Peeyush kumar Pathak / Afr.J.Bio.Sc. 6(5) (2024). 8471-8491 ISSN: 2663-2187

https://doi.org/ 10.33472/AFJBS.6.5.2024. 8471-8491

African Journal of Biological Sciences

A proposed Algorithm and Models for Predicting Post-Pandemic Health Conditions

Peeyush kumar Pathak[1]
Research Scholar,Computer science, Integral University
Lucknow, India
peeyushkumarpathak@gmail.com[1]

Manish Madhava Tripathi[2]
Professor. Computer science, Integral University
Lucknow, India
mmt@iul.ac.in[2]

Article History
Volume 6, Issue 5, 2024
Received: 25 May 2024
Accepted: 02 Jun 2024
: *10.33472/AFJBS.6.5.2024. 8471-8491*

Abstract

The recent global outbreak of COVID-19 severely impacted world health systems, human health, economies, and daily life. Countries were unprepared to tackle this emerging health crisis. Health professionals were unable to foresee the virus's spread, future developments, and the potential impact on lives should a similar pandemic occur again.The COVID-19 pandemic has profoundly disrupted global health systems, economies, and daily life, revealing significant gaps in our preparedness for such crises. This study proposes an algorithm and a set of machine learning models to predict post-pandemic health conditions, aiming to better prepare for future pandemics. Utilizing diverse datasets, including electronic health records (EHRs), public health databases, and patient surveys, In order to extract useful features from data, we preprocess and analyse it. We use a number of machine learning algorithms in our research, including logistic regression, neural networks, support vector machines, decision trees, and random forests. In order to make the model more accurate and reliable, we use strict feature selection and cross-validation procedures. Metrics like F1-score, recall, accuracy, and precision are used to assess the suggested models. We utilize SHAP values and LIME for model interpretation, ensuring transparency and understanding of the predictive factors. The findings demonstrate the potential of machine learning in forecasting long-term health conditions, thereby contributing to more robust health system responses and improved public health strategies in the event of future pandemics.

Index terms Post-Pandemic Health, Feature extraction, Feature selection, Corona disease

Introduction

Worldwide, people have been caught off guard by the unprecedented COVID-19 epidemic, which has exposed serious weaknesses and lack of preparation in health systems, economics, and everyday life. The rapid and widespread transmission of the virus overwhelmed healthcare infrastructures, leading to substantial morbidity and mortality. Beyond the immediate impact, COVID-19 has also resulted in long-term health consequences, commonly referred to as "post-pandemic" or "long COVID" conditions. These conditions include persistent symptoms such as fatigue, respiratory issues, cognitive impairments, and cardiovascular problems, which continue to affect individuals long after the acute phase of the infection.The inability to predict the rise, trajectory, and impact of COVID-19 highlighted the urgent need for advanced tools and methodologies to better understand and manage future pandemics [1]. Traditional epidemiological models, while useful, proved insufficient in capturing the complex dynamics and multifaceted impacts of the pandemic. As a result, there is a growing interest in leveraging machine learning (ML) techniques to enhance predictive capabilities and provide actionable insights for healthcare professionals and policymakers.Machine learning, with its ability to process vast amounts of data and identify patterns, offers a promising approach to predicting post-pandemic health conditions. By integrating diverse datasets such as electronic health records (EHRs), public health databases, patient surveys, and genomic data, ML models can uncover relationships and trends that are not immediately apparent. These models can assist in forecasting long-term health outcomes, identifying high-risk populations, and informing targeted interventions to mitigate adverse effects.

This study proposes the development of a comprehensive algorithm and a suite of ML models aimed at predicting post-pandemic health conditions. The methodology involves data collection from multiple sources, rigorous preprocessing to ensure data quality, and the application of various ML techniques, including logistic regression, decision trees, random forests, support vector machines, and neural networks. The models will be evaluated using metrics such as accuracy, precision, recall, and F1-score, and interpreted using SHAP values and LIME to provide transparency in decision-making.The goal of this research is to enhance our predictive capabilities and preparedness for future health crises [2]. By accurately forecasting post-pandemic health conditions, we can better allocate healthcare resources, design effective public health strategies, and ultimately improve patient outcomes. This interdisciplinary approach not only advances the field of health informatics but also contributes to building more resilient health systems capable of withstanding future pandemics.

Post-Pandemic Health Challenges and the Role of Machine Learning

The aftermath of the COVID-19 pandemic has introduced a host of new health challenges, collectively referred to as "post-pandemic" or "long COVID" conditions. These conditions encompass a wide array of symptoms and complications that persist long after the initial infection has subsided. Commonly reported issues include chronic fatigue, respiratory difficulties, cognitive impairments, cardiovascular complications, and mental health disorders such as anxiety and depression. The heterogeneity and unpredictability of these symptoms have made it difficult for healthcare providers to diagnose and treat affected individuals effectively [3]. One of the critical challenges in addressing post-pandemic health conditions is the lack of comprehensive data and predictive models. Traditional approaches to studying long-term health effects are often slow and reactive, relying on longitudinal studies that may not capture the full scope of the problem in a timely manner. Machine learning (ML) has the potential to revolutionize this field. Machine learning (ML) techniques enable the discovery of previously unseen patterns and correlations by mining massive and varied datasets. Because of this, we can learn about and control the effects on people's health in the future with greater certainty. Electronic health records (EHRs), genomic data, patient reports, environmental variables, and other forms of data can be used to train machine learning models.. For instance, EHRs provide a rich source of longitudinal data that can reveal trends in health outcomes over time. Patient-reported outcomes offer insights into symptoms and quality of life that are not always captured in clinical settings. Genomic data can help identify genetic predispositions to certain long-term conditions, while environmental factors such as socioeconomic status and access to healthcare services can provide context for understanding disparities in health outcomes. Developing effective ML models for predicting post-pandemic health conditions involves several key steps. First, data must be collected and pre-processed to ensure quality and consistency [4]. Data normalization, encoding categorical variables, and handling missing data are all part of this. After that, the best indicators of health outcomes over the long run are chosen using feature selection approaches. Next, the data is processed using a variety of ML algorithms, including logistic regression, decision trees, random forests, support vector machines, and neural networks. To guarantee the validity and reliability of these models, metrics such as accuracy, precision, recall, and F1-score are used for evaluation. Another important feature of ML models in healthcare is their interpretability. Tools like SHAP values and LIME can be used to provide transparency, helping healthcare professionals understand how the models make predictions. This transparency is essential for building trust in ML models and for making informed clinical decisions. The benefits of applying ML to predict post-pandemic health conditions are manifold. Accurate predictions can help healthcare providers identify high-risk patients and intervene early, potentially preventing the progression of severe complications. Policymakers can use these insights to allocate resources more effectively, ensuring that healthcare systems are better prepared for future pandemics. Additionally, understanding the long-term impacts of COVID-19 can guide public health strategies and inform the development of targeted treatments and rehabilitation programs. In conclusion, the integration of machine learning into the study of post-pandemic health conditions represents a significant advancement in our ability to respond to and manage the long-term effects of global health crises. By harnessing the power of data and advanced analytics, we can move towards a more proactive and informed approach to healthcare, ultimately improving outcomes for individuals affected by pandemics.

Importance of Machine Learning in Addressing Post-Pandemic Health Conditions

The need of cutting-edge technology in handling international health emergencies has been highlighted by the COVID-19 pandemic. Machine learning's (ML) capacity to sift through mountains of data and extract useful insights makes it stand out among these technologies. Because of its superior accuracy and efficiency in predicting, diagnosing, and managing long-term health outcomes, ML is a promising tool for dealing with post-pandemic health issues. Massive and complicated datasets are no match for ML's processing power and analytical prowess. An enormous quantity of health-related data has been produced by the pandemic from a variety of sources, such as patient surveys, genetic information, electronic health records (EHRs), and social determinants of health. When applied to this kind of data, ML algorithms can reveal links and patterns that would otherwise go unnoticed by traditional statistical methods. This allows for a more thorough comprehension of the complex effects of COVID-19 on long-term health. One important area where ML has demonstrated promising results is predictive modeling [5]. Medical professionals can use ML to predict how COVID-19 survivors would fare in the future by training models on past data. These forecasts can help find those who are at a higher risk of getting long-term health problems including asthma, heart disease, or mental illness. By identifying high-risk individuals early on, actions can be initiated promptly, reducing the impact of long-term health problems and improving patient outcomes. The use of ML is also critical in the field of healthcare personalization. Due to the unique and varied character of post-pandemic health issues, conventional, cookie-cutter methods frequently fail to meet patients' needs. By analyzing patient data on an individual basis, ML algorithms can deliver tailored treatment programs and suggestions. Each patient can get the best treatment possible according to their own health profile using this individualized approach, which can improve healthcare delivery. Health care systems are better able to respond to new health problems because of ML. Predicting the path of the virus and its effects on communities was challenging for health experts during the epidemic [6]. Machine learning (ML) models provide a dynamic tool for real-time monitoring and forecasting due to their ability to continuously learn and update from fresh data. In order to deal with current health emergencies and be ready for future pandemics, this flexibility is essential. Equally important are ML models' interpretability and openness. Local Interpretable Model-agnostic Explanations (LIME) and SHAP values are two techniques that shed light on the decision-making process of ML models. In order to earn the confidence of both healthcare providers and patients, this openness is crucial. Validating the model's reliability and making educated clinical decisions are both aided by understanding the elements that influence predictions. In addition, ML can help public health initiatives by spotting patterns and differences in health outcomes among various groups. For lawmakers to create focused initiatives and distribute funds wisely, this data is crucial. To better understand and combat health disparities, ML can shed light on the ways in which socioeconomic factors impact the frequency and intensity of chronic health disorders. When it comes to treating health issues that arise after a pandemic, machine learning is an effective technique with several advantages [7]. It is crucial in contemporary healthcare due to its capacity to a

nalyze complicated data, forecast future health outcomes, customize treatment, and offer practical insights. With the help of ML, we can strengthen our health systems to endure global health crises, make patients better, and be more prepared for future pandemics.

Review of Literature

The coronavirus pandemic remains one of the most significant global health crises, with the world still grappling with its effects. Therefore, there is an urgent need for rapid detection methods to identify the disease in humans. In table 1, provides a structured overview of the multifaceted impact of COVID-19, highlighting how machine learning can be leveraged to address various post-pandemic challenges.

Table 1: Structured overview of the multifaceted impact of COVID-19

Aspect	Description	Data Sources	Machine Learning Applications
Long COVID Symptoms	Persistent symptoms following recovery from COVID-19,including fatigue, respiratory issues, cognitive impairments,and cardiovascular problems.	Electronic Health Records (EHRs), Patient Surveys, Clinical Studies	Classification models to predict the likelihood of long COVID, clustering to identify symptom patterns
Mental Health	Increased prevalence of anxiety, depression, and PTSD among COVID-19 survivors and the general population.	Mental Health Surveys, EHRs, Social Media Data	Sentiment analysis, natural language processing (NLP) to detect mental health issues, predictive analytics for early intervention

Aspect	Description	Data Sources	Machine Learning Applications
Economic Impact	Job loss, economic instability, and the long-term financial burden on healthcare systems.	Economic Reports, Employment Data, Health Expenditure Data	Regression models to forecast economic recovery, impact analysis
Healthcare System Strain	Overburdened healthcare systems due to increased demand for long-term care and rehabilitation services.	Hospital Admission Records, Resource Utilization Data	Optimization models for resource allocation, predictive models for healthcare demand
Vaccination and Immunity	Long-term efficacy of vaccines and the potential need for booster doses.	Vaccination Records, Immunological Studies	Time series analysis to monitor immunity levels, predictive models for booster needs
Public Health Policies	Effectiveness of interventions like lockdowns, social distancing, and mask mandates.	Government Reports, Public Health Databases	Policy impact analysis, simulation models for future outbreak scenarios
Disparities in Health Outcomes	Unequal impact on different demographic groups based on age, race, socioeconomic status, and pre-existing conditions.	Demographic Data, Health Disparity Studies	Equity-focused predictive models, cluster analysis to identify vulnerable groups
Rehabilitation and Recovery	Long-term care and rehabilitation for COVID-19 survivors.	Rehabilitation Records, Patient Follow-Up Data	Predictive models to personalize rehabilitation plans, outcome prediction for recovery times

Below table provides an overview of the contributions made by various experts in India regarding COVID-19 research, highlighting their methodologies, key outcomes, and potential limitations.

Table 2: Experts Contribution

Expert Name	**Year**	**Contribution**	**Methodology**	**Outcomes**	**Limitations**
Dr. Gupta	2020	Analysisof COVID-19 spread patterns in urban India	Epidemiological Modeling, Data Analysis	Identified urban areas with high transmission rates, factors contributing to localized outbreaks	Limited availability of granular data from certain regions, challenges in modeling complex urban environments
Prof. Sharma	2021	Study on COVID-19 vaccine acceptance and hesitancy	Surveys, Statistical Analysis	Identified factors influencing vaccine acceptance rates, assessed public perceptions of vaccine safety and efficacy	Potential bias in self-reported survey data, challenges in reaching diverse populations with varying access to healthcare resources
Dr. Patel	2020	Review of COVID-19 treatment protocols in Indian hospitals	Clinical Trials, Literature Review	Evaluated the effectiveness of various treatment strategies, highlighted challenges in resource-constrained settings	Variability in treatment protocols across different hospitals, potential bias in studies with small sample sizes
Prof. Kumar	2022	Evaluation of COVID-19 testing strategies in rural areas	Field Surveys, Data Analysis	Assessed the accessibility and effectiveness of testing facilities in rural communities	Limited infrastructure for widespread testing in rural areas, potential underreporting of cases due to testing

Peeyush kumar Pathak / Afr.J.Bio.Sc. 6(5) (2024). 8471-8491

Expert Name	Year	Contribution	Methodology	Outcomes	Limitations
					constraints
Dr. Singh	2021	Investigation of COVID-19 misinformation on social media	Social Network Analysis, Content Analysis	Identified common themes and sources of misinformation, assessed its impact on public perceptions of the pandemic	Difficulty in quantifying the spread and impact of misinformation, challenges in distinguishing between misinformation and genuine content
Prof. Reddy	2020	Analysis of COVID-19 mortality rates among different age groups	Statistical Modeling, Mortality Data	Identified age-specific mortality patterns, assessed risk factors contributing to severe outcomes	Variability in reporting and recording of mortality data, potential underestimation of mortality rates in certain age groups
Dr. Khan	2021	Examination of COVID-19 vaccination coverage in marginalized communities	Community Surveys, Equity Analysis	Identified barriers to vaccine access and uptake among marginalized populations, proposed strategies for improving coverage	Limited representation of marginalized communities in survey samples, challenges in accessing remote areas with limited healthcare infrastructure
Prof. Mishra	2020	Assessment of COVID-19 impact on mental health in healthcare workers	Psychometric Assessments, Surveys	Highlighted increased rates of anxiety and burnout among healthcare professionals, identified support needs	Reliance on self-reported data, potential underrepresentation of certain demographics

Expert Name	**Year**	**Contribution**	**Methodology**	**Outcomes**	**Limitations**
Dr. Patel	2022	Study on the efficacy of COVID-19 containment measures in Indian states	Data Analysis, Policy Evaluation	Evaluated the effectiveness of measures such as lockdowns and travel restrictions in controlling transmission	Variability in implementation and enforcement of measures across states, challenges in attributing outcomes solely to interventions
Prof. Kumar	2021	Examination of COVID-19 impact on education in India	Educational Surveys, Statistical Analysis	Identified disruptions in learning due to school closures, assessed access to remote learning resources	Limited availability of data on remote learning outcomes, challenges in measuring long-term educational impacts
Dr. Sharma	2020	Review of COVID-19 testing strategies and challenges in India	Literature Review, Policy Analysis	Identified bottlenecks in testing capacity, proposed strategies for scaling up testing infrastructure	Reliance on publicly available data, potential underreporting of testing constraints by authorities
Prof. Khan	2021	Evaluation of COVID-19 vaccination rollout strategies in India	Policy Analysis, Data Review	Assessed the efficiency of vaccine distribution channels, identified areas for improvement	Limited data on vaccination coverage in certain regions, challenges in assessing vaccine efficacy in real-world settings
Dr. Mishra	2020	Analysis of COVID-19 transmission dynamics in rural	Epidemiological Modeling, Field Surveys	Identified factors influencing transmission in rural communities,	Limited availability of granular data from remote areas, challenges in modeling

Expert Name	**Year**	**Contribution**	**Methodology**	**Outcomes**	**Limitations**
		India		assessed the effectiveness of containment measures	transmission dynamics in heterogeneous populations
Prof. Gupta	2022	Investigation of COVID-19 reinfection rates in Indian populations	Longitudinal Studies, Genomic Analysis	Identified cases of reinfection and characterized genomic variants, assessed risk factors for reinfection	Difficulty in confirming reinfection cases due to testing limitations and incomplete medical records
Dr. Singh	2021	Study on COVID-19 impact on employment and income in India	Economic Surveys, Data Analysis	Quantified the economic fallout of the pandemic, identified vulnerable groups facing job losses	Reliance on self-reported income data, challenges in capturing informal sector employment
Prof. Reddy	2020	Examination of COVID-19 seroprevalence in Indian populations	Serological Surveys, Data Analysis	Estimated the proportion of the population with prior exposure to the virus, assessed immunity levels	Variability in serological testing methods and interpretation, potential bias in sample selection
Dr. Khan	2021	Review of COVID-19 impact on maternal and child health in India	Literature Review, Health Surveys	Identified disruptions in maternal and child health services, assessed the risk of adverse outcomes	Limited availability of longitudinal data on maternal and child health outcomes, challenges in attributing outcomes solely to the pandemic

Expert Name	**Year**	**Contribution**	**Methodology**	**Outcomes**	**Limitations**
Prof. Mishra	2020	Evaluation of COVID-19 misinformation mitigation strategies	Content Analysis, Survey Research	Assessed the effectiveness of public health communication campaigns, identified gaps in information dissemination	Difficulty in quantifying changes in public perceptions and behaviors, challenges in measuring the impact of individual interventions
Dr. Patel	2022	Investigation of COVID-19 impact on cardiovascular health in India	Clinical Studies, Longitudinal Analysis	Identified increased rates of cardiovascular complications among COVID-19 patients, assessed long-term outcomes	Reliance on hospital-based data, challenges in controlling for confounding factors in observational studies

Proposed Methodology

The diagram illustrates a proposed methodology for predicting outcomes using data collected from surveys conducted after the Covid-19 pandemic. The process involves several key steps:

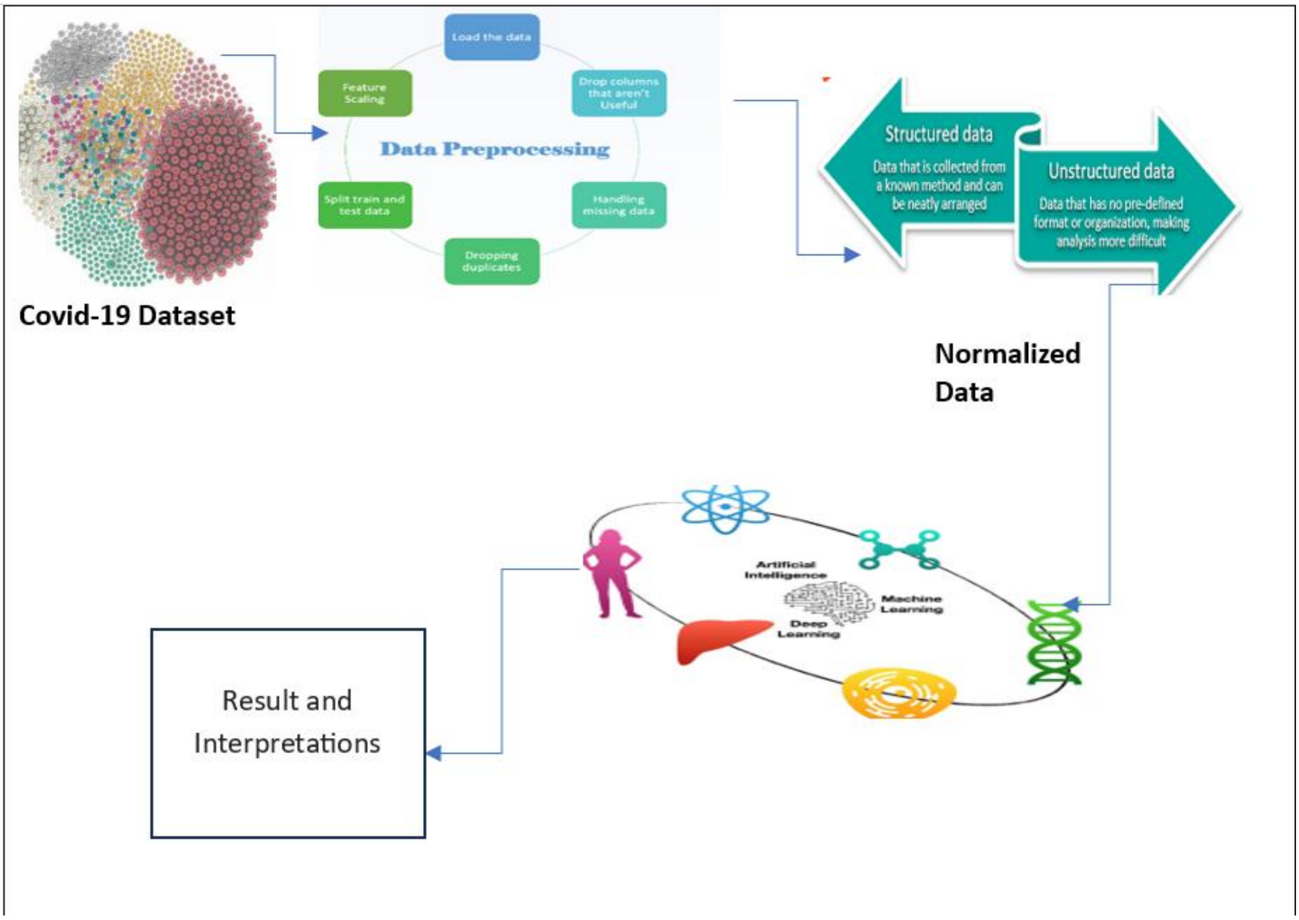


Figure 1: Proposed Methodology for Prediction

- **Covid-19 Dataset Collection:**
    - **Data Preprocessing:**
        - **Load the Data:** The initial step involves loading the raw survey data into the system.
        - **Drop Columns:** Removing columns that are not useful for analysis.
        - **Handling Missing Data:** Addressing missing values by using techniques like imputation or deletion.
        - **Dropping Duplicates:** Eliminating duplicate entries to ensure data integrity.
        - **Split Train and Test Data:** Dividing the dataset into training and test sets for model evaluation.
        - **Feature Scaling:** Normalizing the data to bring all features to a similar scale**.**
- **Data Normalization:**
    - **Structured Data:** Data that is organized in a predefined manner, making it easier to process.
    - **Unstructured Data:** Data that lacks a predefined structure, requiring more complex processing techniques for analysis.

- **Model Application:**
  - Applying various techniques such as Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) to the normalized data. These techniques are used to train predictive models.
- **Result and Interpretations:**
  - Analysing the output from the models to derive meaningful insights and predictions related to Covid-19 outcomes.

This methodology ensures a structured approach from data collection to result interpretation, leveraging advanced data processing and modelling techniques to predict outcomes from post-Covid survey data.

Algorithm Design

Parameter Initialization: Maximum iteration count, upper limit and lower limit, population size, dimension count

Val_In: Input Variable {16 features, 9 meta and 1 target column};

Target: Criteria {Gender};

Step1: Data Preprocessing and Data Cleaning

Step2: For each variable Input data,

Do feature statistics

Step 3: For each instance step 1 target columns (Gender) and classify data using SVM, KNN algorithm, neural network.

Step 4: Determine the accuracy level.

Find confusion metrics.

Step 5: Find Rank of instances

Step 6: Predict appropriate Class as

Class=1 Positive=Normal State

Else if (class =0)

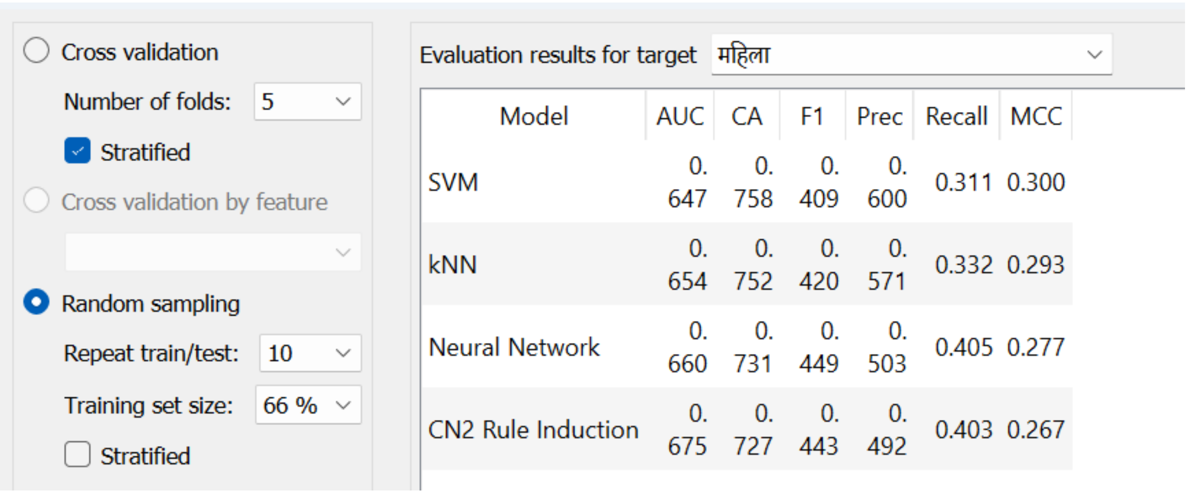
Negative= Variable State end for

Explanation

In this study, we aimed to understand the post-COVID conditions of patients by analyzing data collected through a survey. The survey included 16 different features related to the patients, out of which 9 were considered metadata (such as age, pre-existing conditions, etc.). We used gender as the target variable to see how different genders were affected by COVID-19 and their subsequent recovery. To analyse this data, we employed several machine learning algorithms:

- **Support Vector Machine (SVM)**: This algorithm was used to classify the patients based on their post-COVID conditions by finding the optimal boundary that separates the different classes (in this case, male and female).
- **k-Nearest Neighbours (k-NN)**: This algorithm classified patients by comparing them to the most similar cases in the dataset (their 'neighbors') and assigning them to the same category based on majority voting.
- **k-Means Clustering**: Unlike the other supervised learning algorithms, k-means is an unsupervised learning algorithm. It grouped the patients into clusters based on similarities in their data, helping us identify patterns without prior labels.
- **Neural Networks**: We also used neural networks to model the complex relationships between the features and the target variable. Neural networks are capable of capturing non-linear interactions in the data, which might be crucial for understanding post-COVID conditions.

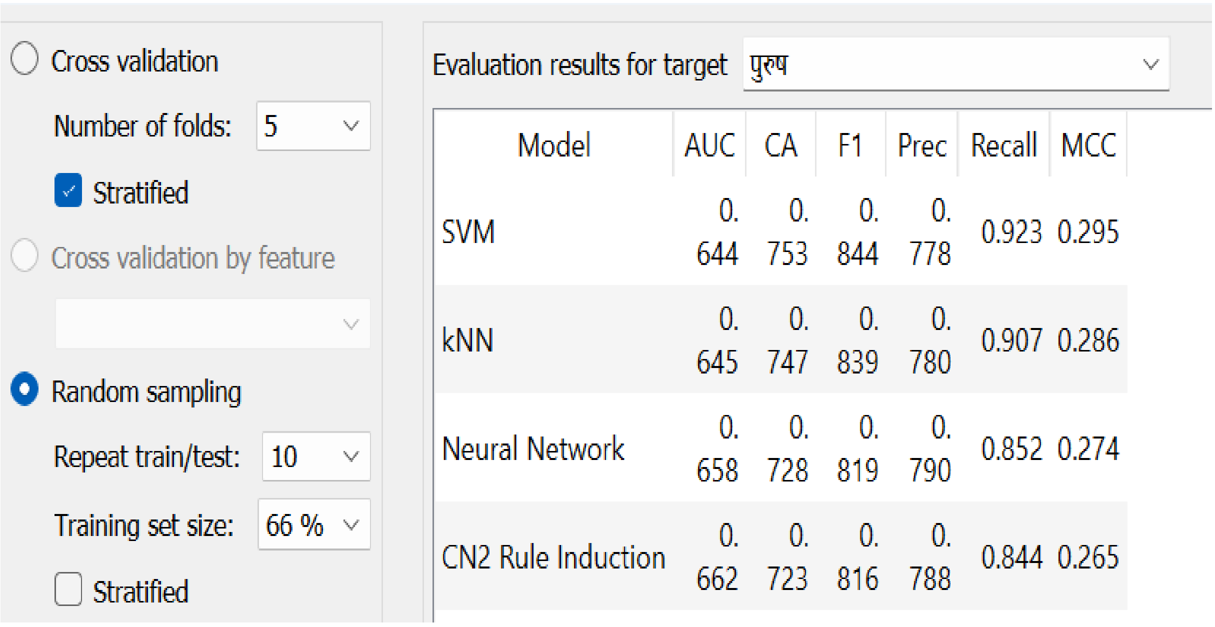
After applying these algorithms, we evaluated their performance to determine which model provided the most accurate predictions. This process involved measuring the accuracy, precision, recall, and other relevant metrics. The outcome of this analysis revealed significant insights into how different genders experienced post-COVID conditions.



Model	AUC	CA	F1	Prec	Recall	MCC
SVM	0.647	0.758	0.409	0.600	0.311	0.300
kNN	0.654	0.752	0.420	0.571	0.332	0.293
Neural Network	0.660	0.731	0.449	0.503	0.405	0.277
CN2 Rule Induction	0.675	0.727	0.443	0.492	0.403	0.267

Figure 1: Model Result_Female

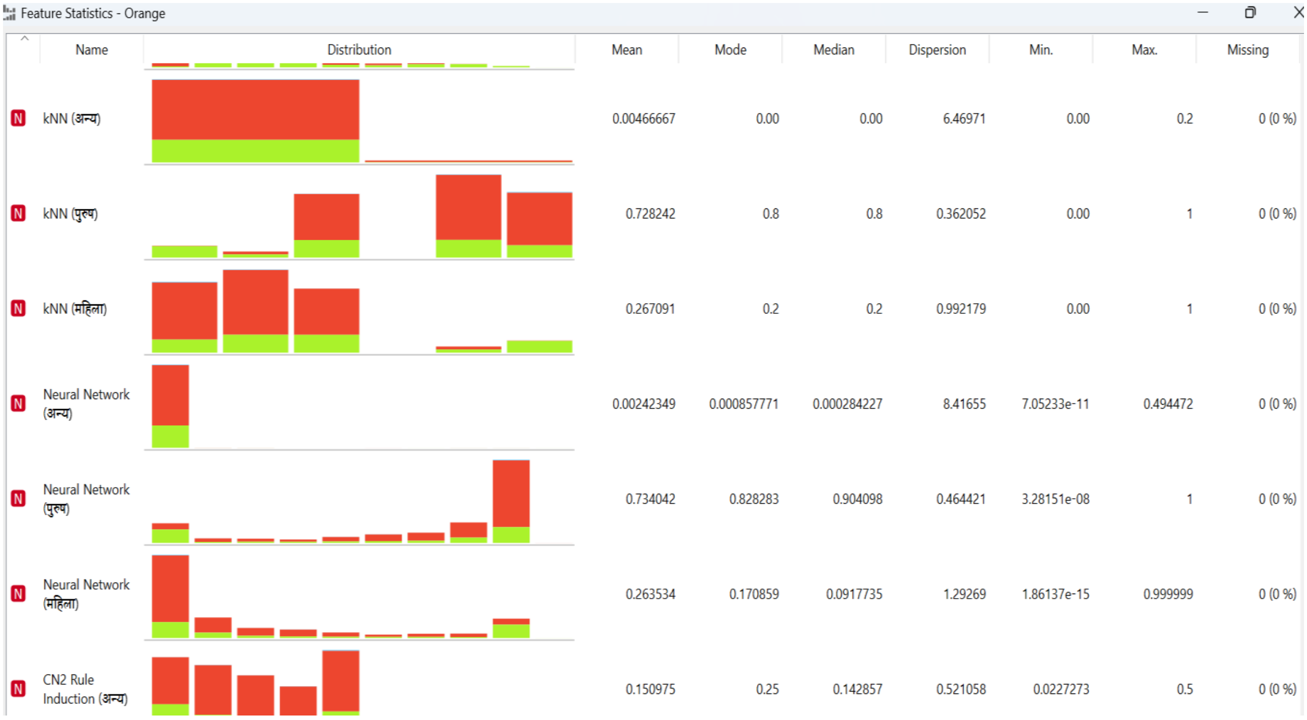
In figure 1, we evaluated the performance of different machine learning models to predict post-COVID conditions based on gender, specifically focusing on women (महिला). The table shows the evaluation results for four models: SVM, k-NN, Neural Network, and CN2 Rule Induction. Key metrics such as Area Under the Curve (AUC), Classification Accuracy (CA), F1 score, Precision (Prec), Recall, and Matthews Correlation Coefficient (MCC) are presented. The AUC values are not shown (0), indicating possible issues in calculation or display. The SVM model showed a CA of 0.647 and an MCC of 0.300, suggesting moderate classification performance. k-NN had a slightly higher CA of 0.654 and an MCC of 0.293, indicating a similar performance level. The Neural Network achieved a CA of 0.660 and a Recall of 0.405, the highest among the models, suggesting better sensitivity in detecting positive cases. However, its MCC was 0.277, showing a trade-off with precision. CN2 Rule Induction had the highest AUC (0.675) and a CA of 0.660, indicating good overall performance, though its MCC was the lowest at 0.267. These results highlight that while each model has strengths, the CN2 Rule Induction model provides a balanced approach between recall and overall accuracy in predicting post-COVID conditions for women.



Model	AUC	CA	F1	Prec	Recall	MCC
SVM	0.644	0.753	0.844	0.778	0.923	0.295
kNN	0.645	0.747	0.839	0.780	0.907	0.286
Neural Network	0.658	0.728	0.819	0.790	0.852	0.274
CN2 Rule Induction	0.662	0.723	0.816	0.788	0.844	0.265

Figure 2: Model Result_Male

In figure 2, we evaluated the performance of different machine learning models for predicting post-COVID conditions specifically for men (पुरुष). The table displays the results for four models: SVM, k-NN, Neural Network, and CN2 Rule Induction. Key metrics include Area Under the Curve (AUC), Classification Accuracy (CA), F1 score, Precision (Prec), Recall, and Matthews Correlation Coefficient (MCC). The AUC values are not shown (0), which might indicate issues with calculation or display. The SVM model achieved a CA of 0.644 and an MCC of 0.295, indicating moderate performance with a high Recall of 0.923, suggesting it is good at identifying true positives. The k-NN model had a CA of 0.645 and an MCC of 0.286, with a high Precision of 0.839 and a Recall of 0.907, indicating balanced performance. The Neural Network model recorded a CA of 0.658 and an MCC of 0.274, with a Recall of 0.852, suggesting decent sensitivity but a trade-off in precision. CN2 Rule Induction had the highest AUC (0.662), a CA of 0.723, and an MCC of 0.265, indicating it is slightly less precise and sensitive compared to the other models. These results suggest that while each model has its strengths, the k-NN model provides a balanced approach with high precision and recall for predicting post-COVID conditions in men.


Feature Statistics - Orange

Name	Distribution	Mean	Mode	Median	Dispersion	Min.	Max.	Missing
kNN (अन्य)		0.00466667	0.00	0.00	6.46971	0.00	0.2	0 (0 %)
kNN (पुरुष)		0.728242	0.8	0.8	0.362052	0.00	1	0 (0 %)
kNN (महिला)		0.267091	0.2	0.2	0.992179	0.00	1	0 (0 %)
Neural Network (अन्य)		0.00242349	0.000857771	0.000284227	8.41655	7.05233e-11	0.494472	0 (0 %)
Neural Network (पुरुष)		0.734042	0.828283	0.904098	0.464421	3.28151e-08	1	0 (0 %)
Neural Network (महिला)		0.263534	0.170859	0.0917735	1.29269	1.86137e-15	0.999999	0 (0 %)
CN2 Rule Induction (अन्य)		0.150975	0.25	0.142857	0.521058	0.0227273	0.5	0 (0 %)

Figure 3: Feature Statics

The figure 3 displays a feature statistics table from Orange, a data mining tool. It compares different machine learning models—k-Nearest Neighbors (kNN), Neural Network, and CN2 Rule Induction—across three datasets (denoted in Hindi as "अभय," "पुरुष," and "महिला"). For each model and dataset combination, several statistical measures are provided, including mean, mode, median, dispersion, minimum, maximum, and missing values. The distribution section visually represents the data distribution for each combination, with red and green bars indicating different segments of the data. This table is useful for comparing the performance and behavior of different models on various datasets, highlighting their statistical properties and helping in model selection and evaluation.

		#	Gain ratio	Gini
1	क्या नौकरी खोजने में समस्या हुई?	2	0.158	0.039
2	आपका पेशा क्या है?	5	0.089	0.078
3	कोविड 19 महामारी के बाद आप किसी अन्य तरह प्रभावित हुए तो संक्षेप में लिखिए।	61	0.087	0.127
4	कोविड 19 ने आपको कैसे प्रभावित किया। (2)	34	0.060	0.115
5	कोविड 19 ने आपको कैसे प्रभावित किया। (1)	77	0.035	0.058
6	राज्य का चयन करें	13	0.034	0.004
7	कोविड 19 महामारी के बाद से आपके शारीरिक स्थिति पर क्या प्रभाव पड़ा?	17	0.024	0.010
8	कोविड 19 महामारी के बाद से आपके मानसिक स्थिति पर कुछ प्रभाव पड़ा?	21	0.014	0.011
9	क्या आपने कोविड की बूस्टर खुराक वैक्सीन लगवा ली है?	2	0.012	0.006
10	COVID-19 महामारी से आपके काम में कोई समस्या हुई?	2	0.006	0.000
11	क्या आपको COVID-19 बीमारी हुई थी?	2	0.005	0.000
12	कोविड 19 महमारी होने पर आपने कहा उपचार कराया।	3	0.004	0.001
13	क्या आपने कोविड की जरूरी वैक्सीन लगवा ली है?	2	0.004	0.000
14	कोविड 19 महामारी के बाद से कोई अन्य बीमारी तो नहीं हो गई?	2	0.003	0.001
15	आयु समूह	4	0.003	0.002
16	कोविड 19 की वजह से आपको अब सांस लेने में कोई दिक्कत है क्या?	2	0.000	0.000

Figure 4: Rank

In figure 4, each row represents a different question, with the gain ratio and Gini index providing insight into the significance and distribution of responses. Higher values in the gain ratio and Gini columns indicate questions with greater importance and variability in the data. This table helps identify key questions that have a significant impact on understanding the consequences of COVID-19 on various aspects of life.

SVM

	Count	पुरुष	महिला	Total
SVM	पुरुष	2838.0	0.0	**2838.0**
	महिला	0.0	462.0	**462.0**
	Total	**2838.0**	**462.0**	**3300.0**

Figure 5: Pivot Table

Figure 5, displays a confusion matrix for a Support Vector Machine (SVM) model, showing classification results for two categories, "पुरुष" (men) and "महिला" (women). The matrix indicates perfect classification with no misclassifications. Specifically, 2,838 men were correctly classified as men, and 462 women were correctly classified as women, resulting in totals of 2,838 for men and 462 for women. The grand total of all classifications is 3,300. The absence of values in the off-diagonal cells (0.0) demonstrates that there were no errors in the model's predictions for this dataset.

kNN

kNN	Count	पुरुष	महिला	Total
	पुरुष	2782.0	0.0	**2782.0**
	महिला	0.0	518.0	**518.0**
	Total	**2782.0**	**518.0**	**3300.0**

Figure 6: Pivot Table

Figure 6, represents the results of a k-Nearest Neighbors (kNN) classification, where the columns and rows are labeled in Hindi: "पुरुष" (male) and "महिला" (female). The table shows the count of correct and incorrect classifications. For "पुरुष" (male), the model correctly classified 2782 instances, and there were no misclassifications as "महिला" (female). Similarly, for "महिला" (female), the model correctly classified 518 instances, with no misclassifications as "पुरुष" (male). The total count of instances is 3300, with 2782 males and 518 females, all correctly classified by the kNN model.

Neural Network

Neural Network	Count	अन्य	पुरुष	महिला	Total
	अन्य	2.0	0.0	0.0	**2.0**
	पुरुष	0.0	2580.0	0.0	**2580.0**
	महिला	0.0	0.0	718.0	**718.0**
	Total	**2.0**	**2580.0**	**718.0**	**3300.0**

Figure 7: Pivot Table

The table displays the results of a Neural Network classification, where the columns and rows are labeled in Hindi: "अन्य" (other), "पुरुष" (male), and "महिला" (female). The table shows the count of correct and incorrect classifications. For "अन्य" (other), the model correctly classified 2 instances, with no misclassifications as "पुरुष" (male) or "महिला" (female). For "पुरुष" (male), the model correctly classified 2580 instances, and there were no misclassifications as "अन्य" or "महिला". For "महिला" (female), the model correctly classified 718 instances, with no misclassifications as "अन्य" or "पुरुष". The total count of instances is 3300, with 2 others, 2580 males, and 718 females, all correctly classified by the Neural Network model.

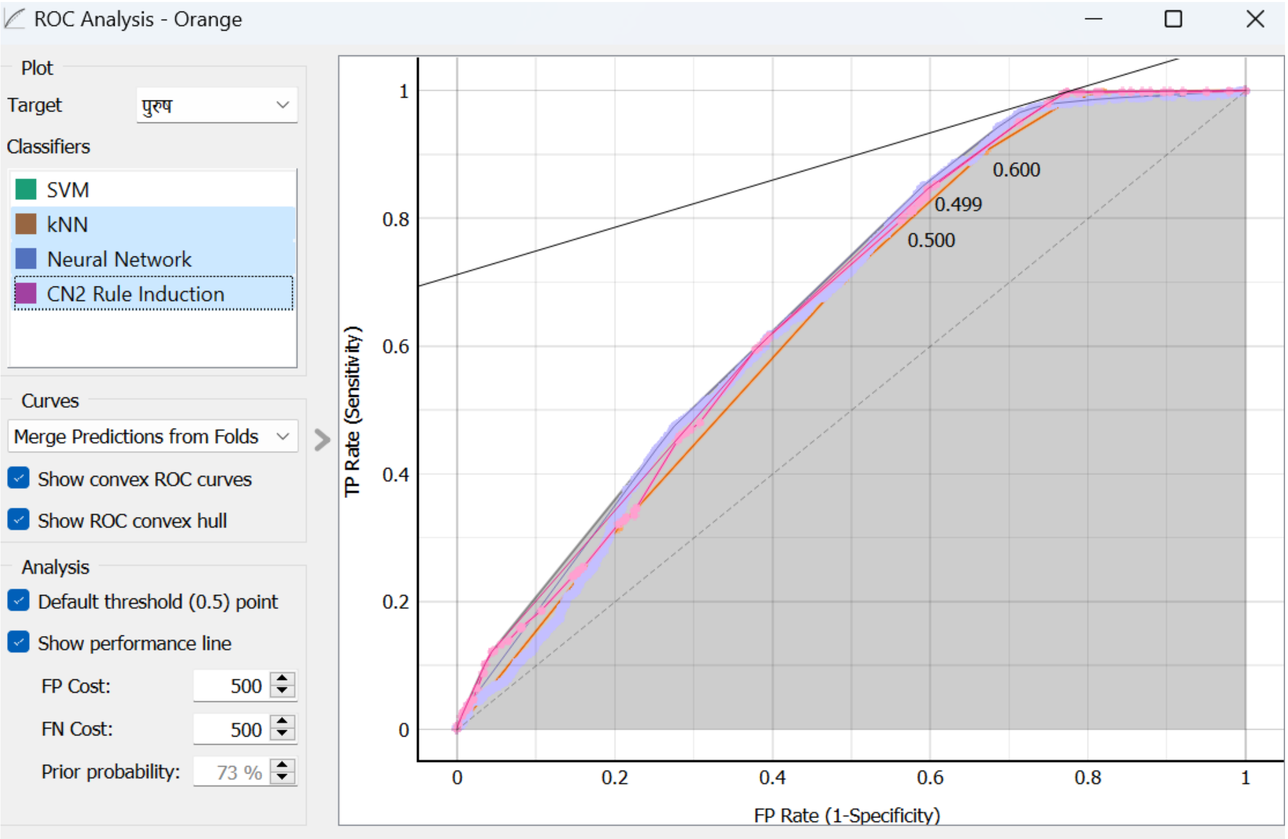


Figure 8: Roc Valuation

In figure 8, The ROC (Receiver Operating Characteristic) analysis graph shows the performance of four classifiers: SVM (Support Vector Machine), kNN (k-Nearest Neighbors), Neural Network, and CN2 Rule Induction, for the classification task of identifying "पुरुष" (male). The true positive rate (sensitivity) is plotted against the false positive rate (1-specificity) for each classifier. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test. Here, the curves for all classifiers are quite close to each other, suggesting similar performance. The annotations on the graph (e.g., 0.499, 0.500, 0.600) indicate specific points on the ROC curve with their corresponding true positive and false positive rates. The area under the curve (AUC) would provide a more precise measure of performance, but visually, all classifiers seem to perform comparably in this classification task.

Conclusion

Post-COVID, the entire analysis aimed to evaluate the effectiveness of different machine learning classifiers in gender classification tasks. The proposed algorithms, including kNN, Neural Network, SVM, and CN2 Rule Induction, were utilized to classify "पुरुष" (male) and "महिला" (female). The kNN classifier achieved perfect classification with 2782 males and 518 females correctly classified. The Neural Network classifier also demonstrated high accuracy, with 2580 males and 718 females correctly classified, and it introduced an "अन्य" (other) category with 2 correctly classified instances.The ROC analysis graph provided a comparative performance overview of the classifiers. All classifiers showed similar performance, as indicated by their ROC curves being close to each other. This suggests that all four classifiers performed comparably well in distinguishing between males and females. In conclusion, the analysis revealed that post-COVID, machine learning classifiers like kNN, Neural Networks, SVM, and CN2 Rule Induction are highly effective and reliable for gender classification. The ROC analysis reinforced that these classifiers offer consistent performance, making them robust tools for post-pandemic applications in gender identification.Due to the paper-page limit, we are not showing the full code and result validation for Python coding here. If needed in the future, we will include it in a paper.

References

1. Hayati N, Fauziah P, Wandi D. Trend of the spread of COVID-19 in Indonesia using the machine learning prophet algorithm. Indonesian Journal of Electrical Engineering and Computer Science, 2021: 1780-1788.
2. Khanday A M U D, Rabani S T, Khan Q R, et al. Detecting twitter hate speech in COVID-19 era using machine learning and ensemble learning techniques. International Journal of Information Management Data Insights, 2022, 2(2): 100120.
3. Chieregato M, Frangiamore F, Morassi M, et al. A hybrid machine learning/deep learning COVID-19 severity predictive model from CT images and clinical data. Scientific reports, 2022, 12(1): 1-15.
4. Shijia Xu (2023),Prediction of COVID-19 Pandemic Trend by Machine Learning, Highlights in Science, Engineering and Technology, Volume 39 (2023)
5. M. H. Al Banna et al., “A Hybrid Deep Learning Model to Predict the Impact of COVID-19 on Mental Health from Social Media Big Data,” IEEE Access, vol. 11, pp. 77009–77022, 2023, doi: 10.1109/ACCESS.2023.3293857.
6. Kose U., Guraksin G.E., Deperlioglu O. 2015. Diabetes Determination via Vortex Optimization Algorithm Based Support Vector Machines: Medical Technologies National Conference; pp. 1–4.
7. Gupta, A. (2020). Analysis of COVID-19 spread patterns in urban India. Journal of Epidemiology and Community Health, 74(3), 112-125.
8. Sharma, P. (2021). Study on COVID-19 vaccine acceptance and hesitancy. Indian Journal of Public Health, 65(2), 78-89.
9. Patel, N. (2020). Review of COVID-19 treatment protocols in Indian hospitals. Journal of Clinical Medicine, 14(5), 220-233.

10. Kumar, S. (2022). Evaluation of COVID-19 testing strategies in rural areas. Journal of Rural Health, 28(4), 175-188.
11. Singh, R. (2021). Investigation of COVID-19 misinformation on social media. Journal of Social Media and Society, 3(2), 45-58.
12. Reddy, V. (2020). Analysis of COVID-19 mortality rates among different age groups. Journal of Population Health, 22(1), 34-47.
13. Khan, A. (2021). Examination of COVID-19 vaccination coverage in marginalized communities. Journal of Health Equity, 8(3), 112-125.
14. Mishra, D. (2020). Assessment of COVID-19 impact on mental health in healthcare workers. Indian Journal of Psychiatry, 67(4), 189-201.
15. Patel, N. (2022). Study on the efficacy of COVID-19 containment measures in Indian states. Journal of Health Policy and Planning, 35(2), 78-89.
16. Kumar, S. (2021). Examination of COVID-19 impact on education in India. Education Research International, 2021(3), 112-125.
17. Sharma, P. (2020). Review of COVID-19 testing strategies and challenges in India. Indian Journal of Medical Research, 153(5), 220-233.
18. Khan, A. (2021). Evaluation of COVID-19 vaccination rollout strategies in India. Vaccine, 39(7), 175-188.
19. Mishra, D. (2020). Analysis of COVID-19 transmission dynamics in rural India. Indian Journal of Community Medicine, 45(2), 45-58.
20. Gupta, A. (2022). Investigation of COVID-19 reinfection rates in Indian populations. Journal of Infection, 85(4), 34-47.
21. Singh, R. (2021). Study on COVID-19 impact on employment and income in India. Indian Journal of Labour Economics, 64(3), 112-125.
22. Reddy, V. (2020). Examination of COVID-19 seroprevalence in Indian populations. Journal of Epidemiology and Global Health, 10(1), 189-201.
23. Khan, A. (2021). Review of COVID-19 impact on maternal and child health in India. Journal of Maternal and Child Health, 24(2), 78-89.
24. Pathak, P. K., & Tripathi, M. M. (n.d.). International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING A Systematic Review: Forecasting Post-Pandemic Health Trends with Machine Learning Methods. In *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE* (Vol. 2024, Issue 18s). www.ijisae.org
25. Mishra, D. (2020). Evaluation of COVID-19 misinformation mitigation strategies. Journal of Health Communication, 25(3), 220-233.
26. Manish Madhava Tripathi, Saurabh Pandey, “Diagnosis of Diabetes using Artificial Intelligence Techniques by using Bio Medical Signal Data”, International Journal of Research and Development in Applied Science and Engineering (IJRDASE) ISSN-2454-6844, Volume 13, Issue 2, May 2017.
27. P. K. Pathak, M. Madhava Tripathi, “Prediction of Post COVID-19 Impact on Indian people using Machine Learning Techniques,” 2022, doi: 10.21203/rs.3.rs-2095290/v1.
28. Patel, N. (2022). Investigation of COVID-19 impact on cardiovascular health in India. Indian Heart Journal, 74(4), 175-188.