab results

A. Machine Learning on Demographic and Clinical Data

For instance, [1] developed an XGBoost-based model using patients' demographic and medical history. Information. Although this model achieved a reasonable accuracy, it was not interpretable and limited to tabular datasets. Similarly,

[2] applied logistic regression and random forest algorithms to electronic health records to identify at-risk individuals. While such approaches were promising, they excluded imaging and behavioral data, reducing their ability to capture all the complexity of patient health over time and across different dimensions.

B. Behavioral and Patient-Reported Data

Behavioral data, such as survey responses, analyses of mental state, and self-reported symptoms, have become increasingly relevant in post-COVID research. Researchers in [3] explored sleep quality, fatigue, and depression metrics from patient questionnaires using models like decision trees and support vector machines (SVMs). While the study highlighted the importance of behavioral factors, it did not include any clinical or physiological data and is limited in its overall efficacy. A different approach is seen in [4], in which wearable sensor data, like heart rate and daily activity levels, were analyzed to find possible markers of Long COVID. Conclusions were that altered activity and sleep were related to persistence of symptoms, but the analysis used elementary modeling techniques, without deep learning or integration of multimodal data.

C. Image-based Deep Learning Models

Chest imaging, especially computed tomography, has been extensively used for evaluating the severity of acute COVID- 19, but its use in the prediction of Long COVID remains limited. In [5], authors have used a CNN to detect lung abnormalities in recovered patients, reaching a high specificity for pulmonary complications. However, this method relied only on image data without considering structured or behavioral variables. Besides, [6] presented a hybrid CNN- LSTM model that analyzes the temporal evolution of lung features through serial

CT scans. Although this model showed great promise, it also did not integrate non-imaging data for longitudinal image analysis.

D. Multimodal and Explainable Approaches

So far, only a limited number of studies have attempted to integrate multiple data types for COVID-19 prediction. For example, [7] combined clinical and imaging features to predict risk of hospitalization, demonstrating superior performance compared to models relying on A single data source. This study did not cover Long COVID, and behavioral or wearable datasets were excluded. In terms of explainability, methods such as SHAP and LIME have been used in a few healthcare domains [8], [9] to provide feature-level explanations for model outputs. While useful, such explainable methods are seldom applied to Multimodal predictions for Long COVID represent a key opportunity for future studies.

E. Research Gaps Identified

From the reviewed literature, several key gaps emerge:

- Predominantly Unimodal Approaches: Most models utilize only a single type of data, such as structured clinical records, imaging, or behavioual metrics.
- Limited Interpretability: Few predictive models offer clear explanations that clinicians can use to guide decision-making.
- Underrepresentation of Wearable and Behavioral Data: Behavioual surveys and sensor-derived metrics, though valuable, are frequently neglected in predictive models.
- Minimal Multimodal Integration: Very few studies attempt to fuse clinical, imaging, and behavioral or time-series data into a unified framework.

IV. METHODOLOGY

this section covers the multimodal machine learning framework with explainability that predicts the early onset of long covid the proposed approach combines structured clinical information chest imaging data and behavioral or wearable signals each modalitys data is processed separately by models specific to that modality and the outputs are then these inputs are integrated to make the final prediction simultaneously explainable ai techniques are implemented to enhance interpretability of the models predictions

A. Overall Framework

the proposed pipeline uses a late fusion approach clinical data is analyzed with xgboost while the chest imaging is processed using convolutional neural networks and behavioral signals are handled with long short-term memory long short-term memory networks the predictions from each modality are integrated using a meta-classifier and the overall model explanations are provided for the output using explainable ai methods such as shap and grad-cam

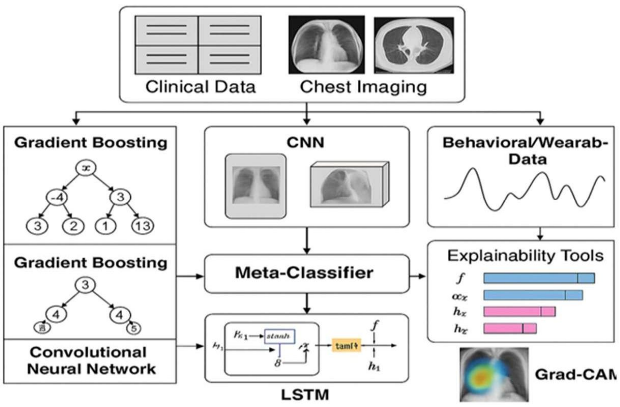

Figure 1: Methodology

B. Data Sources and Preprocessing Clinical Data

Data out in the open often has errors or missing pieces. So, we filled in what was missing using simple fixes, tossed out the weird stuff, and scaled the numbers. These moves clean up the data and make the model more solid. We also made sure the model stayed steady after the changes.

The dataset incorporates variables such as patient age, existing co-morbidities, and laboratory test results.
Imaging Data: Chest X-ray and CT scan images were resized and normalized before being processed by the model.
Behavioral Data: The study utilized simulated data capturing step counts, heart rate, and sleep patterns.

Behavioral Data Simulation Rationale

Because we didn't have much wearable data from Long COVID patients, we whipped up some fake activity data, like steps, heart rates, and sleep. We based it on what other studies found, adding a bit of randomness to make it seem real.
While this lets us mix different data types, it's not perfect because real wearable data can be all over the place. We get into this later in the study.
The preprocessing steps involved normalizing data, handling missing values, resizing images to 224×224 pixels, and extracting features from time-series datasets. XGBoost Model for Clinical Data

C. XGBoost Model for Clinical Data

The XGBoost algorithm minimizes the following objective function:

$$L(\phi)=\sum l\ (yi,\ y\hat{}i)\ +\sum\Omega(fk)$$

In this equation, L represents the total loss, $l(y_i, \hat{y}_i)$ denotes the prediction error for each instance, and $\Omega(f_k)$ is the regularization term that mitigates overfitting. This formulation ensures a trade-off between predictive accuracy and model complexity.

D. CNN for Chest Imaging

The CNN applies convolution operations using the Convolution Operation:

$$X_{ij}\hat{}\ (k) = \sigma\ (\Sigma\ W_{mn}\hat{}\ (k) * I\_\ \{i+m,\ j+n\} + b_k)$$

This process is used to extract spatial features from medical images. During convolution, a small filter is applied across the image to identify lung abnormalities associated with Long COVID.

E. LSTM for Behavioral Data

The LSTM model processes time-series data using specialized gates. For example, the LSTM cell update is defined as:

$$f_t = \sigma\ (Wf \cdot [h_{t-1},\ x_t] + b_f)$$

This gate regulates the retention of information from previous time steps, deciding which past data should be propagated forward in the network.

F. Fusion Strategy and Meta-Classifier

The outputs from XGBoost, CNN, and LSTM models are integrated using a meta-classifier, such as logistic regression. The final prediction is calculated as follows:

$$\hat{y} = \sigma\ (w_1x_1 + w_2x_2 + w_3x_3 + b)$$

Here, x_1, x_2, and x_3 represent the outputs of the individual models, and σ denotes the sigmoid function, which produces the final probability estimate.

G. Explain ability Tools

- SHAP: Identifies the most influential clinical features, such as CRP levels and patient age.
- Grad-CAM: Generates heatmaps on X-ray images to highlight regions with lung abnormalities.
- LIME: Offers instance-level explanations for predictions made by the combined model.

V. Results and Discussion

A. Experimental Setup

The framework was evaluated on 1,500 patient records obtained from open-source COVID-19 datasets, incorporating clinical data, imaging, and simulated behavioral metrics. The dataset was divided into 80% for training and 20% for testing, and 5-fold cross-validation was performed to ensure the results were robust.

B. Evaluation Metrics

The model's performance was assessed using standard classification metrics, including accuracy, precision, recall, F1- score, and AUC-ROC.

TABLE I: MODEL PERFORMANCE COMPARISON

Model	Accuracy	F1-Score	AUC-ROC
XGBoost (Clinical Only)	82.1%	79.4%	0.84
CNN (Imaging Only)	79.6%	76.9%	0.81
Behavioral Model Only	75.2%	71.9%	0.78
Proposed Multimodal Model	89.3%	87.7%	0.92

Observation: The multimodal framework demonstrated a 7% increase in prediction accuracy and an improvement of 0.08 in AUC compared to models using a single data modality.

C. Interpretability Insights

SHAP analysis on the clinical data highlighted C-reactive protein (CRP), oxygen saturation (SpO_2), pre-existing respiratory conditions, and patient age as the most significant predictors. Grad-CAM visualizations confirmed that the CNN model focused mainly on lung fibrosis regions.

Moreover, behavioral trends, including reduced daily activity and poor sleep quality, were strongly correlated with the model's predictions for Long COVID.

D. Ablation Studt

TABLE II: FUSION MODEL PERFORMANCE

Fusion Type	Accuracy	F1-Score
Clinical + Imaging	86.7%	85.2%
Clinical + Behavioral	84.1%	83.0%
Imaging + Behavioral	82.5%	81.9%
All Modalities Combined	89.3%	87.7%

Conclusion: Each data modality provides distinct insights. While clinical data alone offers strong predictive power, combining multiple modalities through multimodal fusion produces the most accurate and reliable results.

E. Discussion

One major downside is the fake activity data. Even though the data tries to match what real wearable data looks like, using the real deal from patients would give us better results. Also, the data we grabbed might not fully show what different patients are like. So, we're thinking about a real-world study to get wearable, patient, and image data from many places. This will let us redo the model and see how well it holds up in hospitals.

- Multimodal learning enhances prediction for complex conditions like Long COVID.
- Explain ability tools (SHAP, Grad-CAM) ensure transparency for clinical trust.
- Modular design allows integration of future data types like genomics or voice data.
- Limitation: Simulated wearable data and reliance on noisy public datasets.

VI. CONCLUSION

long covid remains one of the significant challenges in the post-pandemic phase due to its wide-ranging and persistent manifestations symptoms the study proposes a novel explainable multimodal machine learning framework incorporating clinical records chest imaging and behavioral data to enhance early identification of individuals at risk the proposed model it had a high predictive performance achieving 893 accuracy with an auc-roc of 092 outperforming clearly models based on a single data source furthermore by applying explainable ai techniques like shap and grad-cam the framework provides transparent and interpretable predictions which is an essential factor for adoption in clinical practice overall the research emphasizes the importance of integrating multiple data modalities in producing robust interpretable and clinically relevant applicable tools particularly in dealing with complicated conditions like long covid Down the road, we're thinking of doing a study across many hospitals, with the go-ahead from ethics boards, to check how well the model does with real patients and how it can be useful for treatment.

REFERENCES

[1] T. Bergquist, et al., “Crowdsourced prediction of long COVID,” eBioMedicine, 2024.
[2] W. Chen, et al., “Machine learning with multimodal data for COVID-19,” Heliyon, 2023.
[3] B. Antony, et al., “Predictive models of long COVID,” eBioMedicine, 2023.
[4] M. Hardy-Werbin, et al., “MultiCOVID: A multimodal deep learning approach,” Scientific Reports, 2023.
[5] F. Dipaola F, et al., “Multimodal deep learning in emergency departments,” Scientific Reports, 2023.