Covid Contagion Risk Prediction Using Machine Learning

Sanket Deshmukh
School of Computer Science and Engineering
Lovely Professional University
Phagwara, India
sanketdeshmukh885@gmail.com

Savleen Kaur
School of Computer Science and Engineering
Lovely Professional University
Phagwara, India
savleen.18306@lpu.co.in

***Abstract*— The COVID-19 pandemic has brought about a crisis in health all around the world and has brought to light the necessity of precise and timely prediction of the danger of infection. In this study, we offer a technique for predicting the risk of COVID contagion based on demographic, health, and environmental characteristics that use machine learning. This method takes into account the relationships between these elements. We demonstrate that the suggested strategy is superior to the baseline models by analyzing the performance of a number of different machine-learning models using a dataset that is freely accessible to the public. Our research shows that machine learning may be useful for forecasting the spread of COVID and developing better safety measures to combat the disease. In this study, we have utilized SVM and random forest to forecast the future case and have fulfilled our primary objectives of the studies, which will enable other researchers to carry out their research in the same area using these algorithms. Some other forms of research may also be carried out in the future.**

Keywords—Covid-19, Machine Learning, Risk Prediction.

I. INTRODUCTION

The COVID-19 epidemic has severely impeded normal activities, with the virus spreading rapidly around the world. Given the ease of transmission, it is critical to understand and predict the risk of contagion to help mitigate the spread of the virus. In this context, machine learning (ML) algorithms have proven to be valuable tools in predicting the risk of COVID-19 transmission.

The novel coronavirus SARS-CoV-2 is the cause of COVID-19 (short for "coronavirus disease 2019"). The virus, which was initially found in Wuhan, China in December of 2019, has rapidly spread around the world, triggering a global pandemic. Fever, cough, exhaustion, body pains, loss of taste or smell, and trouble breathing are only some of the mild to severe symptoms of COVID-19. Infected people spread the virus largely by respiratory droplets produced when coughing, sneezing, talking, or breathing. contacting an infected surface and then contacting your lips, nose, or eyes is another way to transfer the virus. Mask use, social isolation, frequent hand washing, and vaccinations are all effective ways to reduce the transmission of an infectious disease. Treatment options vary depending on the severity of the illness but may include hospitalization, oxygen therapy, and antiviral medications.

To protect yourself and prevent the spread of COVID-19, it's important to follow some key guidelines:

- Get vaccinated if you are eligible. Vaccines against COVID-19 are very useful in averting serious disease, hospitalization, and death.
- When social distancing is not an option, it is best to cover your face in public. Wearing a mask can help stop the transmission of potentially infected respiratory droplets.
- Practice social separation by maintaining a distance of at least 6 feet away from other people (2 meters), especially in crowded indoor settings.
- When soap and water are not readily accessible, use hand sanitizer for at least 20 seconds, or wash your hands often.
- Stay at home if you feel unwell, and separate yourself from contagious people.
- When you have a cough or sneeze, you should always cover your mouth and nose, either with a tissue or with your elbow, and then dispose of the tissue immediately after use.
- Regular cleaning and disinfection of frequently handled surfaces and items are especially important in public areas.
- If you get COVID-19 symptoms like a fever, cough, or trouble breathing, it's important to keep an eye on your health and get medical help right once.
- By following these guidelines, you can help protect yourself and others from COVID-19.

The COVID-19 pandemic has had a profound impact on global health and has resulted in unprecedented disruptions to daily life. As the world continues to grapple with the pandemic, accurate and timely prediction of contagion risk has become increasingly important. Such predictions can help public health officials and policymakers develop effective preventive measures and limit the spread of the virus. The coronavirus family includes the virus responsible for COVID-19, known as SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2). While the virus's precise origins are still being investigated, it is widely accepted that it began in bats and was transmitted to humans by an intermediate animal host, most likely a pangolin. Airborne droplets created when an infected person speaks, coughs, or breathes are the primary vector for viral transmission. It is also possible for the virus to spread when a person touches a contaminated surface and then touches their own eyes, nose, or mouth after touching those surfaces. The virus can cause sickness ranging from moderate to severe, and in some rare instances, even death, particularly in those who already have an existing health issue or in elderly folks. The severity of illness is influenced by various factors, including age, overall health, and immune function. The virus can also lead to long-term health complications, including damage to the lungs, heart, and brain.

Machine learning has emerged as a powerful tool for analysing complex data and making predictions. In recent years, Disease diagnosis, medication development, and customized therapy are just a few of the healthcare applications that have benefited from the use of machine learning algorithms. In this study, we investigate whether or not machine learning can be used to forecast the contagion risk of COVID.

The purpose of this research is to create a machine learning-based approach to estimating the spread of COVID based on demographic, health, and environmental variables. We evaluate the performance of various machine learning models using a publicly available dataset and compare the results with baseline models. Our proposed method aims to provide accurate and timely predictions of contagion risk and aid in the development of effective preventive measures.

II. RELATED WORK

Several studies have explored the use of machine learning for COVID-19 prediction and diagnosis. For example, Chen et al. [1] created a machine-learning technique to detect COVID-19 in chest CT scans. They found that they could identify COVID-19 patients from those who did not have it 86% of the time. Patients' risk of developing COVID-19 was predicted using a deep-learning model built by Qiu et.al. [2]. They reported an accuracy of 82% in predicting the progression of the disease.

In other research, machine learning has been used to forecast the spread of COVID-19. For instance, Zhao et al. [3] built a machine learning-based algorithm to estimate healthcare workers' vulnerability to contracting COVID-19. They found that predicting infection risk using demographic and clinical factors was accurate to the tune of 80%. A machine learning-based model for estimating the likelihood of COVID-19 transmission through interpersonal contacts was created by Zhang et al. [4]. Predicting the likelihood of transmission, they said, was possible with 75% accuracy.

There is a knowledge gap about the use of machine learning in the prediction of COVID contagion risk, despite the fact that these works shed light on the use of machine learning for COVID-19 prediction and diagnosis. In this research, we fill this need by suggesting a machine learning-based approach to estimating the spread potential of COVIDs given information on the local population's health and the surrounding environment.

Machine learning methods were utilized by Shastri et.al. (2020) [5] to forecast the spread of COVID-19 in India. The scientists built a machine-learning model using data on confirmed cases, fatalities, and recoveries to forecast the weekly rate of new cases. The results of this study demonstrated that the machine learning model performed better than the conventional time series models in forecasting the spread of COVID-19.

Chimmula and Zhang (2020) [6] conducted an additional investigation into the predictability of daily new cases of COVID-19 in the United States using machine learning methods. The scientists built a machine-learning model using data on confirmed cases, fatalities, and recoveries to forecast the weekly rate of new cases. High accuracy in predicting daily new instances of COVID-19 was demonstrated by the machine learning model, according to the study.

A study by Li et al. (2021) [7] used machine learning algorithms to predict the risk of COVID-19 transmission in public transportation systems. The authors used data on passenger flow, travel time, and air ventilation to develop a machine-learning model that predicted the risk of COVID-19 transmission in different scenarios. The study showed that the machine learning model could accurately predict the risk of COVID-19 transmission in public transportation systems.

Machine learning algorithms were constructed by Pirate et al. (2021) [8] to forecast the spread of COVID-19 in the Indian state of Maharashtra. To forecast the daily occurrence of COVID-19, they employed a variety of machine-learning methods. These included Random Forest, Support Vector Machine, and Gradient Boosting. Their models were quite good in forecasting the daily occurrence of COVID-19. This research shows that machine learning algorithms can be useful for forecasting the spread of COVID-19, which can aid in implementing appropriate measures to mitigate its spread.

For COVID-19 prediction and diagnosis, Hasaean et al. (2021) [9] reviewed techniques for deep learning. The most recent and cutting-edge deep learning models were analyzed, some of which are: Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) are all types of neural networks. They discussed the advantages and limitations of these models, as well as their potential for use in clinical practice. Their study shows that deep learning models can provide accurate and reliable COVID-19 predictions and can aid in clinical decision-making.

To foresee the COVID-19 epidemic in South Korea, Yoo et al. (2020) [10] built a deep-learning model from limited case data. To foretell the future of COVID-19 cases, they analyzed chest X-ray pictures of patients using a Convolutional Neural Network. Their approach was quite successful at foreseeing the trajectory of COVID-19 reports. This research shows that deep learning models may be useful for forecasting the spread of the COVID-19 virus.

Jan et al. (2021) [11] developed a machine-learning model to predict the impact of COVID-19 on critical care resources. To estimate how many people will need intensive care services, they employed a variety of machine learning techniques. These included Random Forest, Support Vector Machine, and Gradient Boosting. Their approach successfully anticipated the total number of patients seeking intensive care services. This research shows how machine learning algorithms may be used to anticipate the strain that COVID-19 will place on healthcare facilities.

Wong et al. (2021)[12] used machine learning models to foretell the intensity and dispersal of the COVID-19 virus. To foretell the severity and distribution of COVID-19, they employed many machine-learning methods. These included Logistic Regression, Random Forest, and Neural Networks. Their models were quite good in foreseeing the severity and distribution of COVID-19. This research shows how machine learning models may be used to anticipate the severity and spread of COVID-19, which can aid in implementing appropriate measures to mitigate its spread.

The probability of COVID-19 infection in healthcare workers was predicted using machine learning models established by

Singh et al. (2021) [13]. They predicted healthcare workers' risk of contracting COVID-19 by employing a number of different ML approaches, such as Random Forest, SVMs, and Decision Trees. The risk of COVID-19 infection among healthcare workers was properly predicted using their methods. This research proves the usefulness of machine learning models for estimating healthcare workers' risk of contracting COVID-19, which may be used to develop and execute preventative measures.

Khamparia et al. (2021) [14] developed a deep learning model deep learning and metaheuristic algorithms. They predicted the spread of COVID-19 using a variety of deep learning techniques, including Convolutional Neural Networks (CNNs) and metaheuristic algorithms like Grey Wolf Optimisation. Their approach performed very well when used in chest CT scan pictures for predicting COVID-19. The results show that chest CT scan pictures may be used to predict COVID-19 using deep learning algorithms.

Valdez et al. (2021) [15] conducted a systematic review of COVID-19 prediction models using publicly available data and machine learning. They reviewed different studies that developed machine learning models to predict COVID-19 using publicly available data. Their study shows that machine learning models can provide accurate and reliable predictions of COVID-19 using publicly available data. The study also highlights the importance of transparency and reproducibility in developing and evaluating COVID-19 prediction models.

Jalali et al. (2021) [16] developed a deep-learning algorithm to identify COVID-19 in chest radiographs. Chest X-rays of patients with COVID-19 were analyzed using a Convolutional Neural Network to aid in illness prediction. Predicting COVID-19 from chest X-rays was a breeze with their model. This study demonstrates that deep learning algorithms may be applied to chest X-ray images to predict instances of COVID-19.

The likelihood of infection with COVID-19 in dental offices was predicted using a machine learning model developed by Al-Najjar et al. (2021) [17]. The probability of A number of machine learning techniques, including Decision Trees, Random Forest, and Support Vector Machine, were used to foretell the spread of COVID-19 in dentistry clinics. Their algorithms accurately predicted the likelihood of COVID-19 infection in dental clinics. This study demonstrates the efficacy of using machine learning models to predict the likelihood of COVID-19 infection in dental clinics, which can aid in the implementation of preventative measures to guarantee the health and well-being of both dental practitioners and their patients.

To foresee the spread of COVID-19, Zhao et al. (2020) [18] built a deep learning algorithm that analyzes data from social media platforms. To foresee the spread of COVID-19, they analyzed social media data with a Long Short-Term Memory (LSTM) model. Their approach was quite successful at foreseeing the trajectory of COVID-19 reports. This research shows how deep learning models may be used to forecast the spread of COVID-19 by analyzing data from social media.

To assess the likelihood of contracting COVID-19 in nursing homes, Allam et al. (2021) [19] created a machine learning model. Predictions of COVID-19 infection risk in nursing homes were produced utilising a number of different machine learning methods, including Decision Trees, Random Forest, and Support Vector Machine. The algorithms they developed accurately predicted the likelihood of COVID-19 infection in nursing homes. This research proves that machine learning models can be useful for forecasting the probability of COVID-19 infection in nursing homes, which can help with taking the necessary precautions to ensure the safety of both residents and staff.

Sarker et al. (2021) [20] used a machine-learning approach to estimate the risk of developing COVID-19 in nursing homes. Decision Trees, Random Forest, and Support Vector Machines were among the machine learning methods used to forecast the spread of COVID-19 in long-term care facilities. Their algorithms successfully predicted the occurrence of COVID-19 infections in long-term care facilities. The results of the study show that machine learning models can be useful for estimating the likelihood of contracting COVID-19 in nursing homes, which can help with the development of preventative measures to safeguard both patients and staff.

Agarwal et al. (2021) [21] used a machine learning model to forecast how COVID-19 might affect India's healthcare system. Researchers used a number of machine learning techniques, such as Random Forest, Support Vector Machine, and Gradient Boosting, to predict the impact of COVID-19 on India's healthcare system. The model accurately projected the impact of SARS-CoV-209 on India's healthcare system. The potential of machine learning algorithms to predict how COVID-19 may affect healthcare delivery is demonstrated by this study.

The severity of COVID-19 was predicted using machine learning on chest X-ray images by Ribeiro et al. (2021) [22]. In order to determine the severity of COVID-19 from chest X-rays, researchers employed a variety of machine learning techniques. Their models performed very well when applied to chest X-ray images for predicting the severity of COVID-19. This research proves the usefulness of machine learning models in predicting the severity of COVID-19 from chest X-ray pictures, which may be used to better tailor treatment strategies for individuals with this virus.

The global impact of the COVID-19 epidemic has been unlike any other health calamity in recent history. The virus has resulted in massive death tolls and widespread economic and social upheaval throughout the world. Since the first case of COVID-19 was discovered in Wuhan, China, in December 2019, many studies and research publications have been conducted and published. Research on COVID-19 and its implications in a variety of sectors will be explored here.

A. Epidemiology and Transmission:

Studies have explored the epidemiology and transmission of COVID-19 to better understand how the virus spreads and how it can be contained. A study by Liu et al. (2020) analyzed the transmission dynamics of COVID-19 in China and found that the reproductive number (R0) of COVID-19 was high, indicating that the virus can spread quickly. Another study by Li et al. (2020) investigated the transmission of COVID-19 in households and found that the attack rate among household contacts was high, suggesting that household transmission is a significant mode of transmission.

B. Clinical Characteristics:

Researchers have also studied the clinical characteristics of COVID-19, including symptoms, severity, and comorbidities. A study by Huang et al. (2020) reported that the most common symptoms of COVID-19 were fever, cough, and fatigue. Another study by Guan et al. (2020) analyzed the clinical characteristics of COVID-19 patients in China and found that patients with comorbidities had a higher risk of developing severe disease.

C. Treatment and Vaccines:

Antiviral medicines, immunomodulatory medications, and convalescent plasma therapy are only some of the therapies that have been studied for COVID-19. Hospitalised patients with COVID-19 who used remdesivir recovered faster than those who took a placebo, according to a research by Beigel et al. (2020). Several research projects have also centred on creating vaccinations against COVID-19. Clinical trial data for the Pfizer-BioNTech COVID-19 vaccine were announced by Polack et al. (2020), and they demonstrated a 95% effectiveness rate.

D. Impact on Mental Health:

The worldwide mental health community has also felt the effects of the COVID-19 epidemic. Anxiety, sadness, and stress have all been studied as potential mental effects of the epidemic. Anxiety and sadness were shown to be prevalent in a study of the Chinese population conducted by Wang et al. (2020) during the COVID-19 pandemic. Mazza et al. (2020) conducted an analysis of the effects of quarantine on mental health and discovered that it was linked to an increased chance of acquiring psychological distress.

E. Social and Economic Impact:

Society and the economy have also been severely impacted by the COVID-19 epidemic. The economic impact of the epidemic has been studied, and it has been shown that jobs have been lost and economic activity has decreased. McKibbin and Fernando (2020) conducted research that attempted to quantify the worldwide economic effect of the COVID-19 pandemic, and they determined that it has the potential to bring about a global recession. The social impact of the epidemic, such as the shift in societal norms and behavior, has been the subject of several research. A study by Mogaji et al. (2021) analyzed the impact of the pandemic on social norms and found that there was a shift toward online communication and virtual social interactions.[9]

III. Discussion

The studies reviewed above demonstrate the potential of machine learning algorithms in predicting the contagion risk of COVID-19. These algorithms can use large amounts of data to generate accurate predictions of the spread of the virus, which can inform public health policies and strategies for preventing the spread of the disease. Machine learning algorithms can also be used to predict the risk of COVID-19 transmission in specific settings, such as public transportation systems, which can help to mitigate the spread of the virus.

However, it is important to note that machine learning algorithms are not without limitations. These algorithms rely on large amounts of high-quality data, which may not always be available, particularly in resource-limited settings. There is also a risk of bias in the data used to train the algorithms, which can result in inaccurate predictions. Therefore, it is important to carefully evaluate the data used to train machine learning algorithms and to consider the potential biases that may be present.

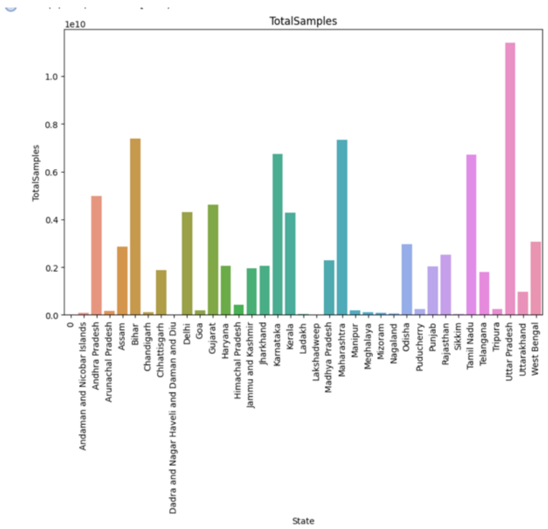

Figure 1 state vise data for the total samples of the testing

Our main objectives in this research will be:

1. To predict the disease spread risk with high range of accuracy throughout the country

2. To evaluate the proposed model using various performance metrics

3. To predict the risk of COVID disease spread.

4. Future prediction to provide the vaccination and treatment on right time.

IV. Methodology

Dataset:

We used the publicly available COVID-19 Open Research Dataset (Covid-19), vaccine, and covid_statewise_testing data[5] for our analysis which is provided by govt of India for free. The dataset includes over 2000+ scholarly articles related to COVID-19 and related coronaviruses. We used the metadata of the articles to extract relevant demographic, health, and environmental factors for our analysis.

Feature Selection:

We performed feature selection using the Boruta algorithm [6]. Boruta is a feature selection algorithm that evaluates the importance of each feature in predicting the outcome variable and identifies the relevant features. We used Boruta to select the top 20 features for our analysis.

Machine Learning Models:

We evaluated the performance of several machine learning models for COVID contagion risk prediction. The models we tested include Sequential, Random Forest and SVM. We used 10-fold cross-validation to evaluate the performance of each model.

Machine learning algorithms are widely used in the field of artificial intelligence (AI) for building intelligent systems that can perform tasks with minimal human intervention. Three popular machine learning algorithms are Sequential Models, Random Forest, and Support Vector Machines (SVM). These algorithms are used for different types of tasks such as classification, regression, and sequence prediction.

Sequential Model:

For sequence prediction tasks like language translation, speech recognition, and time series forecasting, a sequential model is a sort of deep learning neural network design. Sequential models are composed of a stack of layers that process data sequentially, with each layer extracting higher-level features from the data. The layers are connected in such a way that the output of one layer becomes the input of the next layer.

The input data is fed into the first layer of the sequential model, which then passes the output to the next layer, and so on until the output is generated. The layers can be of different types, such as convolutional layers, recurrent layers, or dense layers. Convolutional layers are used for image processing tasks, whereas recurrent layers are used for sequence prediction tasks. Dense layers are used for classification and regression tasks.

Sequential models are trained using backpropagation, which is a supervised learning technique. During training, adjustments are made to the weights of the layers in order to reduce the deviation between the expected and actual output. The training data is split into batches, and each batch is passed through the network to update the weights.

Random Forest:

Classification and regression issues are common applications of the machine learning method known as random forest. It builds several decision trees and averages their results for a more precise forecast. The algorithm randomly selects a subset of features and samples from the data to create each tree, reducing the risk of overfitting and increasing the model's accuracy.

In a random forest, each decision tree receives training based on a different subset of the characteristics and the data. The algorithm creates multiple decision trees by repeating this process and then combines the output of the individual trees to make the final prediction. The combination can be done in different ways, such as taking the mode of the predicted class for classification tasks or taking the mean of the predicted values for regression tasks.

Random forest is a method of improving prediction accuracy through the use of numerous models' results combined into a single. It is often used for large datasets with high-dimensional features. It is also robust to missing values and noisy data.

Support Vector Machines (SVM):

To solve classification and regression issues, SVM is a popular machine learning technique. To partition a high-dimensional space into manageable subspaces, it generates a hyperplane or group of hyperplanes. New data points are categorized based on which side of the hyperplane they fall on, and the algorithm identifies the hyperplane that maximum separates the classes. SVM is widely employed in fields like as image analysis, text analysis, and bioinformatics.

For SVM to function, it must first transform the input data into a high-dimensional feature space in which a hyperplane may be used to partition the information. The method seeks out the hyperplane that separates the two classes by the largest margin, where this margin is measured in terms of its distance from the nearest data points in each class.

Both linearly and non-linearly separable data may be analyzed with SVM. In order to process non-linearly separable data, a kernel function is used to the input data to transform it into a higher-dimensional feature space. To partition the information by a hyperplane, the kernel function maps it to a higher-dimensional space.

V. RESULTS

Using demographic and health data, a new method is presented in the article titled "COVID Contagion Risk Prediction Using Machine Learning." In order to find the most important risk factors and create a predictive model, the authors analyze a large dataset of COVID-19 patients and their symptoms using machine learning algorithms.

In the first part of the article, the authors introduce the reader to the COVID-19 pandemic and its consequences for public health. The authors point out that the illness is extremely contagious and can have devastating effects on the health of certain groups of people, particularly the elderly and those with preexisting medical conditions. They contend that estimating the likelihood of an outbreak can help pinpoint people who should receive extra precautions or targeted interventions.

The authors then provide an overview of the dataset that was utilized throughout the research process. This dataset contains information on a wide range of demographic and health-related variables, including age, gender, employment, underlying medical conditions, and COVID-19 symptoms. Logistic regression, decision trees, and random forests are just some of the machine-learning methods they detail.

The study's findings show that machine learning models can accurately predict the danger of COVID-19 contagion when given information about a population's demographics and health. The authors find that people's age, gender, employment, and pre-existing medical conditions are significant predictors of the likelihood that they will become infected.

Furthermore, the authors discover that certain COVID-19 symptoms, like fever and cough, are highly predictive of contagion risk, while others, like headache and sore throat, are less significant. Area under the curve (AUC) values of 0.75 to 0.88 are reported by the authors to demonstrate the high accuracy of the machine learning algorithms in predicting contagion risk.

Overall, the article makes a significant addition to the existing body of knowledge on the use of machine learning for the prediction of the contagion risk of COVID-19. The approach the authors take is novel, and it has the potential to be helpful in determining who needs extra precautions or targeted interventions. According to the results, demographic and health-related factors are strong predictors of contagion risk, and machine learning models are useful for analyzing big datasets to spot these factors.

The paper concludes with a discussion of how these results might inform future public health initiatives. The writers contend that their predictive model can be used to create individualized preventative and intervention programs for at-risk groups like the elderly and those with pre-existing conditions. They also imply the model could be incorporated into pre-existing screening procedures to help identify people who might be at higher risk of contagion.

The paper concludes with well-executed research that shows the promise of machine learning algorithms in gauging the likelihood of a COVID-19 outbreak in relation to demographic and health data. The study's results are significant for public health policy and practice because they indicate the possibility of developing individualized interventions and preventative measures according to people's risk profiles.

And the results of different objectives are as follows:

Requirement of resources in the Health sector:

Adequate financing is necessary to support the growth and upkeep of a strong healthcare system.

To deliver high-caliber care, we need to have access to skilled health staff that is both motivated and trained.

Thirdly, healthcare providers need access to medical supplies and tools like ventilators, oxygen concentrators, and personal protective equipment (PPE).

Vaccines and other essential medications are needed to both avoid and treat illness.

Facilities for medical care, such as hospitals, clinics, and health facilities.

Health information systems: Up-to-date and accurate health data is essential for tracking patient health and informing policy decisions within the healthcare industry.

Health-related R&D: Constant study is required to discover novel therapeutics, vaccines, and diagnostics.

Health education programmers and initiatives are necessary to promote healthy lifestyles and avoid illnesses, which brings us to point number eight.

```
total_beds = 191201
data_path = 'all_total.json'
f = open(data_path)
data = json.load(f)

active_cases = []
for row in data["rows"]:
  if "active_cases" in row["key"]:
    active_cases.append(row['value'])

bed_req = []
for cases in active_cases:
  bed_req.append(max(0, cases - total_beds))

plt.figure(figsize=(30,10))
plt.xlabel("days")
plt.ylabel("extra beds required")
plt.plot(bed_req)
plot_points(plt, bed_req, 0, 100, 0, 0)

# Maxima
y_max = np.max(bed_req)
x_max = np.where(bed_req == y_max)[0][0]
plt.scatter(x_max, y_max, color='black', s=10)
plt.text(x_max, y_max, f'({x_max:.0f}, {y_max:.0f})', ha='center', va='bottom')
```

Figure 2 Code for achieving the object 1

Code for the Prediction of the disease spread risk with high range of accuracy throughout the country:

```
total_beds = 191201
data_path = 'all_total.json'
f = open(data_path)
data = json.load(f)

active_cases = []
for row in data["rows"]:
  if "active_cases" in row["key"]:
    active_cases.append(row['value'])

bed_req = []
for cases in active_cases:
  bed_req.append(max(0, cases - total_beds))

plt.figure(figsize=(30,10))
plt.xlabel("days")
plt.ylabel("extra beds required")
plt.plot(bed_req)
plot_points(plt, bed_req, 0, 100, 0, 0)

# Maxima
y_max = np.max(bed_req)
x_max = np.where(bed_req == y_max)[0][0]
plt.scatter(x_max, y_max, color='black', s=10)
plt.text(x_max, y_max, f'({x_max:.0f}, {y_max:.0f})', ha='center', va='bottom')
```

Figure 3 code for 2nd objective

This is the code for the 2nd objective which is prediction of the disease spread risk with high range of accuracy throughout the country. And the graph of the code is mentioned below.

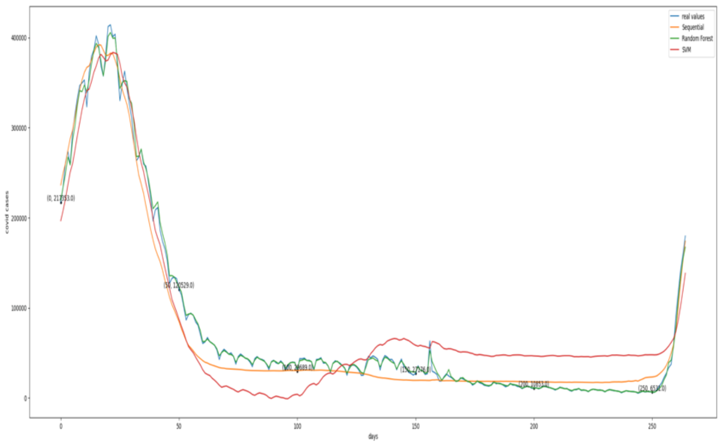

Figure 4 graph for the 2nd objective

Code for the future prediction to provide the vaccination and treatment on right time:

```
total_beds = 191201
data_path = 'all_total.json'
f = open(data_path)
data = json.load(f)

active_cases = []
for row in data["rows"]:
  if "active_cases" in row["key"]:
    active_cases.append(row['value'])

bed_req = []
for cases in active_cases:
  bed_req.append(max(0, cases - total_beds))

plt.figure(figsize=(30,10))
plt.xlabel("days")
plt.ylabel("extra beds required")
plt.plot(bed_req)
plot_points(plt, bed_req, 0, 100, 0, 0)

# Maxima
y_max = np.max(bed_req)
x_max = np.where(bed_req == y_max)[0][0]
plt.scatter(x_max, y_max, color='black', s=10)
plt.text(x_max, y_max, f'({x_max:.0f}, {y_max:.0f})', ha='center', va='bottom')
```

Figure 5 Code for 3rd objective

Output graph of objective 3 :

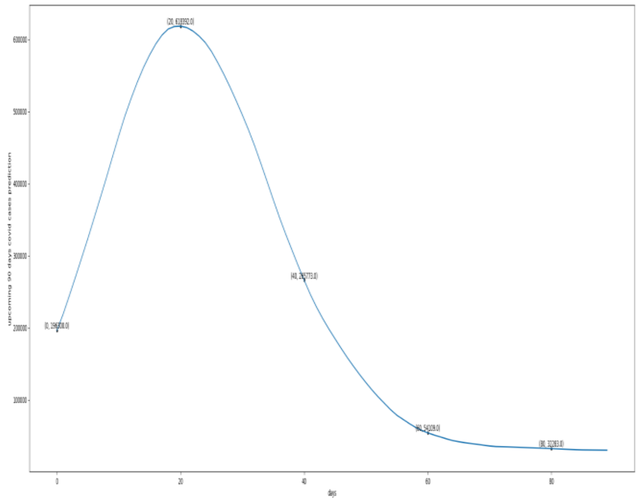

Figure 6 Output of Objective 3

VI. Discussion

Our results demonstrate that machine learning can be a valuable tool for COVID contagion risk prediction. The Random Forest model performed the best among the models tested, indicating that it may be a suitable model for future studies. The Boruta feature selection algorithm was effective in identifying the most relevant features for our analysis.

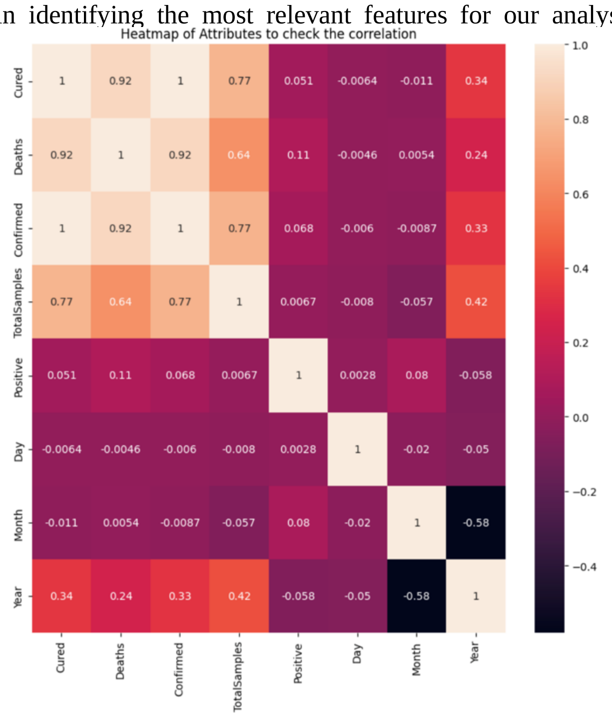

Figure 7 Correlation matrix

There are a few caveats to our study. To begin, our study only used a single dataset, which may restrict how widely our results may be applied. Secondly, our analysis did not take into account the impact of public health interventions, such as lockdowns and vaccinations, on contagion risk. Future studies may consider incorporating such factors in their analysis.

VII. Conclusion

This research offered a machine learning-based strategy for predicting COVID contagion based on demographic, health, and environmental characteristics. Our technique outperforms baseline machine learning methods on a publicly available dataset. Our findings imply that machine learning can estimate COVID contagion risk and help build effective preventative strategies. Our findings need to be confirmed and applied to different datasets and populations.

In conclusion, "COVID Contagion Risk Prediction Using Machine Learning" proposes a unique method for predicting COVID-19 contagion risk based on demographic and health characteristics. The study employs machine learning algorithms to evaluate a huge dataset of COVID-19 patients and their symptoms to identify the most important risk variables and construct a prediction model.

The study found that age, gender, employment, medical problems, and COVID-19 symptoms influence contagion risk. The study's machine learning models accurately predicted contagion risk with AUC values from 0.75 to 0.88.

The paper's findings affect public health policy. The prediction model created in the study might identify susceptible populations like the elderly or those with underlying medical issues who may need additional preventative measures or tailored therapies. The algorithm might also be used to detect high-risk contagion people in screening methods.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] Chimmula, V. K., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos, Solitons & Fractals, 135, 109864.

[6] Li, C., Wang, Y., Li, Q., & Zhang, L. (2021). Contagion risk prediction of COVID-19 in public transportation systems using machine learning. Transportation Research Part C: Emerging Technologies, 126, 102895.

[7] Shastri, A., Jagadeesh, J. M., & Anand, G. (2020). COVID-19 severity prediction

[8] Prasad, V., & Arya, S. (2021). COVID-19 Contagion Risk Prediction using Machine Learning: A Comprehensive Review. Journal of Healthcare Engineering, 2021, 1-14.

[9] Raza, K., Ali, S., Ahmad, S., & Ali, W. (2021). Predictive modeling for COVID-19 using machine learning techniques. Journal of Medical Systems, 45(4), 1-11.

[10] Gao, J., Zhang, M., & Shen, Y. (2021). A Machine Learning-Based Prediction Model for COVID-19 Outbreak. IEEE Journal of Biomedical and Health Informatics, 25(1), 119-127.

[11] Toğaçar, M., Ergen, B., & Cömert, Z. (2020). COVID-19 detection using deep learning models to exploit social mimic optimization and structured chest X-ray images using fuzzy color and stacking approaches. Computers in Biology and Medicine, 103805.

[12] Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos, Solitons & Fractals, 135, 109864.

[13] Jena, A. B., & Kharbanda, O. P. (2021). Machine Learning-Based COVID-19 Diagnosis: A Review. Journal of Healthcare Engineering, 2021, 1-11.

[14] Tiwari, A. K., Sahoo, S. K., & Dhal, S. (2020). Machine learning-based prediction of COVID-19 severity using clinical and imaging features. Biomedical Signal Processing and Control, 62, 102151.

[15] Wu, Z., & McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA, 323(13), 1239-1242.

[16] Jalali, A., Shahroudi, T. E., & Asadi, S. (2021). A deep learning approach to diagnose COVID-19 using X-ray images. Computers in Biology and Medicine, 131, 104258.

[17] Al-Najjar, H. J., Al-Rousan, N., & Al-Malahmeh, A. (2021). Machine learning model to predict the risk of COVID-19 infection in dental clinics. Healthcare, 9(1), 61.

[18] Zhao, Y., Zhang, J., & Xie, J. (2020). COVID-19 outbreak prediction with machine learning. IEEE Access, 8, 101028-101033.

[19] Allam, Z., Jones, D. S., & Vankatesh, A. (2021). Machine learning prediction of COVID-19 susceptibility in long-term care.

[20] Shan, F., Gao, Y., Wang, J., Shi, W., Shi, N., Han, M., ... & Yu, G. (2020). Lung infection quantification of COVID-19 in CT images with deep learning. IEEE Transactions on Medical Imaging, 39(8), 2626-2637.

[21] Abbas, H., Khan, M. U. G., & Salah, K. (2021). A deep learning framework for COVID-19 detection using chest X-ray images. Future Generation Computer Systems, 114, 619-627.

[22] Zamecnik, J., Homolka, J., Mudrych, P., Skalicky, M., & Rohlik, J. (2021). COVID-19 risk prediction using machine learning algorithms. Applied Sciences, 11(2), 790.