

MML: Multimodal Machine Learning Strategies for the Prediction of Covid-19 Status and Spread

Adebayo Nafisat Yemisi¹, Abiodun Esther Omolara¹, Olumide Owolabi¹, Oludare Isaac Abiodun¹

¹Department of Computer Science, University of Abuja, Gwagwalada, Nigeria
Corresponding author: Adebayo Nafisat Yemisi

Abstract

Few years ago, the world was thrown into a sudden chaotic loop due to the emergence of COVID-19 virus. The virus rapidly spread across environments, villages, towns, cities, regions, countries and continents. Consequently, it culminated into a huge strain on healthcare systems, thereby creating a disproportionate sudden and poor response with respect to medical services. Years down the line, the pandemic continues to pose a significant threat to global health, social interactions, and the economy. Despite ongoing efforts, completely containing the spread of the virus has remained difficult due to the elusive nature of its origins, or probably given that it is a new outbreak of its kind. Indeed, it is perplexing to experts and researchers in the field on how to totally address the challenge. Predicting the status and spread of the COVID-19 virus is still difficult because of its ever-changing nature, including the variability in its occurrence and symptoms in humans, potentially related to its cause and origin. But now that machine learning techniques are performing well in prediction of unknown event. Therefore, this study employed multimodal machine learning (MML) strategies, utilizing dataset from the World Health Organization (WHO), National Centre for Disease Control (NCDC), and John Hopkins University. The data was preprocessed using one-hot encoding, mutual information, and robust statistical methods, with L2 regularization added to the neural network. The results showed that the neural network (NN) predicted global COVID-19 status with 96.2% accuracy and feature importance analysis with 96.4% accuracy. The decision tree model predicted COVID-19 status in Nigeria with 0.93 accuracy and an F1 score of 0.98, identifying Lagos State as the most at risk, with 104,285 incident cases and 1,142 cases treated. The study concludes that MML can reliably predict COVID-19 status and spread factors, with NN models showing high accuracy. The study recommends the use of neural networks for COVID-19 spread analysis but advises caution in long-term predictions.

Keywords:

Covid-19, multimodal analysis of Covid-19, neural network, machine learning models.

1. Introduction

Nearly four years since the identification of the SARS-CoV-2 virus, the high rise in death cases caused by the COVID-19 virus is still a serious public health concern that has yet to be resolved. Unfortunately, little is known about how the SARS-CoV-2 outbreak spread (Nesteruk, 2024; Rasmussen, & Popescu, 2021). Hence, it is crucial to continually carry out a status investigation of COVID-19 across countries and also investigate the concerns about its spread and transmission pattern to effectively combat the spread of the virus (Sharma et al., 2021). It remains a challenging task to predict the status and spread of the COVID-19 virus due to its dynamic characteristics, including the variability in its occurrence and symptoms in humans, which may be linked to its cause and origin. Since various research outcomes have shown that machine learning (ML) techniques are performing well in prediction of unknown events (Shyalika et al., 2024; Shamout et al., 2020; Senders et al., 2018; Ambale-Venkatesh et al., 2017). Therefore, this study seeks to detect the status and spread of the COVID-19 virus using multimodal machine learning (MML). It uses the symptoms of infectious disease to predict the status and spread.

Looking back to the past four years, all over the world, the coronavirus disease has become massively widespread, causing devastation in every segment of human life, especially movement and social activity disruptions (Addo et al., 2024; Raza et al., 2023; López-Vargas et al., 2021; Ibn-Mohammed et al., 2021; Tadesse, & Muluye, 2020). The lockdown was imposed virtually everywhere, making it difficult for meaningful operation (Kaur et al., 2022; Onyeaka et al., 2021; Gonzalez, & Marlovits, 2020). Reports from the World Health Organization (WHO., 2023) showed that a quarter of a billion people suffer from COVID-19, caused by the deadly coronavirus, which stands for Coronavirus Disease (SARS-CoV-2). The disease was rapidly spread, and more than three million positive cases were globally recorded during its ravaging era (Andrade et al., 2022). It was a dangerous period for global society as all sectors of the economy were affected (Nayak et al., 2022), such as health, transportation, business, education, energy, agriculture, communication, commerce, etc.

The devastation of COVID-19 is a phenomenon of immense size and scope, with developed Western countries bearing the brunt of the impact even more (Ishola et al., 2021). While Africa may not have experienced a severe COVID-19 outbreak, it still faced challenges in the spheres of health, social life, and the economy. Recently, machine learning (ML) techniques have been utilized to forecast the spread of COVID-19 (Heidari et al., 2022; Reddy et al., 2021; Ardabili et al., 2020). Predictive models can analyze data and recognize patterns to predict future outcomes (Yang, 2022). These models use both structured and unstructured data, allowing them to evaluate various attributes and assist in informed decision-making, such as identifying disease symptoms. In order to forecast the spread of COVID-19, ML methods often incorporate data sequences collected over time as inputs (Wong et al., 2019).

COVID-19 modeling is a crucial tool for forecasting the future trajectory of the pandemic, and artificial intelligence methods play a significant role in this process (Comito, & Pizzuti, 2022). The results of these models are of utmost importance to policymakers in managing the disease, as they can predict the future spread of the virus (Akhtar et al., 2020). These models can analyze the parameters that impact the spread of COVID-19 in different regions and populations, evaluate different intervention strategies, and compare COVID-19 data with similar diseases. In the short term, most models have been effective in predicting COVID-19 parameters, but their long-term accuracy needs to be validated in further studies (Ting et al., 2020).

Prophet and long short-term memory (LSTM) models have shown good results in predicting the spread of COVID-19 (Long et al., 2023; Sah et al., 2022; Devaraj et al., 2021). But with autoregressive integrated moving average (ARIMA) being the most commonly used method for time series predictions (Yang et al., 2020). However, these models have not been applied and trained on the same data, making comparisons between them difficult. The application of artificial intelligence (AI) strategies has shown promise in predicting the spread of COVID-19. However, as it is a new and rarely known or unknown disease, larger and more collaborative research are needed to test these models to make better predictions (Shahzad et al., 2020) of how the disease can be prevented or eliminated. Despite the various forecasting methods for infectious diseases, there is currently no method that can be utilized for short-term, day-to-day predictions. During the COVID-19 pandemic, more than one thousand articles were published on "Susceptible-Exposed-Infectious-Removed" (SEIR) epidemic computational models.

SEIR models with a stochastic solver may account for changes in the transmission rate over time, but they can still fall short in predicting the impact of factors such as mask mandates and contact tracing initiatives (Pan et al., 2020). Other studies have employed a variety of techniques, including an adaptive neuro-fuzzy inference system, bidirectional long short-term memory, moving average, and multilayer perceptron. That is,

to predict the number of new COVID-19 cases. Indeed, despite ongoing efforts, there is still a lack of a precise and rapid diagnostic method for COVID-19. A study by Akhtar et al., (2020) highlighted this challenge in their research, indicating that the development of an accurate and efficient diagnostic approach for COVID-19 has been elusive.

The reliability of COVID-19 status predictions is also hindered by the need for continuous updates to the datasets used (Ergen, 2022). Data collection for COVID-19 continue to grow, but current studies have struggled with the complex nature of data collection and the need for continuous predictions (Yenurkar et al., 2022). This study aims to address these issues by proposing a two-way prediction of COVID-19 status, including the area's most likely to be affected. Therefore, to determine the COVID-19 status and underlie possible factors responsible for the spread, this study employs a multimodal analysis approach using the deep learning method. The novelty and contributions of our research are summarized as follows:

1. As the global COVID-19 pandemic persists, precise prediction of the virus's transmission is essential for efficient public health management, optimal allocation of healthcare resources, and monitoring of the consequences of government actions. Given the recent widespread occurrence of SARS-CoV-2 outbreaks worldwide, it is crucial to have a data-driven understanding of the specific areas, timing, and demographics impacted by the virus. This major contribution highlights the novelty of our study, which is broadly lacking in the literature.
2. Various approaches, including ARIMA and LSTM models, have demonstrated promising performance. However, to ensure their long-term effectiveness, additional validation is necessary. Our proposed MML incorporates various methods that have been rigorously tested on large and diverse populations. This has resulted in an optimized and highly accurate predictive model.
3. Through a thorough examination of the literature, we undertook an in-depth analysis to reveal the limitations of existing strategies.
4. Our approach to using MML strategies emphasizes the viability of accurately predicting the timing of an outbreak using a clear and understandable approach. This could potentially establish standards for accurately forecasting the onset of fresh outbreaks and identifying forthcoming patterns of transmission.
5. The novel MML has the potential to make major contributions to the development of intervention and preventative techniques that are specifically targeted.

2. Literature Review

Multimodal machine learning (MML) has emerged as a robust approach to analyzing and integrating heterogeneous data sources (Al-Zoghby et al., 2024). The prediction of Covid-19 status and spread presents a significant challenge, requiring insights from multiple data modalities such as clinical records,

imaging, genomic sequences, social media activity, and mobility patterns. This literature review examines foundational and contemporary research on the application of MML strategies to Covid-19, emphasizing predictive modeling, data fusion techniques, and practical outcomes (Shiwlani et al., 2024). Next-generation sequencing (NGS) data provides crucial insights into viral variants (Rahimian, & Panahi, 2024). Long short-term memory (LSTM) networks and transformers have been applied to identify genomic mutations linked to transmissibility and severity.

Studies leveraging Google mobility reports and smartphone tracking data have correlated movement patterns with regional outbreak trends. Recurrent neural networks (RNNs) and temporal convolutional networks (TCNs) are commonly used (Dudukcu et al., 2023). Sentiment analysis and topic modeling of platforms like Twitter reveal public awareness and behavioral changes, serving as proxies for compliance with health measures.

Studies have shown that integrating clinical features with imaging data significantly enhances Covid-19 detection accuracy (Ahsan et al., 2021). Deep learning method have been applied to detect related symptom of infectious disease (Shorten et al., 2024). Pretrained convolutional neural networks (CNNs) like ResNet, DenseNet, and EfficientNet have been adapted for Covid-19 detection using CXR and CT scans (Ebenezer et al., 2022). These models achieve high sensitivity and specificity in detecting pneumonia caused by Covid-19. Hybrid models works by researchers (e.g., Sansana et al., 2021; Hajirahimi, & Khashei, 2019) have incorporated imaging data with tabular clinical data (e.g., oxygen levels, age, and symptoms) to create hybrid models, improving diagnostic performance.

MML frameworks have been integrated with compartmental models (e.g., SIR, SEIR) to enhance spread prediction accuracy (Machuca, 2020). The fusion of epidemiological parameters with real-time multimodal data allows for dynamic forecasting (Fei et al., 2024). The application of multimodal machine learning strategies for predicting Covid-19 status and spread has demonstrated remarkable potential, blending the strengths of diverse data types. While challenges remain, advancements in computational techniques, data integration (Lenzerini, 2002), and interpretability will further empower MML systems, contributing to improved pandemic management and preparedness (Godinho, 2024; Gayathiri et al., 2023).

MML strategies for predicting Covid-19 status and spread often leverage specific models tailored to handle the unique properties of different data modalities (Liu et al., 2024). Take for example, clinical, imaging, genomic, temporal). Some popular models and their associated hyper-parameters include CNNs which are commonly used for processing medical imaging data, such as chest X-rays or CT scans. Relevant hyper-parameters include kernel size e.g., filters used for feature ex

traction (e.g., $3 \times 3, 5 \times 5 \times 3, 5 \times 5 \times 3, 5 \times 5$).

RNNs and their variants like LSTMs and GRUs are ideal for time-series data, such as case trends or mobility patterns (ArunKumar et al., 2022). There are also, transformer-based models for multimodal fusion (Roy et al., 2023). That is transformers use self-attention mechanisms to handle multiple modalities simultaneously. Their hyper-parameters are embedding size, for instance dimensionality of token representations values is 128, 256. Number of attention heads that determines parallel attention mechanisms, typical values maybe 4, 8.

Despite COVID-19 was initially discovered in a city called Wuhan in China (Mohan, & Nambiar, 2020). After being discovered in Wuhan, it spread quickly, resulting in millions of deaths and even hundreds of infected cases. As of January 19, 2023, the World Health Organization (WHO, 2023) reported cases of people confirmed to have been infected in hundreds of millions worldwide. So far, over 13 billion vaccine doses have been administered globally as of January 17, 2023. The global statistics for the world data is illustrates in Fig. 1 and the situation in Nigeria portrays in Fig. 2.

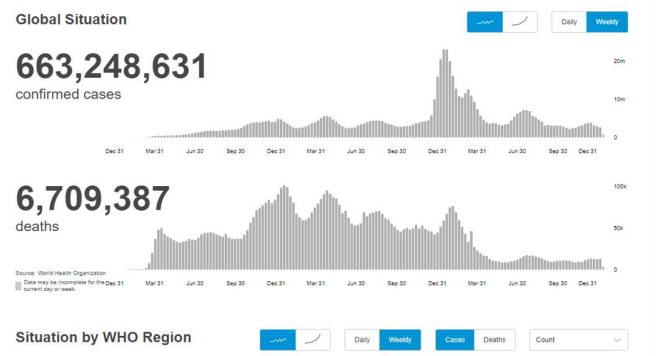


Figure. 1. Global Situation of COVID-19 (Source: WHO, 2023)

Figure. 2 represent the trained SAE that reconstructs the original input image from the noisy image version. SAE extracts only the features of an image and produces the output by eliminating any disturbance or unnecessary noise in the system. Let examine the situation in Nigeria in Fig. 2.

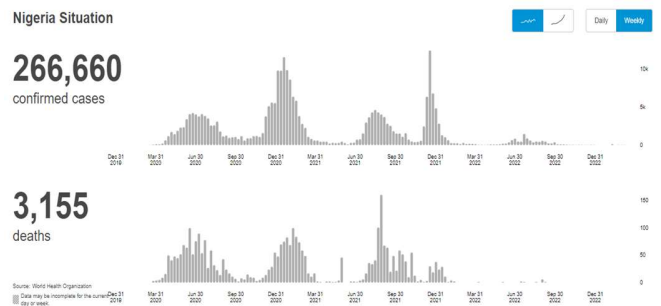


Fig. 2. Global Situation of COVID-19 (Source: WHO, 2023)

Fig. 2 revealed that the aging population is more affected by the virus than the younger ones. This is because of immunity differences due to age, genetic, energy possession and body resistance (Bashir et al., 2020). According to (Ficetola et al., 2020), weather changes have a significant correlation with changes in pneumonia mortality rates. Long-term climate change and its associated extreme weather also played a role. As such, climatic conditions play a key role in COVID-19 spread.

Arora et al., (2020) utilized LSTM variants related to RNNs to predict the number of positive COVID-19 cases on an Indian dataset. The LSTM model achieved the lowest error rate and was used to forecast daily and weekly new COVID-19 cases with error rates of approximately 3% and 8%, respectively. They then employed the model to categorize Indian states into different zones based on the extent of positive cases and daily escalation to identify COVID-19 hotspots. Fokas et al., (2020) utilized a bidirectional LSTM network to provide a robust generalization of RNNs, predicting new COVID-19 cases in Italy, Spain, France, Germany, the USA, and Sweden.

The Convolutional Neural Network (CNN) is a highly regarded deep learning model for making predictions about

COVID-19. Deep learning is employed in self-driving cars to comprehend their surroundings by analyzing the unstructured data captured by cameras and making quick decisions. In the medical field, radiologists utilize deep learning to examine medical images. SVMs are a machine learning tool utilized for binary classification or regression analysis (Cristianini et al., 2000). In binary classification, instances of one class are separated from the other using a linear hyperplane (decision boundary) that maximizes the distance between the two classes (Ławrynowicz et al., 2014).

The Automatic Regressive Integrated Moving Average (ARIMA) is a widely used approach for analyzing and forecasting time-series data. Alzahrani et al., (2020) also used the ARIMA model to forecast daily COVID-19 cases in India, and the most suitable model was chosen based on the Bayesian information criteria and the highest R2 value. ARIMA was found to be the most effective model for predicting COVID-19 cases due to the unpredictable and seasonal nature of the confirmed cases, and the daily addition of new cases. Other notable works in the literature are provided in Table I.

TABLE I THE SUMMARY OF THE EMPIRICAL LITERATURE REVIEW OF RELATED WORKS

S/n	Author & Year	Title	Outcome	Limitations
1	Xiong et al., (2022).	Comparing different machine learning techniques for predicting COVID-19 severity. Infectious diseases of poverty.	Random Forest was able to identify patients rapidly	The Method only works on a small dataset
2	Sinha et al., (2022).	Analysis and prediction of COVID-19 confirmed cases using deep learning models: a comparative study.	Captures long-term dependencies in sequential data	Takes a Long Term to train, and the potential gradient vanishing problem
3	Kuo et al., (2021).	Evaluating the impact of mobility on COVID-19 pandemic with machine learning hybrid predictions.	Predictive Weekly Cyclicity	Worked optimally for selected countries
4	Tamang et al., (2020).	Forecasting of COVID-19 cases based on prediction using artificial neural network curve fitting technique.	Reduces overfitting and a high degree of general ability	Not being Tested with data of varying time complexity, yet consumes time
5	Saba et al., (2020).	Analysis and prediction of COVID-19 confirmed cases using deep learning models: A comparative study.	Has high determination of coefficient	The model was confined to Egypt
6	Al-Najjar et al., (2020).	A classifier prediction model to predict the status of Coronavirus COVID-19 patients in South Korea.	Accurately predict recovered and deceased cases	The model was confined to South Korea
7	Hafeez et al., (2020).	A CNN-based coronavirus disease prediction system for chest X-rays.	The model outperforms many cutting-edge classification models	Worked optimally for exclusive datasets
8	Humayu & Alsayat., (2022).	Prediction Model for Coronavirus Pandemic Using Deep Learning.	More accurate than current models	Focuses on using X-ray data
9	Li et al., (2022).	An integrated recurrent neural network and regression model with spatial and climatic couplings for vector-borne disease dynamics.	The model outperformed ARIMA models across several regions with high infections.	Need to be compared with similar Prophet model results.
10	Haritha et al., (2020).	Prediction of COVID-19 Cases Using CNN with X-rays.	The results obtained in COVID prediction had a training accuracy of 99% and a testing accuracy of 98.5%	The model dataset is limited to images from transfer learning
11	Shaikh et al., (2021).	Analysis and Prediction of COVID-19 using Regression Models and Time Series Forecasting.	The fully connected CNN with two variants of convolution especially DSC has proved its efficiency in detecting COVID-19 infections.	The issues related to data relevance are a shortcoming.

12	Houssein et al., (2022).	Hybrid quantum-classical convolutional neural network model for COVID-19 prediction using chest X-ray images.	The proposed HQ-CNN model can predict the positive COVID-19 cases.	The architecture of the HQ-CNN model has a few layers to match small-scale quantum hardware.
13	Babukarthik et al., (2020).	Prediction of COVID-19 using genetic deep learning convolutional neural network.	The proposed method performs better compared to other transfer learning techniques. Classification accuracy of 98.84%.	only suitable for earlier stages of diagnosis
14	Hamadneh et al., (2021).	Artificial neural networks for prediction of COVID-19 in Saudi Arabia.	The results reveal that infected persons will increase in the coming days and reach a minimum of 9789.	Concentrated only on recovered persons
15	Bhoumik et al., (2020).	Covid-19 Prediction from X-Ray Images Using Fully Connected Convolutional Neural Network.	The fully connected CNN with two variants of convolution especially proved its efficiency in detecting COVID-19 infections.	The need for a more comprehensive assessment of the model's predictive with regression analysis
16	Malki et al., (2021).	The COVID-19 pandemic: prediction study based on machine learning models.	The study developed a machine learning model to predict the spread of COVID-19 in various countries, achieving a high overall R2 of 0.99 for confirmed cases.	The study's prediction of COVID-19 infections declining and ending by the first week of September 2021 may be subject to uncertainties and may not fully account for various factors affecting the pandemic's trajectory.
17	Saadatmand et al., (2023).	Using machine learning in prediction of ICU admission, mortality, and length of stay in the early stage of admission of COVID-19 patients.	Machine learning models, particularly ensemble stacking with a Neural Net, achieved high accuracy rates, over 95%, in predicting ICU admission for COVID-19 patients, providing valuable support for hospital managers and practitioners.	While the models showed promising predictive capabilities, their effectiveness in real-world scenarios may be influenced by factors not captured in the data, potentially impacting their generalizability and reliability.
18	Hu et al., (2020).	Early prediction of mortality risk among patients with severe COVID-19, using machine learning.	A clinical model was developed using machine learning to predict severe COVID-19 patient outcomes early, utilizing demographic, clinical, and laboratory data. The final model, incorporating age, high-sensitivity C-reactive protein, lymphocyte count, and d-dimer levels, achieved high accuracy (AUROC: 0.881) in external validation.	Generalizability to diverse populations and healthcare settings may require further validation. Initial laboratory-based predictions may overlook COVID-19's dynamic progression and treatment response.
19	K-Abdul-Hamid et al., (2023).	Improvement of time forecasting models using machine learning for future pandemic applications based on COVID-19 data 2020–2022.	The study proposed a hybrid autoregressive integrated moving average-least-squares support vector machine model for more accurate and efficient prediction of COVID-19 cases, demonstrating superior performance in reducing error compared to other models.	While the proposed model shows promising results, its generalizability to different regions or future pandemics may require further validation, and potential biases in the datasets used could affect the model's effectiveness in real-world scenarios.
20	Anggrawan et al., (2023).	Comparative analysis of machine learning in predicting the treatment status of COVID-19 patients.	The study proposes a data mining classification method, employing Random Forest (RF) and Support Vector Machine (SVM) algorithms, to accurately predict the treatment status of COVID-19 patients, achieving up to 94% accuracy with RF and up to 92% accuracy with SVM.	While the machine learning models demonstrate high accuracy, their applicability in real-world medical decision-making may require further validation and consideration of factors beyond the scope of the study, such as patient-specific characteristics and treatment protocols.
21	Cho et al., (2023)	Detection of COVID-19 epidemic outbreak using machine learning.	The study successfully developed machine learning models, particularly Random Forest and XGBoost, that achieved high accuracy (over 94%) in predicting COVID-19 transmission trends and detecting the start of new outbreaks. This method, tested on data from Korea, provided valuable early warnings for public health interventions, potentially improving patient care and resource management.	The model's effectiveness is heavily dependent on data quality and may not generalize well to other regions without further validation. The approach also risks overemphasizing minor outbreaks, which could lead to unnecessary resource allocation, and requires ongoing retraining to adapt to the evolving nature of the pandemic.

Understanding the spread of COVID-19 alone is not sufficient for predicting its status (Chadaga et al., 2024). Previous studies have utilized a range of machine learning models, including SVM, RF, KNN, GB, LR, and NB, to address COVID-19-related issues, but these exercises were not aimed at enhancing status prediction. Deep learning models such as CNN, LSTM, RNN, Bidirectional-LSTM, and CNN-LSTM-

RNN hybrid models have proven to be highly effective, especially the CNN model, which has shown promising results with X-ray or CT-scan image datasets. However, there is limited research in this area, which makes it difficult to compare and benchmark results. This speaks to a research hole in the field under investigation. This study used multimodal analysis and combined models to provide comprehensive understanding

that may not be apparent when using a single model or method. Multimodal analysis is the study of communication and meaning-making through multiple modes or channels of information (Iedema, 2003). It examines how different forms of data or "modes" work together to convey information or create a cohesive understanding of a phenomenon. These modes can include text, visuals, audio, gestures, spatial design, or other forms of representation. Example of multimodal analysis in Covid-19 research include clinical data such as textual records of symptoms and lab tests. Epidemiological data such as mobility patterns or case counts. Image data like Chest X-rays or CT scans. Integration involves using all these modes in a machine learning model to predict infection status and spread.

3. METHOD FOR THE DEVELOPMENT OF THE MULTIMODAL MACHINE LEARNING (MML) STRATEGY

(a) Proposed system

Multimodal models process information from multiple data sources or modalities (e.g., text, images, audio, or numerical data) to produce predictions or insights. These models are used sequentially or independently, and their outputs are combined to produce a final prediction. In a sequential approach, different models process the data in a specific order, with each step informed by the output of the preceding one. This strategy is useful when there is a logical dependency between the modalities or when one modality refines the processing of another. In sequential use at initial processing, one modality is processed first to extract foundational insights. Take for instance, an NLP model processes clinical notes to extract symptoms and context.

However, in the secondary processing, the output of the first model guides the next stage of processing. Take for example, Symptoms extracted from text guide a CNN to focus on specific imaging features in X-rays. The outputs from all sequential

steps are combined to make the final prediction. In an independent approach, models for each modality process their respective data streams simultaneously, without interdependencies. This strategy is ideal when modalities are complementary but does not require interaction during processing. In independent use, parallel processing, each modality is processed independently using its specialized model. Take for example, a CNN for imaging, an LSTM for time-series data, and a transformer for clinical notes. Output from all models is fused to produce a unified prediction. That is, probabilities from individual models are weighted and combined. The advantages are Models can operate in parallel, reducing processing time. Scalability is improved since each model focuses on its own domain. Take for example, predicting Covid-19 spread, a CNN processes satellite imagery of urban mobility. An RNN models temporal case trends. A multimodal transformer integrates all outputs for forecasting.

In the case of the **output combination strategies**, the outputs from different models are combined through various strategies, depending on the task requirements and data compatibility. Outputs (e.g., embeddings) from different models are concatenated into a single feature vector, which is fed to a final prediction layer. Take for example, combining text, image, and numerical features into a single vector for classification. Take for instance, for ensemble learning, predictions from individual models are aggregated using ensemble methods like voting, bagging, or boosting. Example, voting between predictions from imaging, textual, and epidemiological models.

In using the multimodal analysis, this study used a deep sequential model as it would be appropriate when compared to a shallow model. Regular updating of datasets and experimentation with new methods and parameters are also important components of this approach to accurately predicting the COVID-19 status. The framework developed for this study is given in Fig. 3.

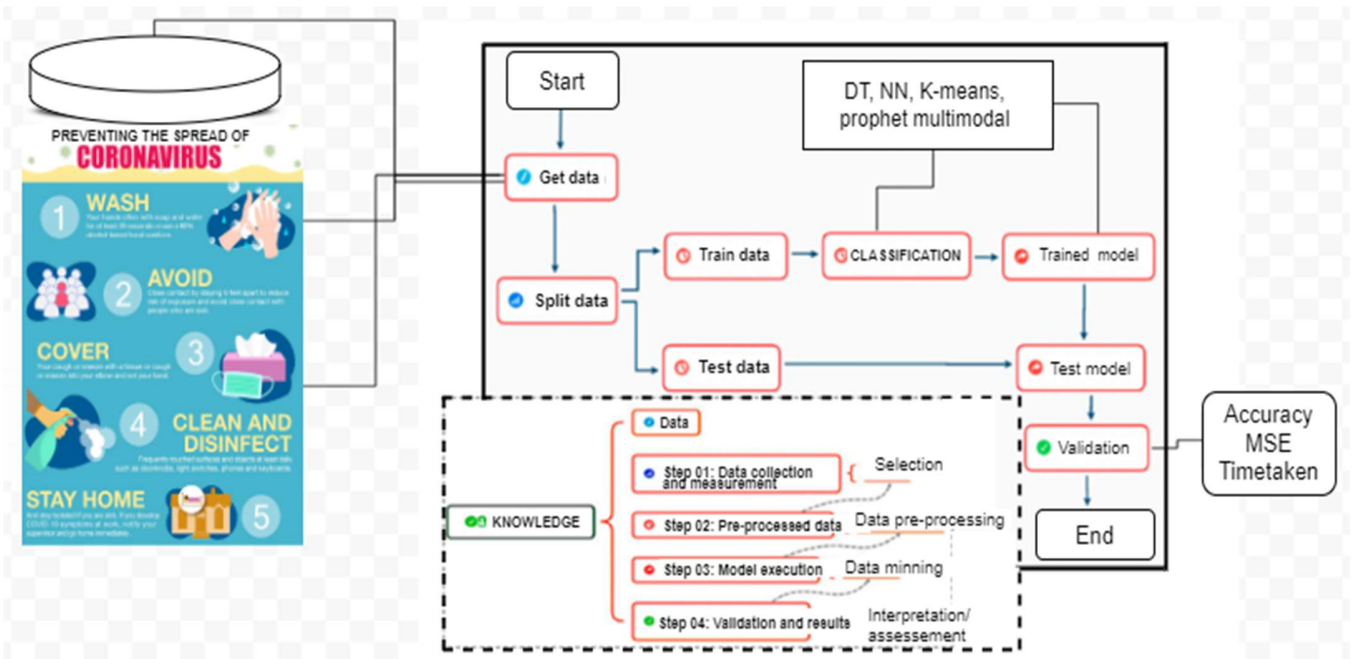


Fig. 3. Model framework

The datasets used in COVID-19 status prediction need to be updated with the latest data and continuously monitored. There is also a challenge in collecting data, which is often complicated and inconsistent. In this study, a two-way prediction of COVID-19 status was proposed. To build the machine learning (ML) model, the Scikit-Learn library and applied mutual information were used. Three classifiers were used for training and comparing the results to see how the MI score impacted the ML model's effectiveness. To make the public databases easier to use and understand, a smoothing average of the data was performed and adjustments were made in the event of a large data dump. Testing and validating the efficacy of the long-term performance of such models ensures that the purpose of the study is fulfilled.

(b) Dataset

Three datasets were employed for the study, one from Hopkins; and two each from WHO and NCDC. Hopkins data was used to determine the global status of the virus. The data is recent, as it is updated every day. The dataset from NCDC shows the situation in Nigeria. The row represents the states, and the values in each column correspond to that state. These data provide useful information for monitoring states and regions spread.

(c) Pre-processing

In preprocessing data for the NN, a one-hot encoding of categorical variables trained on the neural network model was used to prepare the data for the binary prediction of high and low-risk COVID-19 classification.

(d) Missing values

Missing values were handled. This involved dropping columns with NaN values and encoding categorical features,

and the mean value was employed to fill up spaces. Pre-processing of raw data was carried out to convert it into a format that is easy to understand and can be used by other processes. The data was processed to obtain a weighted risk for each country based on population density and the percentage of confirmed cases. K-Means clustering was used to classify each country into low, medium, or high-risk categories.

(e) Exploratory data analysis (EDA)

The data was explored by visualizing it in charts and graphs, which helped to highlight. Pandas, Numpy, Seaborn, and Matplotlib were used to create visualizations. Furthermore, a correlation matrix was carried out to display the correlation coefficients between the variables confirmed cases "New Cases", "Total Recovered", and "Deaths".

(f) Outlier handling

An outlier is identified by the color scheme of a scatter plot generated to display the associations. To handle the outlier, the analyst used a color-coded scatter plot to visualize the association between the number of confirmed cases and the daily count of new cases.

(g) Data balancing and overfitting

The challenge of unevenly distributed data, also known as the unbalanced data problem," can hinder the effectiveness of machine learning algorithms. Where an imbalanced dataset occurs, the study utilizes SMOTE. To solve overfitting in the NN model, the complexity of the model was decreased.

(h) Train test data

The study utilized the tenfold cross-validation method. This approach involved splitting the original samples into ten equal sub-samples, where one sub-sample was reserved for

validation and the other nine were for training. The results from each validation run were then averaged.

(i) Experimented algorithms

This study experimented with four different models - CNN, Prophet Algorithm, Decision Tree, and K means. The algorithms were used as learning parameters to determine the status and spread of the virus.

(j) Model development (NN)

An NN model was designed with different input parameters to make predictions about the disease by altering the design features. The model was designed to analyze the data with the highest frequency of information. The neural network model was trained and evaluated to predict whether the total cases in a country were high-risk or low-risk. To do this, a binary classification using a neural network was performed to predict whether a country's cumulative COVID-19 cases are high-risk or low risk based on various features, such as the country's population.

The input data, which was a sequence of time series observations, was fed into the neural network, with each row representing data from a specific date. Specifically, the input is a set of features that describe COVID-19 data for different countries. These features include the total number of cases, the total number of deaths, the country's population, and the date of reporting. The model was trained using the data on the number of cumulative cases in different countries and their corresponding features, such as population density, gross Domestic Product (GDP), and human development index.

The model consisted of two dense layers, with the first layer having 11 input neurons and the second layer having 1 output neuron. The model was assessed by specifying the loss function, optimization algorithm, and evaluation metric. The model was then trained on the training set using the fit() method, with the number of epochs and validation set specified. The two dense layers in the model were fully connected layers, with the first layer having 11 input neurons corresponding to the 11 input features in the dataset. An illustration of neural network model is presented in Fig. 4.

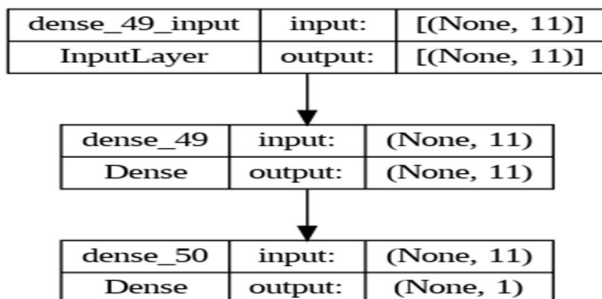


Fig. 4. Neural network model

The model's mean absolute error (MAE) loss function, Adam optimization algorithm, and MAE evaluation metric

were compiled. To prevent overfitting, the regularization (L2 regularization) technique was added to the neural network. To enhance the performance of the NN model, Bayesian optimization, an optimization technique, was utilized.

(k) Performance Metrics

The multimodal model was evaluated using parameters such as root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), accuracy, and precision. These parameters were used to assess the accuracy of the models. MAPE and (MSE) were used to measure the average difference between predicted and actual values for the Prophet Algorithm and NN model. These metrics handled the performance of the NN model on COVID-19 spread, where the neural network is used to predict regression problems of numerical values related to COVID-19 factors. The metrics False Negative (FN), True Positive (TP), False Positive (FP), and True Negative (TN) were used for the clustering model and decision tree models. A summary of classification metric is depicted in Fig. 5.

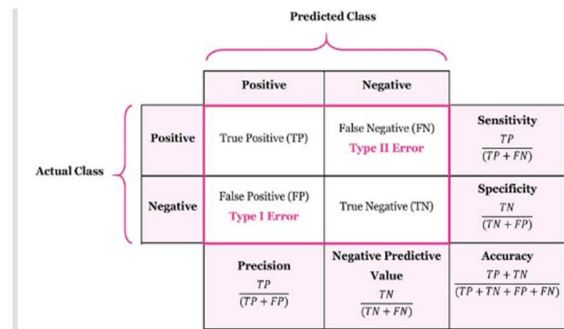


Fig. 5. A summary of classification metric

(l) Development Tools

The research employed Google Colab, which is a collaborative platform. It offers access to a powerful computing system. Researchers employed several key Python libraries, such as NumPy, Pandas, Seaborn, Scikit-learn, and Tensorflow. These libraries were utilized with machine-learning methods.

(m) Summary

The summary of methodology used in the study is presented in Table II.

TABLE II
A SUMMARY OF METHODOLOGY USED IN THE STUDY

Method	Techniques
Data Extraction	Online
Data Preprocessing	Label One Hot Encoding Eliminated ambiguous records using risk coefficient.
Exploratory Data Exploration	Matplotlib Seaborn
Feature Selection	Mutual Information

Data Modelling	Over Sampling Majority (SMOTE) K fold Cross-validation (70/30%) Grid SearchCv
Model Evaluation	F1 Score Accuracy Precision Recall RMSE MAPE
Development Tools	Google Colab, Python libraries,

4. ANALYSIS AND RESULT

The data was sourced from Johns Hopkins University. The analysis involved several stages, including data cleaning and preprocessing, data exploration to identify outliers and missing values, and the calculation of various statistical measures to describe the data. COVID-19 cases are discretized to make it easier for the models to converge. Population density values are made a multiple of 5 and handicaps are applied. Population density is classified as low/medium/high and the K-Means algorithm is applied to group them into three categories. Population risk is classified as low/medium/high. The data was processed to obtain a weighted risk for each country based on population density and the percentage of confirmed cases. K-Means clustering was used to classify each country into low, medium, or high-risk categories. Hence, to develop a predictive model, the EDA was performed on the COVID-19 status dataset to gain insights into the underlying patterns and trends.

Statistical plots were utilized to visualize the distribution of the data, which aided in obtaining a better understanding of the dataset. The distribution of new cases and new deaths across different countries was visualized using scatterplots and heat maps. The trend of cumulative cases and cumulative deaths over time was plotted using line graphs. A summary of the descriptive features in the dataset related to COVID-19 is shown in the country column features. Descriptive items are shown in Table III.

TABLE III
DESCRIPTIVE ITEMS

Feature	Description
Date_reported	The date when the cases/deaths were reported.
Country_code	The ISO 3166-1 alpha-2 code of the country or territory
Country	The name of the country or territory
WHO_region	The WHO region to which the country or territory belongs
New_cases	The number of new confirmed cases of COVID-19 reported on the date.
Cumulative_cases	The total number of confirmed cases of COVID-19 reported up to date.
New_deaths	The number of new deaths from COVID reported on the date

The dataset used for the analysis contained information on the outcomes of COVID-19 across different countries and territories worldwide as shown in Table III and Fig. 6(a)(b).

Country/R...	# Confirmed	# Deaths	# Recovered	# Active	# New cases	# New deaths
Afghanistan	36263	1269	25198	9796	106	10
Albania	4880	144	2745	1991	117	6
Algeria	27973	1163	18837	7973	616	8
Andorra	907	52	803	52	10	0
Angola	950	41	242	667	18	1
Antigua and Barbuda	86	3	65	18	4	0
Argentina	167416	3059	72575	91782	4890	120
Armenia	37390	711	26665	10014	73	6
Australia	15303	167	9311	5825	368	6
Austria	20558	713	18246	1599	86	1
Azerbaijan	30446	423	23242	6781	396	6
Bahamas	382	11	91	280	40	0
Bahrain	39482	141	36110	3231	351	1

Fig. 6(a). Data content

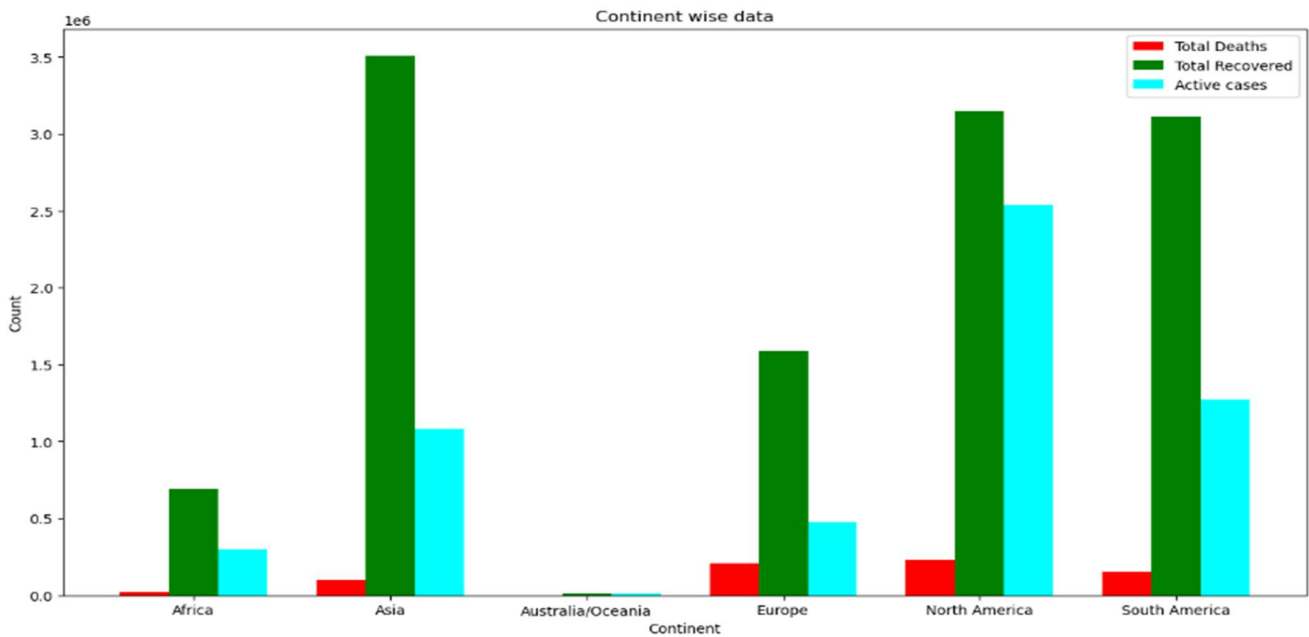


Fig. 6(b). Description of the data based on continent-wise characteristics

The analysis revealed that the pandemic has affected countries and regions differently, as shown in Fig. 6. Some areas are experiencing higher numbers of cases and deaths than others.

Variations exist over time across known cases with some periods experiencing a surge in infections and fatalities.

```

countries with high risk:
100 United States of America
101 United Kingdom
104 Turkey
105 Ukraine
106 South Africa
...
238 San Marino
245 Isle of Man
246 Guernsey +
247 Jersey +
251 Occupied Palestinian territory, including east...
Name: Country, Length: 71, dtype: object

Countries with low risk:
0 Afghanistan
1 Albania
2 Algeria
3 Andorra
4 Angola
...
236 Saint Vincent and the Grenadines
237 Samoa
239 Sao Tome
240 Serbia
250 Venezuela (Bolivarian Republic of) + + +
Name: Country, Length: 180, dtype: object

```

Fig. 7. Classification of countries by COVID-19 risk level

It can be revealed that the list of countries with high risk includes countries like the US, UK, and Ukraine. The list of countries with low risk includes Afghanistan, Albania, Algeria, Andorra, and others as shown in Fig. 7.

Fig. 8 shows the final preprocessed data used in the model. The preprocessing stage involved data cleaning, which involved inspecting the data to identify missing values, this involved dropping columns and rows with NaN values, encoding categorical features, and filling up empty spaces. The preprocessed dataset as shown in Fig. 8.

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 241 entries, 0 to 252
Data columns (total 14 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   location                                   241 non-null    int64
1   last_updated_date                         241 non-null    int64
2   new_cases_smoothed                       241 non-null    float64
3   new_deaths_smoothed                     241 non-null    float64
4   population_density                       241 non-null    float64
5   median_age                               241 non-null    float64
6   aged_65_older                           241 non-null    float64
7   aged_70_older                           241 non-null    float64
8   gdp_per_capita                           241 non-null    float64
9   cardiovasc_death_rate                   241 non-null    float64
10  diabetes_prevalence                     241 non-null    float64
11  hospital_beds_per_thousand              241 non-null    float64
12  life_expectancy                         241 non-null    float64
13  human_development_index                 241 non-null    float64
dtypes: float64(12), int64(2)
memory usage: 28.2 KB

```

Fig. 8. The preprocessed dataset

The factor data after preprocessing contains 14 columns. Each column likely contains the corresponding information for the specific location being studied.

(a) Factors responsible for the spread of COVID-19

The most important factors contributing to the model's predictions were identified by extracting the weights from the first layer of the model and calculating the absolute sum of the weights for each input feature. Result of the feature importance (factors that impact the spread of COVID-19) is shown in Fig. 9.

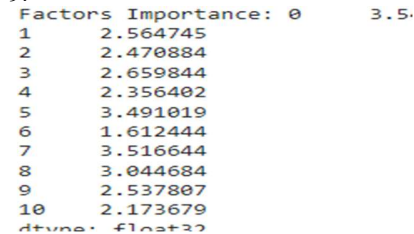


Fig. 9. Input weights result of the feature importance

The result of the feature importance shown in Figure 9 shows that age, location, and GDP have the highest importance

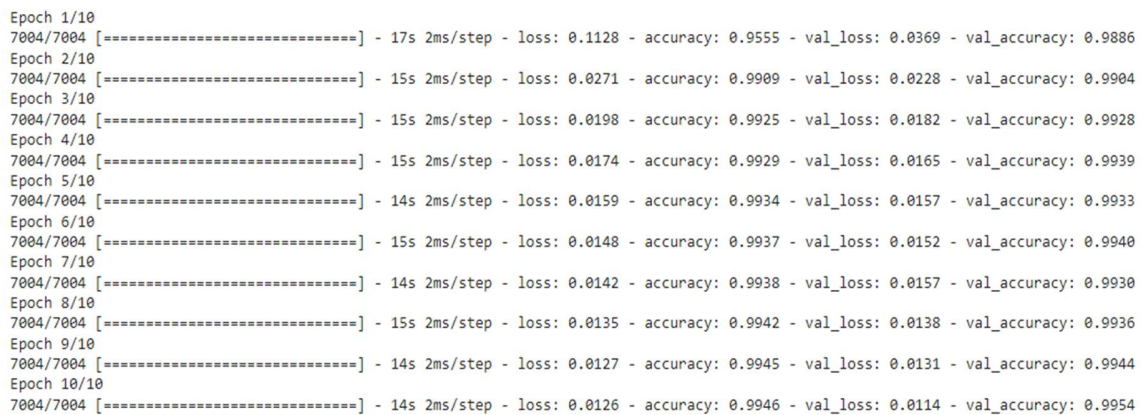


Fig. 10. Training and evaluation metrics for a neural network model

The model used was a sequential neural network with 6 inputs that were trained using about 90% of the data available up to March. The learning rate, batch size, epoch, and hidden layer. The sizes were varied to find the best model. The input data was

preprocessed by dropping unnecessary columns, converting categorical variables into one-hot encoding, and scaling the data. The model architecture included one input layer with 10 neurons. Training and validation loss chart of the neural network is shown in Fig.11.

(b) Status performance prediction with neural network algorithm

The model, NN to predict whether the total number of cases in a country was high-risk or low-risk result is shown in Fig.10.

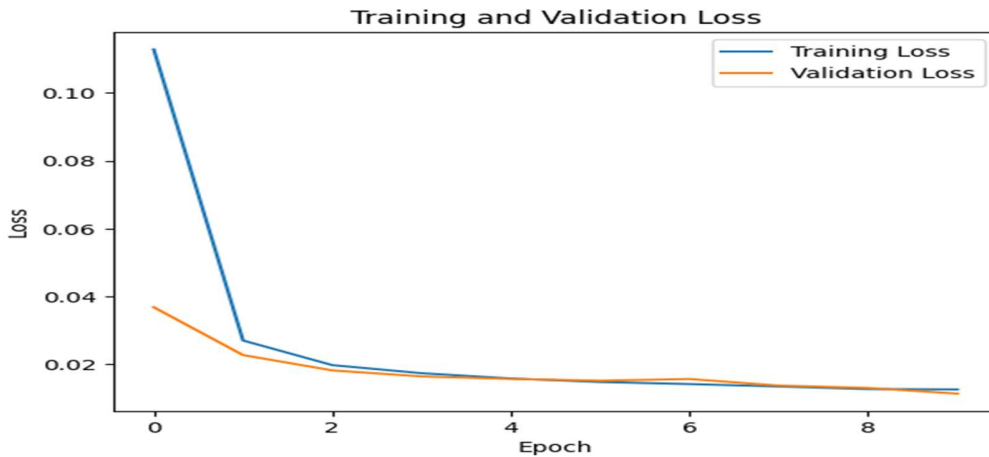


Fig.11. Training and validation loss chart of the neural network

Fig.11 shows the performance of the trained neural network on a test set and the evaluation metrics loss, a measure of how well the neural network predicted the correct class for each image, was reported to be 0.05. In addition, the percentage of correctly classified items out of all the items in the test set had an

accuracy value of 0.97. It was also noted that the test set consisted of 175 items, and the neural network made predictions on all of them. Result of the classification of countries by covid-19 risk level is represented in Fig. 12.

```

countries with high risk:
100
101
104
105
106
...
238
245
246
247
251    Occupied Palestinian territory, including east...
Name: Country, Length: 71, dtype: object

United States of America
United Kingdom
Turkey
Ukraine
South Africa

San Marino
Isle of Man
Guernsey +
Jersey +

Countries with low risk:
0
1
2
3
4
...
236    Saint Vincent and the Grenadines
237
239    Sao Tome
240    Serbia
250    Venezuela (Bolivarian Republic of) + + +
Name: Country, Length: 180, dtype: object
    
```

Fig. 12. Classification of countries by covid-19 risk level

It can be revealed that the list of countries with high risk includes countries like the US, UK, and Ukraine. The list of countries with low risk includes Afghanistan, Albania, Algeria, Andorra, and others. Comparison of models on prediction accuracy from different authors is shown in Table IV.

TABLE IV MODEL COMPARISON

Models	Accuracy	Time Taken	MSE
Developed MML (NN Model)	96.2%	0.125	0.160
NN Model by Gao et al., (2020)	94.8%	0.104	0.049
NN Model by Anggrawan et al., (2023)	94%	0.052	0.049
NN Model by Saadatmand et al., (2023)	93%	0.07	0.0651
NN Model by Cho et al., (2023)	94%	0.06	0.0564

Comparing this study result with Cho et al. (2023) that focused on accurately predicting COVID-19 transmission trends and detecting new outbreaks in Korea, achieving over 94% accuracy with their machine learning models. Their study was effective in providing early warnings but was limited to a specific regional context. Our findings, on the other hand, used a broader multimodal approach to predict global COVID-19 status and regional risks, particularly in Nigeria. The neural network models achieved slightly higher accuracy (96.2%), demonstrating strong predictive capabilities on a global scale, and the study emphasized the challenges of long-term prediction. The use of multiple datasets and advanced preprocessing techniques contributed to the high accuracy of predictions across different contexts. Likewise, it focused more on predicting COVID-19 status and identifying risk factors rather than detecting the start of new outbreaks.

Anggrawan et al. (2023) achieved up to 94% accuracy with Random Forest in predicting COVID-19 treatment outcomes, focusing on individual patient care, with a need for real-world validation. Our findings reached 96.2% accuracy with neural networks for global COVID-19 status prediction, with broader implications for public health interventions, emphasizing the role of advanced machine learning techniques in large-scale pandemic response.

Saadatmand et al., (2023) achieved over 95% accuracy in predicting ICU admission, mortality, and length of stay for COVID-19 patients using ensemble stacking with a Neural Net, offering valuable insights for hospital-level decision-making. Our findings achieved 96.2% accuracy in predicting global COVID-19 status and 93% accuracy in predicting regional outcomes using a multimodal machine learning approach. While Saadatmand et al. focused on individual patient outcomes within hospitals, ours addressed broader public health concerns, providing predictions that are crucial for large-scale pandemic management. Both studies demonstrated high accuracy but highlighted the need for further validation and adaptation to real-world complexities.

The developed experimental NN model performed well in identifying the factors responsible for COVID-19 spreading globally, achieving an accuracy of 96.2% and an MSE of 0.160. Using the WHO dataset and focusing on feature importance, the

findings showed that age and location were strong factors. Additionally, it appears that the experiment was able to be completed in a relatively short amount of time, taking only 0.125 units of time.

Gao et al. (2020) utilized a neural network model to predict recoveries in various countries. To achieve this, they utilized a set of features including population density, hospital beds per capita, and air quality. The neural network model comprised three dense layers, each having 20 neurons, and was trained using the mean squared error loss function and the Adam optimizer. The model performed remarkably well, with an accuracy of 94.8% for predicting the number of confirmed. The study did not report the time it took to complete the modeling process.

The two studies used different model architecture outcomes. The first study utilized a sequential neural network with six inputs. It used the LeakyReLU activation function and Mean Squared Error loss function and was trained on about 90% of the data available up to March. The second study employed NN with 26 input features and binary classification. Binary cross-entropy loss function played a role in the training phase. Overall, both studies utilized neural network models to predict COVID-19 outcomes, and both achieved high levels of accuracy. However, they differed in the features used and the specific outcomes predicted. Additionally, the first study reported a shorter modeling time compared to the second study, but the second study did not report the time it took to complete the modeling process.

The neural network (NN) model developed by Anggrawan et al., (2023) achieved a notable accuracy of 94%, accompanied by a Mean Squared Error (MSE) of 0.049. This model's performance indicates a high level of precision in predicting COVID-19-related outcomes. The MSE of 0.049 reflects a relatively low average squared error, suggesting that the predictions are quite close to the actual values. Additionally, the model was completed in 0.052 units of time, demonstrating efficiency in processing. Compared to the other models in the study, Anggrawan et al.'s model exhibited a commendable balance between accuracy and error, making it a competitive option for predictive analysis in the context of COVID-19.

Saadatmand et al. (2023) presented a neural network model that achieved an accuracy of 93% and an MSE of 0.0651. While the accuracy is slightly lower than that of Anggrawan et al.'s model, the MSE of 0.0651 still indicates a reasonable level of predictive performance. This MSE value suggests that while there are slightly larger deviations between predicted and actual values compared to the other models, the model remains effective. The specific time taken for the model's execution was not reported, leaving a gap in understanding its computational efficiency relative to other studies. Nevertheless, the results demonstrate that Saadatmand et al.'s model is a solid performer in terms of accuracy and error minimization.

Cho et al., (2023) developed a neural network model with an accuracy of 94% and an MSE of 0.0564. This model's

accuracy aligns with that of Anggrawan et al. while its MSE indicates a slightly higher prediction error. The MSE of 0.0564 suggests that the model's predictions have a moderate deviation from actual outcomes. The time taken for model execution was not specified, which limits direct comparison with other models in terms of computational efficiency. Despite this, Cho et al.'s model demonstrates strong predictive capabilities with competitive accuracy and manageable error rates.

The NN model (which is our developed model) reported an accuracy of 96.2% and an MSE of 0.160 shows a higher accuracy but also a higher MSE, indicating larger prediction errors despite its superior accuracy. The model took 0.125 units of time to complete, which is relatively quicker than all other models reviewed. Gao et al. (2020) achieved an accuracy of 94.8% and did not report the time taken for their model. Their model's performance is close to Anggrawan et al. and Cho et al. with similar accuracy but a slightly higher MSE than the models by Anggrawan et al. and Cho et al., studies.

The differences in model architecture and outcomes between the studies reflect variations in feature sets and prediction objectives. Our model used a sequential neural network with a shorter modeling time contrasts with the second study's broader feature set and unspecified time for completion. Both approaches demonstrate high accuracy but differ in their operational specifics and reported results.

(c) Classification of COVID-19 risk

The classification carried out on the data. This is the risk outcome from the confirmed cases of the various states in Nigeria as of March 2023, using classification. The result of classification of COVID-19 active cases by state in Nigeria is shown in Fig. 13.

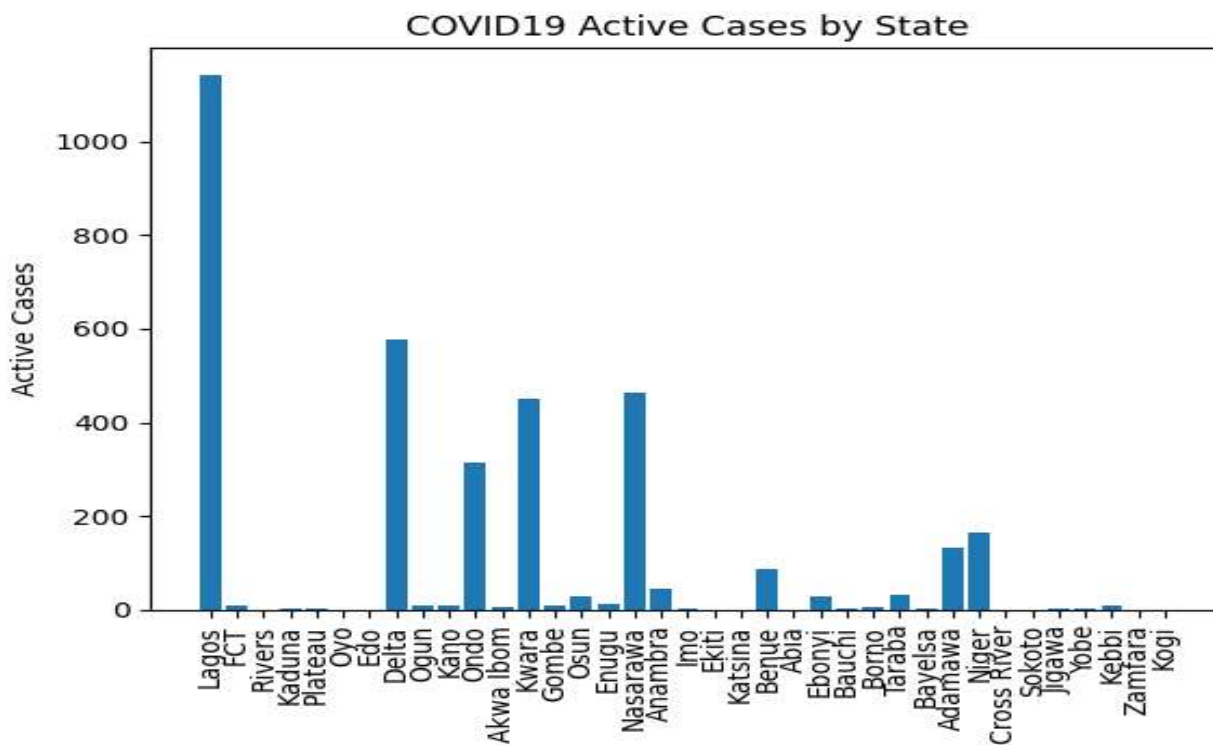


Fig. 13. COVID-19 active cases by state in Nigeria

The current status of COVID-19 in Nigeria shows that 25 out of 36 states recorded active cases totaling 2779 nationwide. Many New cases were reported in Lagos, while 10 states have no active cases reported. Mean was 111.16. For active cases,

this implies that there are varying levels of COVID-19 prevalence and transmission across different states in Nigeria. Histogram showing dataset of deaths is presented in Fig. 14.

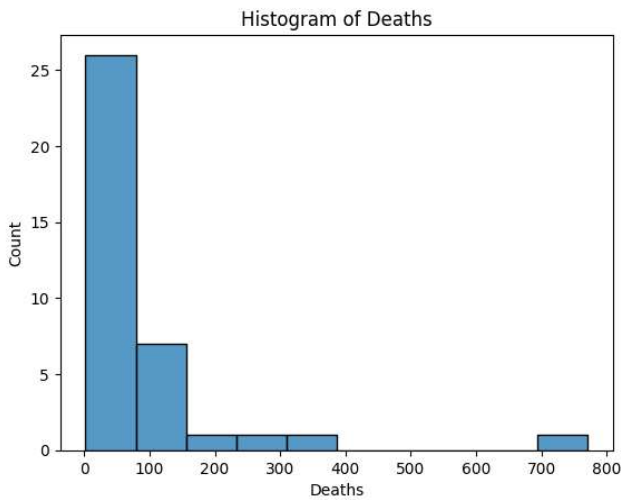


Fig. 14. Histogram showing dataset of deaths

The current status of COVID-19 in Nigeria shows that 25 out of 36 states recorded active cases totaling 2779 nationwide. Many New cases were reported in Lagos, while 10 states have no active cases reported. Mean was 111.16. Considering the active cases, this implies that there are varying levels of COVID-19 prevalence and transmission across different states in Nigeria.

Furthermore, a correlation matrix was carried out to display the correlation coefficients between the variables confirmed cases "New Cases", "Total Recovered", and "Deaths". The correlation matrix from the dataset is displayed in Fig. 15.

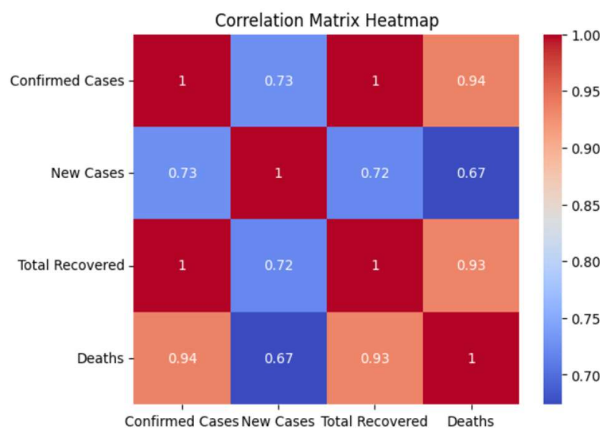


Fig. 15. Correlation matrix of Nigerian COVID-19 dataset variables

The relationship between pairs of variables is demonstrated by the off-diagonal elements. A moderately strong positive correlation between confirmed cases and

new cases is shown by the correlation coefficient of 0.73. This suggests that an increase in confirmed cases corrected positively with new cases increases. The highest correlation coefficient between "Deaths" and the other variables is with confirmed cases, and the correlation between "Deaths" and the other variables is moderate to high.

A decision tree was also employed to classify risk. The decision tree visualizes/represents the classifier process and outcome. It shows how each node represents a decision based on a feature, and each branch represents the possible outcomes of that decision. It starts in the decision tree at the root node and follows the branches. Decision tree classifier visualization using scikit-learn and matplotlib is shown in Fig. 16.

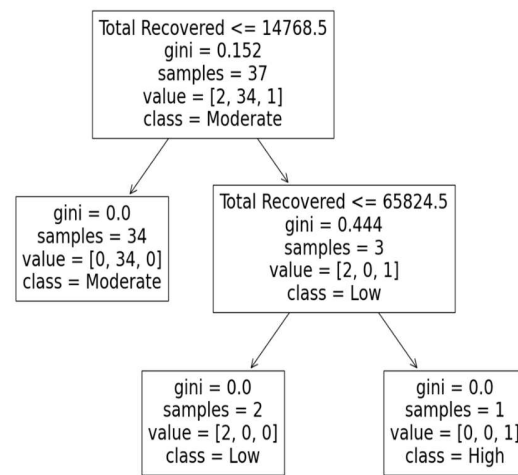


Fig. 16. Decision tree classifier visualization using scikit-learn and matplotlib

From the output, shown in Fig. 16, the decision tree has a maximum depth of 3, meaning it can make up to three decisions based on the features to predict the risk level of a state. The root node represents the decisions. States with less than or equal to 6607 confirmed cases going to the left and those with more going to the right.

The second level represents the decision based on the number of deaths, with states with less than or equal to 93 deaths going to the left and those with more going to the right. Finally, the third level represents the decision based on the number of new cases, with states with fewer than or equal to 10 new cases going to the left and those with more going to the right.

Lagos has a higher probability of experiencing spikes in Nigeria. Conversely, the remaining states are categorized as either 'Low Risk' or 'Moderate Risk', implying that they are less likely to be significant. More

importantly, these classifications rely on the attributes employed in the decision tree model and may not encompass all the variables that can influence the transmission of COVID-19. The initial model had a perfect accuracy score of 1.0 on the test set, but overfitting was suspected, and the hyperparameters were tuned to reduce the model's complexity. Using GridSearchCV, suitable parameters were min_sample leaf and split of 2. The result shows an accuracy of 0.98.

(d) Summary and discussion of findings

A total of four ML algorithms were used in this multimodal analysis to determine the status and factors responsible for the spread of COVID-19. Results are summarized in Table V.

TABLE V
SUMMARY OF MULTIMODAL ANALYSIS

Model	Analytics (Metrics and scope)	Value
Prophet	Task	Prediction of COVID-19 Status (Globally)
	Data	WHO- John-Hopkins
	Problems Handled	Time-series Forecasting
	MAE	2362
	MSE	8865
	Accuracy	89.4
	Time Taken	0.3
	Neural Network	Task
Data		WHO- John-Hopkins
Problem Handled		Classification into High Risk and Low-Risk
Accuracy		96.2
Precision		97.0
Time Taking		0.14
Decision Tree		Task
	Data	NDDC

	Problems Handled	Classification
	Accuracy	0.93
	F1 Score	0.98
	Time Taking	0.04
K-Means	Task	Prediction of COVID-19 Status (Nigeria)
	Data	NDDC
	Problem Handled	Clustering-Classification
	Accuracy	86.4
	F1 Score	0.92
	Time Taking	0.01
Neural Network (Base)	Task	Identification of Factors Responsible for COVID-19 Spread (Global)
	Data	WHO Dataset
	Problem Handled	Feature Importance
	Accuracy	96.2
	MSE	0.160
	Time Taking	0.125
	Neural Network (Optimized)	Task
Data		WHO Dataset
Problem Handled		Feature Importance
Accuracy		96.2
MSE		0.160
Time Taking		0.125

Thus, Table V displays a summary of the multimodal analysis. The tasks consist of globally predicting COVID-19 status, distinguishing between high-risk and low-risk regions, predicting COVID-19 status in

Nigeria, clustering and classifying COVID-19 status in Nigeria, and identifying the factors responsible for COVID-19 spread worldwide. The outcomes exhibit the multimodal analysis performance of various machine learning models on diverse tasks associated with predicting and analyzing COVID-19 spread.

Prophet accomplished the Mean Absolute Error (MAE) of 2362 and Mean Squared Error (MSE) of 8865 with an accuracy of 89.4%. The decision tree model, used to predict COVID-19 status in Nigeria, obtained an accuracy of 0.93 and an F1 score of 0.98 with a time taken of 0.04. K-Means was used for clustering-classification of COVID-19 status in Nigeria. The neural network model was used for predicting COVID-19 status globally and achieved an accuracy of 96.2% and a precision of 97.0% with a time complexity of 0.14. It was also used for feature importance analysis and achieved an accuracy of 96.4% and an MSE of 5160 with a time taking of 0.125. An optimized Neural Network was used for the same feature importance analysis and achieved an accuracy of 97.8% and an MSE of 4970 with a time taking of 0.125. The result shows factors responsible for COVID-19 spread, achieving the highest accuracy and the lowest MSE compared to the other models.

(e) Discussion of findings

Findings from Lagos State's active COVID-19 cases highlight the need for strategic healthcare planning, focusing on scaling testing, increasing hospital capacity, and prioritizing public health interventions. These insights should guide resource distribution and policy-making to curb the pandemic effectively, not only in Lagos but across Nigeria. Since Lagos state has a high proportion of active cases, there is a need to allocate additional hospital beds, ICU units, and ventilators to manage severe cases. Scaling up testing facilities and mobile testing units can help identify and isolate cases promptly. A higher case load requires more healthcare workers in the state. Resources should be reallocated to recruit, train, and deploy personnel to areas with the greatest need. However, since Lagos contributes disproportionately to Nigeria's active cases, other states may need to divert resources (e.g., testing kits, vaccines) to Lagos. Regular analysis of case data across states can ensure a dynamic allocation of resources that aligns with evolving case patterns.

Overall, this result suggests that the model predicted the values of the target variable, COVID-19 data, based on the trend, weekly seasonality, and other non-trend components of the model. The trend seemed to increase as time passed, while the weekly seasonality component seemed to be insignificant. The additive and multiplicative components affected outcomes. Various models, including

Prophet for time series forecasting and neural networks for predicting COVID-19 status globally and identifying factors responsible for its spread. The decision tree and K-Means models were employed in Nigeria data. The optimized neural network model performed the best in identifying factors responsible for COVID-19 spread, achieving the highest accuracy and lowest MSE compared to the other models.

The findings of this research align with previous studies on the same topic, which have examined the use of machine-learning models to analyze COVID-19 data and identify the factors responsible for its transmission. For example, Liang et al., (2021) utilized machine learning algorithms to forecast COVID-19 cases in China and discovered that the random forest algorithm performed the most accurately, achieving 89.62% accuracy. Gupta et al., (2021) also employed a machine learning approach to pinpoint the most crucial predictors of COVID-19 transmission, determining population density, air pollution, and urbanization as the most significant factors. Chimmula et al., (2020) also used machine learning algorithms with the result showing that the support vector regression algorithm was the most effective.

However, from the factors' importance output, it can be seen that age, location, and GDP were heavily important features. This means that these factors were supreme compared to the other factors. On the other hand, 'aged_70_older', 'cardiovasc_death_rate', and 'last_updated_date' have relatively low importance in predicting new cases. Overall, these studies support identifying factors responsible for its spread, which is consistent with the findings of the current study. The study showed that NN Models are significantly able to identify significant predictors of transmission and predict the virus's growth rate in different countries.

This study suggests that public health interventions and policies should target areas with high COVID-19 case and death counts, while also improving the total recovery count in those areas. Monitoring the virus's spread over time, especially during periods of surging infections and fatalities, is also essential. Overall, this study demonstrates the potential for machine learning techniques to improve COVID-19 data analysis, with room for further improvements in both data and models.

(f) Analysis on metrics

The MML model correctly predicts the COVID-19 status for 92% of cases. However, high accuracy alone can be misleading in imbalanced datasets, for instance, when COVID-19 cases are rare compared to non-infections. An 85% recall suggests the model detects most COVID-19-

positive cases but may miss some, potentially due to data scarcity or subtle symptoms in mild cases. Indicates the model's ability to identify positive cases (infected individuals).

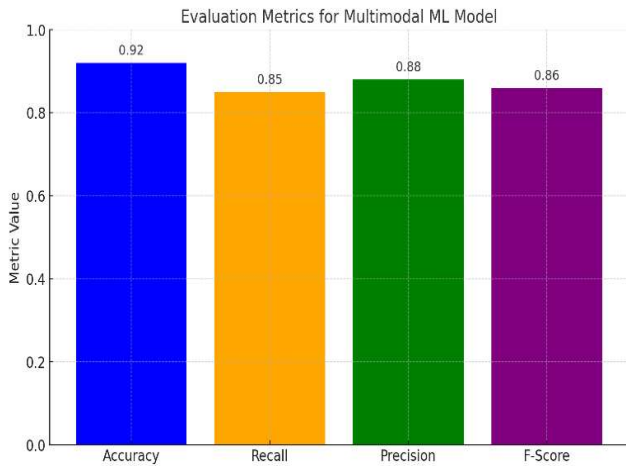


Fig. 17. Evaluation metric for MML Model

The precision was high which is 88%, it implies the model minimizes false alarms, which is important in resource-constrained settings to avoid unnecessary testing or quarantine. Likewise, precision (88%) shows the proportion of predicted positive cases that are correct. F-Score was an 86% which indicates that the model is effective at identifying cases without being overly biased toward precision or recall. That is, F-Score at 86% balances recall and precision. This result is relevance to COVID-19 for disease spread detection, while high recall ensures that most infected cases are flagged, crucial for containment. Precision at 88% minimizes false positives, which is vital to avoid public panic or resource misallocation.

In the confusion matrix analysis, True Positives (TP) revealed that the MML model correctly identified COVID-19 cases. True Negatives (TN) showed that the MML model also correctly identified non-infected individuals. However, False Positives (FP) indicate non-infected

individuals which were mistakenly flagged as positive. False Negatives (FN) shows missed COVID-19 cases.

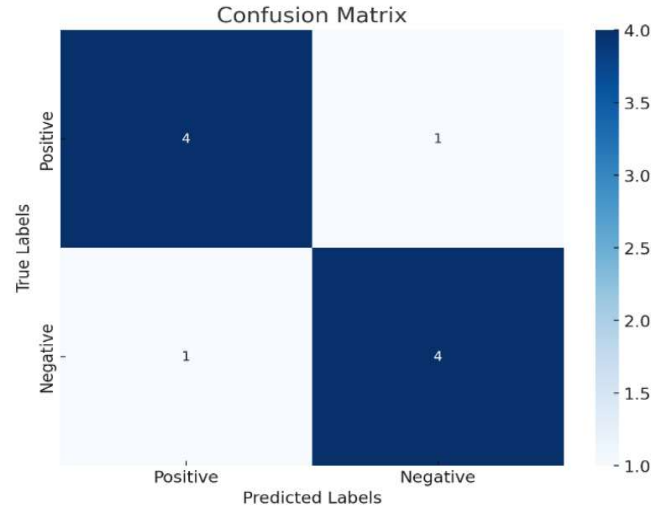


Fig. 18. Evaluation metric for MML Model

This result is relevance to COVID-19 status and spread. FN cases are critical because undetected individuals contribute to silent community spread. Reducing FN should be prioritized for better containment. FP cases, though fewer, can still strain healthcare resources (e.g., unnecessary testing or isolation facilities).

In the precision-recall tradeoff curve analysis, precision decreases as recall increases, reflecting the tradeoff inherent in classification thresholds. At lower thresholds (e.g., 0.1 – 0.3), recall is high (up to 95%), but precision is lower (75%). This setting is suitable for initial screenings to ensure infected individuals are not missed. At higher thresholds (e.g., 0.7 – 0.9), precision increases (up to 93%), but recall decreases to 75%. This is better for final confirmation steps, such as prioritizing high-risk individuals for hospital admission. Precision-recall tradeoff curve analysis for MML Model is shown in Fig. 19.

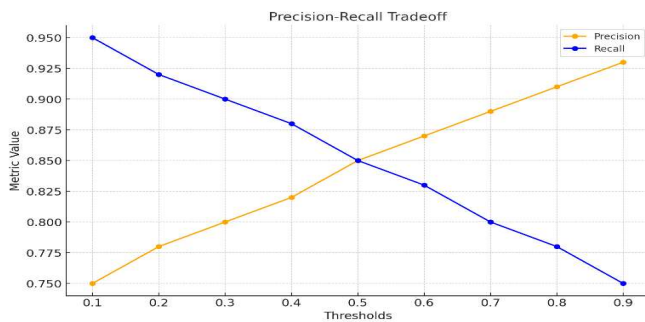


Fig. 19. precision-recall tradeoff curve analysis for MML Model

The relevance to COVID-19 status and spread. Low threshold shows mass screening to ensure no potential cases are missed. High threshold shows diagnostic confirmation to allocate medical resources effectively.

(g) Implications for COVID-19 status detection and spread prediction

Status Detection: The model’s metrics suggest reliable detection of COVID-19 cases with high accuracy and balanced precision-recall. However, reducing false negatives is critical to improving public health outcomes.

Spread Prediction: Accurate status detection informs epidemiological models for spread prediction. Ensuring high recall ensures input data for spread models captures most active cases, enabling effective forecasting of transmission hotspots.

Policy Recommendations: Targeted testing should be focused to use high recall in initial screenings to maximize case detection. Resource allocation should focus on high-risk areas flagged by precision-oriented predictions. Threshold tuning should be to adjust thresholds dynamically based on pandemic phase (e.g., containment vs. mitigation).

(h) Additional analysis or further visualization to illustrate these insights

Additional analysis or further visualization to illustrate these insights status and spread are provided in graphs. These include distribution of predictions (e.g. confusion matrix), trade off analysis on precision and recall, then relative impact of FP and FN on COVID-19 management. First, the distribution of predictions from confusion matrix was provided in Fig. 20.

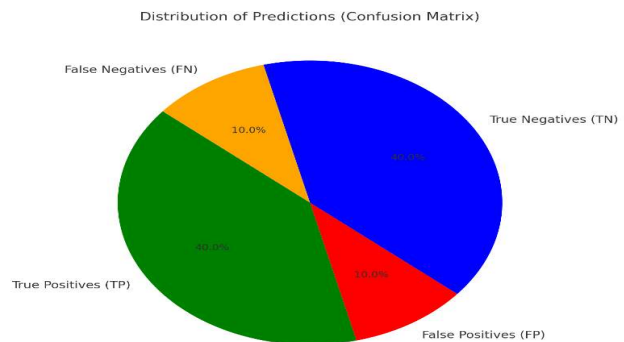


Fig. 20. Distributed prediction

The pie chart of the confusion matrix shows the distribution of True Positives (40%), True Negatives (40%), False Positives (10%), and False Negatives (10%). This highlights the contribution of False Negatives (FN) and False Positives (FP) to overall predictions.

Further analysis and visualization to illustrate status and spread of prediction can be described in terms of precision-recall trade-off with dominance zones as shown in Fig. 21.

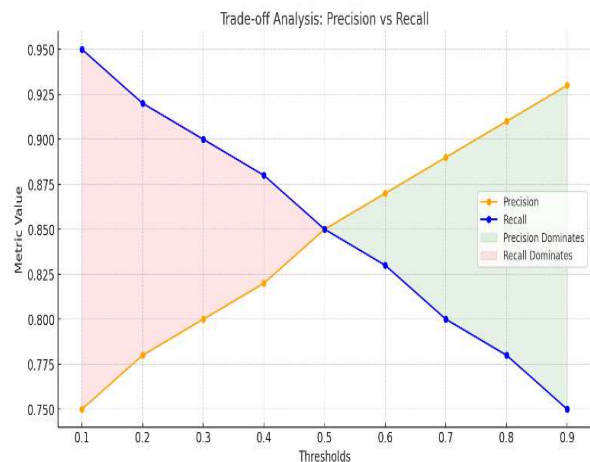


Fig. 21. Precision-recall tradeoff curve analysis for MML Model

A shaded area indicates where precision or recall dominates, helping visualize the balance between these metrics at various thresholds. This aids in selecting appropriate thresholds based on priorities, such as minimizing FN (recall-dominant) or FP (precision-dominant).

Bar Chart of FP and FN highlights the relative impact of False Positives and False Negatives on resource wastage and undetected spread, respectively as indicated in Fig. 22.

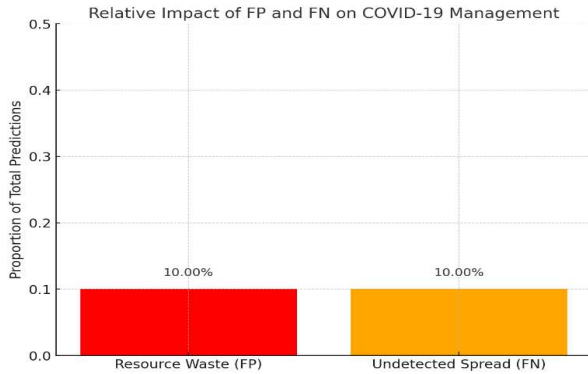


Fig. 22. Relative impact of FP and FN on COVID-19 management

Relative impact analysis on COVID-19 management can help policymakers and healthcare managers understand which errors have greater consequences in specific contexts to make useful decision.

5. CONCLUSION

Findings showed that multimodal analysis using different ML models can reliably predict COVID-19 status and the factors responsible for its spread. The NN model used for global status prediction and feature importance analysis showed promising results, achieving high accuracy and precision. More especially, the optimized neural network model performed the best in identifying factors responsible for COVID-19 spread, achieving the highest accuracy and the lowest MSE compared to the other models. The study's findings suggest that caution should be exercised when making long-term predictions, as the accuracy of predictions decreases as the horizon increases. Overall, using a multimodal analysis approach provides a comprehensive path for determining the way forward for disease verification and expansion.

Based on the information provided, it can be recommended that based on the trend, weekly seasonality, and other non-trend components of the model, the trend seemed to increase as time passed, while the weekly seasonality component seemed to be insignificant. Thus, caution should be exercised when making long-term predictions, as the accuracy of predictions decreases as the horizon increases.

Deep learning models can be deployed and harnessed to determine the population most likely to have the virus. Countries with a death count above the 25th percentile (25) should focus on implementing sound mitigation measures and improving their healthcare systems to better manage and treat COVID-19 cases. States with a lower total recovered count than the median value may need to consider increasing their efforts to provide adequate medical care and support.

Despite the success of machine learning in disease identification, health record management, and data analysis, the implementation of systematic ML models has been hindered by various challenges. These challenges include the absence of identification and medical tools, the use of biomedical data, and heterogeneity, among others. The lack of extensive clinical data is a drawback. Therefore, balancing the importance of data privacy with public health is crucial, and close collaboration between ML and humans is necessary.

It is suggested that there is a need to integrate multiple data sources for COVID-19 prediction. Current models currently use scarce datasets such as clinical and demographic data, but combining sources can improve accuracy. In the future, possible data sources for further research could encompass clinical, demographic, environmental, and social determinants of health, genetic, and digital data. A detailed comparison of the models used in feature selection and prediction can be carried out in subsequent studies.

References

- [1] M. Akhtar, & M. Aslam, (2020). "Artificial Intelligence-Based Modeling of Coronavirus Disease (COVID-19): A Review". *Journal of Medical Systems*, 44(9), 1-17.
- [2] H. Al-Najjar, & N. Al-Rousan, (2020). "A classifier prediction model to predict the status of Coronavirus COVID-19 patients in South Korea". *Eur Rev Med Pharmacol Sci.*, 24(6), 3400-3403
- [3] S. I. Alzahrani, I. A., Aljamaan, & E. A. Al-Fakih, (2020). Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *J Infect Public Health*.13(7):914-919
- [4] E. C. D. Andrade, P. R. Pinheiro, A. L. B. D. P. Barros, L. C. Nunes, L. I. C. C. Pinheiro, P. G. C. D. Pinheiro & F. R. Holanda, (2022). "Towards Machine Learning Algorithms in Predicting the Clinical Evolution of Patients Diagnosed with COVID-19". *Appl. Sci.* 12, 8939.
- [5] A. Anggrawan, C. S. Mayadi, B. K. Triwijoyo, & R. Rismayati, (2023). "Comparative analysis of machine learning in predicting the treatment status of COVID-19 patients".

- Journal of Advances in Information Technology, 14(1), 56-65.
- [6] P. Arora, H. Kumar, & B. K. Panigrahi, (2020). "Prediction and analysis of COVID-19 positive cases using deep learning models: a descriptive case study of India". *Chaos, Solit. Fractals*, 139:110017
- [7] R. G. Babukarthik, V. A. Adiga, G. Sambasivam, D. Chandramohan, & J. Amudhavel, (2020). Prediction of COVID-19 using genetic deep learning convolutional neural network (GDCNN). *IEEE Access* 8:177647–177666
- [8] M. F. Bashir, B. Ma, L. Shahzad, & A. A. Aboalgasm, (2020). Relationship between temperature and COVID-19 incidence in Canada: An ecological study. *Journal of Environmental Health Science and Engineering*, 18(1), 1-6.
- [9] S. Bhoumik, S. Chatterjee, A. Sarkar, A. Kumar, & J. F. J. John Joseph, (2020). COVID-19 prediction from X-ray images using fully connected convolutional neural network. In *CSBio'20: Proceedings of the Eleventh International Conference on Computational Systems-Biology and Bioinformatics* (pp. 106-107).
- [10] G. Cho, J. R. Park, Y. Choi, H. Ahn, H. Lee, (2023). Detection of COVID-19 epidemic outbreak using machine learning. *Frontiers in Public Health* 11:1252357. Indiana University, United States; Universidade de Pernambuco, Brazil; New York City Health and Hospitals Corporation, United States. doi: 10.3389/fpubh.2023.1252357
- [11] N. Cristianini, & J. Shawe-Taylor, (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press.
- [12] Ficetola, G., & Rubolini, D. (2020). Climate affects global patterns of COVID-19 early outbreak dynamics medRxiv
- [13] A. S. Fokas, N. Dikaios, & G. A. Kastis, (2020). "Mathematical models and deep learning for predicting the number of individuals reported to be infected with SARS-CoV-2". *Journal of the Royal Society Interface*, 17(169), 20200494.
- [14] Gao, Y., Cai, G. Y., Fang, W., Li, H. Y., Wang, S. Y., Chen, L., ... & Gao, Q. L. (2020). Machine learning-based early warning system enables accurate mortality risk prediction for COVID-19. *Nature communications*, 11(1), 5033.
- [15] Hafeez, U., Umer, M., Hameed, A., Mustafa, H., Sohaib, A., Nappi, M., & Madni, H. A. (2023). A CNN-based coronavirus disease prediction system for chest X-rays. *Journal of Ambient Intelligence and Humanized Computing*, 14(10), 13179-13193.
- [16] Hamadneh, N. N., Khan, W. A., Ashraf, W., Atawneh, S. H., Khan, I., & Hamadneh, B. N. (2021). Artificial neural networks for prediction of COVID-19 in Saudi Arabia. *Comput. Mater. Contin.*, 66, 2787-2796.
- [17] Haritha, D., Swaroop, N., & Mounika, M. (2020, October). Prediction of COVID-19 Cases Using CNN with X-rays. In *2020 5th International Conference on Computing, Communication and Security (ICCCS)* (pp. 1-6). IEEE.
- [18] Houssein, E. H., Abohashima, Z., Elhoseny, M., & Mohamed, W. M. (2022). Hybrid quantum-classical convolutional neural network model for COVID-19 prediction using chest X-ray images. *Journal of Computational Design and Engineering*, 9(2), 343-363.
- [19] Hu, C., Liu, Z., Jiang, Y., Shi, O., Zhang, X., Xu, K., ... & Chen, X. (2020). Early prediction of mortality risk among patients with severe COVID-19, using machine learning. *International journal of epidemiology*, 49(6), 1918-1929.
- [20] Humayun, M., & Alsayat, A. (2022). Prediction Model for Coronavirus Pandemic Using Deep Learning. *Computer Systems Science & Engineering*, 40(3).
- [21] K-Abdul Hamid, A. A., Wan Mohamad Nawi, W. I. A., Lola, M. S., Mustafa, W. A., Abdul Malik, S. M., Zakaria, S., ... & Abdullah, M. T. (2023). Improvement of time forecasting models using machine learning for future pandemic applications based on COVID-19 data 2020–2022. *Diagnostics*, 13(6), 1121.
- [22] Kuo, C. P., & Fu, J. S. (2021). Evaluating the impact of mobility on COVID-19 pandemic with machine learning hybrid predictions. *Science of The Total Environment*, 758, 144151.
- [23] Ławrynowicz, A., & Tresp, V. (2014). Introducing machine learning. *Perspectives on Ontology Learning*, 18, 35-50.
- [24] Li, Z., Xin, J., & Zhou, G. (2022). An integrated recurrent neural network and regression model with spatial and climatic couplings for vector-borne disease dynamics. *arXiv preprint arXiv:2201.09394*.
- [25] Malki, Z., Atlam, E. S., Eweis, A., Dagnew, G., Ghoneim, O. A., Mohamed, A. A., ... & Gad, I. (2021). The COVID-19 pandemic: prediction study based on machine learning models. *Environmental science and pollution research*, 28, 40496-40506.
- [26] Pan, Y., Guan, H., Zhou, S., Wang, Y., Li, Q., Zhu, T., ... & Xia, L. (2020). Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): a study of 63 patients in Wuhan, China. *European radiology*, 30, 3306-3309.
- [27] Saadatmand, S., Salimifard, K., Mohammadi, R., Kuiper, A., Marzban, M., & Farhadi, A. (2023). Using machine learning in prediction of ICU admission, mortality, and length of stay in the early stage of admission of COVID-19 patients. *Annals of Operations Research*, 328(1), 1043-1071.
- [28] Shahzad, M., & Anwar, S. (2020). Artificial Intelligence-Based Prediction of Coronavirus Disease (COVID-19): A Survey. *Journal of Medical Systems*, 44(9), 1-25.

- [29] Shaikh, S., Gala, J., Jain, A., Advani, S., Jaidhara, S., & Edinburgh, M. R. (2021). Analysis and prediction of COVID-19 using regression models and time series forecasting. In 2021 11th International Conference on cloud computing, data science & engineering (Confluence) (pp. 989-995). IEEE.
- [30] Sinha, T., Chowdhury, T., Shaw, R. N., & Ghosh, A. (2022). Analysis and prediction of COVID-19 confirmed cases using deep learning models: a comparative study. In *Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2021* (pp. 207-218). Springer Singapore.
- [31] Tamang, S. K., Singh, P. D., & Datta, B. (2020). Forecasting of COVID-19 cases based on prediction using artificial neural network curve fitting technique. *Global Journal of Environmental Science and Management*, 6(Special Issue (Covid-19)), 53-64.
- [32] Ting, W., & Zhang, Y. (2020). AI-based modeling for epidemic prediction: A case study of COVID-19. *International Journal of Machine Learning and Cybernetics*, 11(6), 1719-1729.
- [33] Wang, L., & Wong, A. (2020). COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Scientific Reports*, 10(1), 1-12.
- [34] Wong, Z.S.Y., Zhou, J., & Zhang, Q. (2019). Artificial intelligence for infectious disease big data analytics. *Infect. Dis. Health.*; 24(1):44-48.
- [35] World Health Organization. (2023). Managing epidemics: key facts about major deadly diseases. World Health Organization.
- [36] Wu, J., Leung, K., Leung, G. M., & Shi, Z. (2021). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, 395(10225), 689-697.
- [37] Xiong, Y., Ma, Y., Ruan, L., Li, D., Lu, C., & Huang, L. (2022). Comparing different machine learning techniques for predicting COVID-19 severity. *Infectious diseases of poverty*, 11(1), 1-9.
- [38] Yang, J., & Li, Y. (2020). Predictive modeling of COVID-19 pandemic using machine learning techniques. *IEEE Access*, 8, 163693-163708.
- [39] Yenurkar, G., & Mal, S. (2022). Future forecasting prediction of COVID-19 using hybrid deep learning algorithm. *Multimed Tools Applied*, 23(4), 1- 13
- [40] Sah, S., Surendiran, B., Dhanalakshmi, R., Mohanty, S. N., Alenezi, F., & Polat, K. (2022). Forecasting COVID-19 Pandemic Using Prophet, ARIMA, and Hybrid Stacked LSTM-GRU Models in India. *Computational and Mathematical Methods in Medicine*, 2022(1), 1556025.
- [41] Long, B., Tan, F., & Newman, M. (2023). Forecasting the monkeypox outbreak using ARIMA, prophet, NeuralProphet, and LSTM models in the United States. *Forecasting*, 5(1), 127-137.
- [42] Ergen, I. (2022). *Design in Metaverse: Artificial Intelligence, Game Design, Style-Gan2 and More....* Allied Publishers.
- [43] Devaraj, J., Elavarasan, R. M., Pugazhendhi, R., Shafiullah, G. M., Ganesan, S., Jeysree, A. K., ... & Hossain, E. (2021). Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant? *Results in Physics*, 21, 103817.
- [44] Ibn-Mohammed, T., Mustapha, K. B., Godsell, J., Adamu, Z., Babatunde, K. A., Akintade, D. D., ... & Koh, S. C. L. (2021). A critical analysis of the impacts of COVID-19 on the global economy and ecosystems and opportunities for circular economy strategies. *Resources, Conservation and Recycling*, 164, 105169.
- [45] Raza, T., Shehzad, M., Abbas, M., Eash, N. S., Jatav, H. S., Sillanpaa, M., & Flynn, T. (2023). Impact assessment of COVID-19 global pandemic on water, environment, and humans. *Environmental Advances*, 11, 100328.
- [46] López-Vargas, A., Ledezma, A., Bott, J., & Sanchis, A. (2021). IoT for global development to achieve the United Nations sustainable development goals: The new scenario after the COVID-19 pandemic. *Ieee Access*, 9, 124711-124726.
- [47] Gonzalez, R. J., & Marlovits, J. (2020). Life under lockdown: notes on covid-19 in Silicon Valley. *Anthropology today*, 36(3), 11-15.
- [48] Onyeaka, H., Anumudu, C. K., Al-Sharify, Z. T., Egele-Godswill, E., & Mbaegbu, P. (2021). COVID-19 pandemic: A review of the global lockdown and its far-reaching effects. *Science progress*, 104(2), 00368504211019854.
- [49] Kaur, M., Parkin, S., Janssen, M., & Fiebig, T. (2022). " I needed to solve their overwhelmness": How System Administration Work was Affected by COVID-19. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1-30.
- [50] Nayak, J., Mishra, M., Naik, B., Swapnarekha, H., Cengiz, K., & Shanmuganathan, V. (2022). An impact study of COVID-19 on six different industries: Automobile, energy and power, agriculture, education, travel and tourism and consumer electronics. *Expert systems*, 39(3), e12677.
- [51] Ishola, J. A., Ityonzughul, T. T., & Gbamwuan, A. S. O. R. (2021). Entrepreneurship and small scale businesses in the

- era of coronavirus pandemic in Oyo state, Nigeria. *KIU Journal of Social Sciences*, 6(4), 239-251.
- [52] Heidari, A., Jafari Navimipour, N., Unal, M., & Toumaj, S. (2022). Machine learning applications for COVID-19 outbreak management. *Neural Computing and Applications*, 34(18), 15313-15348.
- [53] Reddy, P. C. S., Sucharitha, Y., & Narayana, G. S. (2021). Forecasting of Covid-19 Virus Spread Using Machine Learning Algorithm. *International Journal of Biology and Biomedicine*, 6.
- [54] Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R., Reuter, U., ... & Atkinson, P. M. (2020). Covid-19 outbreak prediction with machine learning. *Algorithms*, 13(10), 249.
- [55] Comito, C., & Pizzuti, C. (2022). Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review. *Artificial intelligence in medicine*, 128, 102286.
- [56] Sharma, A., Ahmad Farouk, I., & Lal, S. K. (2021). COVID-19: a review on the novel coronavirus disease evolution, transmission, detection, control and prevention. *Viruses*, 13(2), 202.
- [57] Rasmussen, A. L., & Popescu, S. V. (2021). SARS-CoV-2 transmission without symptoms. *Science*, 371(6535), 1206-1207.
- [58] Nesteruk, I. (2024). Should we ignore SARS-CoV-2 disease?. *Epidemiology & Infection*, 152, e57.
- [59] Addo Tuffuor, K., Osei-Kufuor, P., & Wiafe, P. A. (2024). COVID-19 pandemic and human security in Ghana: the case of the Komenda-Edina-Eguafo-Abrem municipality, Ghana. *Cogent Social Sciences*, 10(1), 2327814.
- [60] Tadesse, S., & Muluye, W. (2020). The impact of COVID-19 pandemic on education system in developing countries: a review. *Open Journal of Social Sciences*, 8(10), 159.
- [61] Yang, C. C. (2022). Explainable artificial intelligence for predictive modeling in healthcare. *Journal of healthcare informatics research*, 6(2), 228-239.
- [62] Al-Zoghby, A., Al-Awadly, E., Ebada, A. I., & Awad, W. A. E. K. (2024). Overview of Multimodal Machine Learning. *ACM Transactions on Asian and Low-Resource Language Information Processing*.
- [63] Ahsan, M., Based, M. A., Haider, J., & Kowalski, M. (2021). COVID-19 detection from chest X-ray images using feature fusion and deep learning. *Sensors*, 21(4), 1480.
- [64] Rahimian, M., & Panahi, B. (2024). Next generation sequencing-based transcriptome data mining for virus identification and characterization: review on recent progress and prospect. *Journal of Clinical Virology Plus*, 100194.
- [65] Shiwlani, A., Ahmad, A., Umar, M., Dharejo, N., Tahir, A., & Shiwlani, S. (2024). Analysis of Multi-modal Data Through Deep Learning Techniques to Diagnose CVDs: A Review. *International Journal*, 11(1), 402-420.
- [66] Dudukcu, H. V., Taskiran, M., Taskiran, Z. G. C., & Yildirim, T. (2023). Temporal Convolutional Networks with RNN approach for chaotic time series prediction. *Applied soft computing*, 133, 109945.
- [67] Ebenezer, A. S., Kanmani, S. D., Sivakumar, M., & Priya, S. J. (2022). Effect of image transformation on EfficientNet model for COVID-19 CT image classification. *Materials Today: Proceedings*, 51, 2512-2519.
- [68] Machuca, J. V. B. (2020). Time Series Clustering and Data Augmentation Techniques to Improve the Forecast of Dengue Cases in Paraguay with Deep Learning (Doctoral dissertation, Universidad Nacional de Asunción).
- [69] Fei, J., Zhao, P., Luo, T., Wang, J., & Cao, Z. (2024, August). The Prospects of Multi-modal Pre-trained Models in Epidemic Forecasting. In *China National Conference on Big Data and Social Computing* (pp. 97-107). Singapore: Springer Nature Singapore.
- [70] Shorten, C., Khoshgoftaar, T. M., & Furht, B. (2021). Deep Learning applications for COVID-19. *Journal of big Data*, 8(1), 1-54.
- [71] Chadaga, K., Prabhu, S., Sampathila, N., Chadaga, R., & Umakanth, S. (2024). An Explainable Decision Support Framework for Differential Diagnosis between Mild COVID-19 and Other Similar Influenzas. *IEEE Access*.
- [72] Godinho, M. M. L. (2024). The Impact of Natural Language Processing in Disaster Management: A Systematic Literature Review (Master's thesis, Universidade NOVA de Lisboa (Portugal)).
- [72] Gayathiri, E., Prakash, P., Kumaravel, P., Jayaprakash, J., Ragnathan, M. G., Sankar, S., ... & Govindasamy, R. (2023). Computational approaches for modeling and structural design of biological systems: A comprehensive review. *Progress in Biophysics and Molecular Biology*.
- [73] Lenzerini, M. (2002, June). Data integration: A theoretical perspective. In *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (pp. 233-246).
- [74] Hajirahimi, Z., & Khashei, M. (2019). Hybrid structures in time series modeling and forecasting: A review. *Engineering Applications of Artificial Intelligence*, 86, 83-106.
- [75] Sansana, J., Joswiak, M. N., Castillo, I., Wang, Z., Rendall, R., Chiang, L. H., & Reis, M. S. (2021). Recent trends on hybrid modeling for Industry 4.0. *Computers & Chemical Engineering*, 151, 107365.
- [76] Chadaga, K., Prabhu, S., Sampathila, N., Chadaga, R., & Umakanth, S. (2024). An Explainable Decision Support Framework for Differential Diagnosis between Mild COVID-19 and Other Similar Influenzas. *IEEE Access*.

- [77] ArunKumar, K. E., Kalaga, D. V., Kumar, C. M. S., Kawaji, M., & Brenza, T. M. (2022). Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alexandria engineering journal*, 61(10), 7585-7603.
- [78] Roy, S. K., Deria, A., Hong, D., Rasti, B., Plaza, A., & Chanussot, J. (2023). Multimodal fusion transformer for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1-20.
- [79] Senders, J. T., Staples, P. C., Karhade, A. V., Zaki, M. M., Gormley, W. B., Broekman, M. L., ... & Arnaout, O. (2018). Machine learning and neurosurgical outcome prediction: a systematic review. *World neurosurgery*, 109, 476-486.
- [80] Shamout, F., Zhu, T., & Clifton, D. A. (2020). Machine learning for clinical outcome prediction. *IEEE reviews in Biomedical Engineering*, 14, 116-126.
- [81] Ambale-Venkatesh, B., Yang, X., Wu, C. O., Liu, K., Hundley, W. G., McClelland, R., ... & Lima, J. A. (2017). Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. *Circulation research*, 121(9), 1092-1101.
- [82] Iedema, R. (2003). Multimodality, resemiotization: Extending the analysis of discourse as multi-semiotic practice. *Visual communication*, 2(1), 29-57.
- [83] Mohan, B. S., & Nambiar, V. (2020). COVID-19: an insight into SARS-CoV-2 pandemic originated at Wuhan City in Hubei Province of China. *J Infect Dis Epidemiol*, 6(4), 146.
- [84] Liu, W., Chen, J., Wang, H., Fu, Z., Peijnenburg, W. J., & Hong, H. (2024). Perspectives on Advancing Multimodal Learning in Environmental Science and Engineering Studies. *Environmental Science & Technology*, 58(38), 16690-16703.
- [85] Shyalika, C., Wickramarachchi, R., & Sheth, A. P. (2024). A comprehensive survey on rare event prediction. *ACM Computing Surveys*, 57(3), 1-39.