International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

Optimal Visual Predictive Modelling on Covid-19 Zones

A. Geetha Devi [1], Chandra Sekhar Koppireddy[2], V. Dilip Kumar[3], Dr. G. S. N Murthy[4], D. Eswara Chaitanya[5], Lakshmi Ramani Burra[6]

Submitted: 26/01/2024 **Revised:** 04/03/2024 **Accepted:** 12/03/2024

Abstract*:* The COVID-19 pandemic has sparked an unprecedented global crisis, compelling nations to devise effective disease containment and mitigation strategies. Zoning, a critical aspect of public health management, categorizes regions based on COVID-19 severity levels, allowing targeted interventions. In this study, we propose an innovative predictive modelling approach leveraging advanced machine learning algorithms to anticipate and classify regions in India into red, orange, and green zones, reflecting varying levels of COVID-19 threat. Drawing from a vast dataset encompassing diverse socio-economic, demographic, and epidemiological variables. We develop predictive models using K-means clustering and decision tree classification techniques. The paramount significance of predicting COVID-19 zones is a key challenge here. By scrutinizing multivariate data, we identify significant factors influencing the spread and severity of the virus across different regions. This informed approach helps optimize resource allocation and public health interventions, allowing for a proactive response to potential outbreaks. The findings demonstrate the potential of machine learning in augmenting traditional epidemiological methods, aiding policymakers in making informed decisions. By continuously updating our predictive models with the latest data, we enable a dynamic and flexible zoning strategy, aligning with the current state of the pandemic. The projected methodology provides a foundation for developing robust decision-support systems, assisting healthcare authorities and policymakers in effectively navigating the challenges posed by COVID-19. The projected approach advocates for a holistic approach, encompassing healthcare infrastructure improvement and behavioural interventions to manage the pandemic effectively. The visual showing of affected areas with customized colours would report the information when the user interacts with the map.

Keywords: *COVID-19, Zoning, Predictive Modeling, Machine Learning, Public Health Management, Epidemiology, K-means Clustering, Decision Tree Classification, Risk Assessment, and Optimal Predictive Modeling.*

1. Introduction

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has rapidly escalated into a global crisis, challenging healthcare systems, economies, and daily life. The highly contagious nature of the virus and the severity of its impact on public health have necessitated extraordinary measures to contain its spread. Countries worldwide have grappled with varying infection rates and surges, prompting the need for nuanced strategies to mitigate the virus's impact effectively. In India, the battle against COVID-19 has been a multifaceted effort, encompassing testing, contact tracing, quarantine protocols, treatment strategies, and, most importantly, zoning of regions based on the prevalence of the virus. Zoning, classifying regions into red, orange, and green zones, has emerged as a vital approach to stratifying risk levels and tailoring responses accordingly. Red Zones signify areas with a high caseload, requiring stringent containment measures. Orange Zones have a moderate caseload, and Green Zones are relatively safer regions.

This paper focuses on predicting these zones using advanced machine-learning techniques, aiming to enhance the efficiency and accuracy of zone classification. The timely and accurate identification of Red, Orange, and Green Zones is critical for guiding policymakers and public-health authorities in implementing targeted containment strategies. By employing predictive modelling, we intend to develop a robust and adaptable algorithm that incorporates various factors influencing the zone classification, including infection rates, population density, healthcare infrastructure, socio-economic conditions, and mobility patterns.

The pandemic's impact extends beyond immediate health concerns, affecting livelihoods, economies, and mental well-being. Understanding the dynamics of the virus's spread and predicting zone classifications is imperative for strategically deploying resources, focusing healthcare efforts, and safeguarding communities. This research aspires to pay to the ongoing efforts in managing the COVID-19 crisis by providing a data-driven,

1 PVP Siddhartha Institute of Technology, Vijayawada. - INDIA
ORCID ID : 0000-0002-4092-2486
2 Pragati Engineering College - INDIA
ORCID ID : 0000-0002-1697-5289
3 SRKR Engineering College - INDIA
ORCID ID : 0000-0002-5982-8983
4Aditya College of Engineering– INDIA
ORCID ID : 0000-0003-2967-9347
5R.V.R. & J.C.College of Engineering -INDIA
ORCID ID : 0000-0002-5201-2228
6 Koneru Lakshmaiah Education Foundation – INDIA
ORCID ID : 0000-0002-2969-6262
ORCID ID : 0000-0001-9919-4180
** Corresponding Author Email: praveenlurur@email.com*

predictive tool that aids in effective decision-making and contributes to a more resilient public health response in India.

2. Related Work

There were few studies on this recent exploration of COVID-19 disease and its counter effects.

A) Epidemiological Models for COVID-19 Prediction:

Epidemiological models are crucial in understanding and predicting the spread of infectious diseases within a population. [1] The SEIR (Susceptible-Exposed-Infectious-Removed) and SIR (Susceptible-Infectious-Removed) models are widely used in epidemiology. [19] During the COVID-19 pandemic, researchers have extensively employed these models to anticipate disease progression and inform public health strategies.

B) SEIR and SIR Models:

The SEIR model divides the population into compartments: susceptible (S), exposed (E), infectious (I), and removed (R) individuals. It simulates the transition of individuals between these compartments based on parameters like transmission rate and incubation period. [20]

The SIR model, on the other hand, doesn't consider an exposed state; individuals move directly from susceptible to infectious before eventually being removed (through recovery or other outcomes).

C) Parameters and Dynamics:

These models describe disease dynamics using parameters like transmission rate (β), contact rate, recovery rate (γ), and incubation period. Transmission rate is a crucial parameter that determines the speed at which individuals move from being susceptible to infectious.

The models also consider population size, initial infections, and individual interaction.

D) Prediction and Scenario Analysis:

Researchers can estimate parameters and predict future infection trends by fitting these models to real-world data. These predictions are essential for planning healthcare resources, implementing public health measures, and assessing the potential impact of interventions.

Scenario analysis involves altering parameters to simulate the effects of various control measures such as social distancing, lockdowns, or vaccination campaigns.

2.1 Limitations and Challenges:

These models assume homogeneous mixing, meaning everyone in the population has an equal chance of coming into contact with everyone else, which oversimplifies the real-world scenario. Factors like demographics, geographic distribution, and individual behaviours are often not fully considered in traditional.

Table 1. Existing Studies on COVID-19 Zones

Author	References	Purpose / Aim	Key Results
Swaraj, Aman, et al.	Implementation of stacking-based ARIMA model for prediction of COVID-19 cases in India.[2]	Enhance forecasting accuracy by combining multiple ARIMA models.	Accurate prediction of COVID-19 spread in India.
Das, Arghya, et al.	COVID-19: Analytic results for a modified SEIR model and comparison of different intervention strategies. [3]	Analytic results for a modified SEIR model and compare various intervention strategies for combating COVID-19.	Include insights into the effectiveness of different intervention strategies in controlling the spread of COVID-19 based on the modified SEIR model.
Alazab, Moutaz, et al.	COVID-19 prediction and detection using deep learning [4]	Explore the application of deep learning techniques for predicting and detecting COVID-19 cases.	Include the development and evaluation of deep learning models for COVID-19 prediction and detection, demonstrating their effectiveness compared to traditional methods.
Punn, Narinder Singh, Sanjay Kumar Sonbhadra, and Sonali Agarwal.	COVID-19 epidemic analysis using machine learning and deep learning algorithms. [6]	To gain insights into the virus's spread, predict its trajectory, and identify factors influencing its transmission and impact.	Analyzing various aspects of the COVID-19 epidemic, such as prediction, detection, and understanding its dynamics.
Khan, Farhan Mohammad, et al.	Projecting the criticality of COVID-19 transmission in India using GIS and machine learning methods. [9]	The objective is to understand spatial spread dynamics and identify high-risk areas.	Involve insights into the spatial patterns of COVID-19 transmission in India, potentially including identifying high-risk regions & analyzing transmission dynamics.

Reema, Gunti, et al.	COVID-19 EDA analysis and prediction using SIR and SEIR models. [12]	The goal is to understand the disease dynamics and inform public health interventions.	The SIR and SEIR models could forecast future infection rates, recovery rates, and other epidemiological parameters.
Khan, Farhan Mohammad, et al.	Projecting the criticality of COVID-19 transmission in India using GIS and machine learning methods. [13]	By employing Geographic Information Systems (GIS) and machine learning methods, it is likely to assess the severity and potential spread of the virus to inform public health strategies and interventions.	Encompass spatial analysis of high-risk areas, identification of factors contributing to transmission severity, and potentially predictive models for future outbreak scenarios, aiding in effective planning and response strategies.
Saxena, Rahul, Mahipal Jadeja, and Vikrant Bhateja.	Propagation analysis of COVID-19: a SIR model-based investigation of the pandemic. [19]	To conduct a propagation analysis of COVID-19 using a SIR (Susceptible-Infectious-Recovered) model	The estimation of key epidemiological parameters such as transmission rates and recovery rates.

SEIR and SIR models. Understanding and refining these models is crucial for effective pandemic management. Future research may focus on integrating more complex models that consider spatial heterogeneity, behavioural aspects, and the impact of public health interventions to provide a more accurate representation of disease spread and aid in policy formulation.

3. Methodology

The study presents a comprehensive methodology for predicting and categorizing regions in India into red, orange, and green zones, reflecting diverse levels of COVID-19 threat, utilizing advanced machine learning techniques. [7, 8, 14, 15] Beginning with data collection, various socio-economic, demographic, and epidemiological datasets are curated and preprocessed, followed by a meticulous feature selection process to identify influential variables. Exploratory data analysis unveils patterns and relationships within the data, setting the stage for model development. The models are trained and validated to effectively categorize regions by leveraging K-means clustering and decision tree classification. Post-model development, the methodology focuses on result interpretation, identifying key contributing factors to zone classification, and aiding in formulating evidence-based policy decisions. Continuous model updating ensures adaptability to the evolving pandemic scenario while developing a decision support system offers a practical interface for healthcare authorities and policymakers.

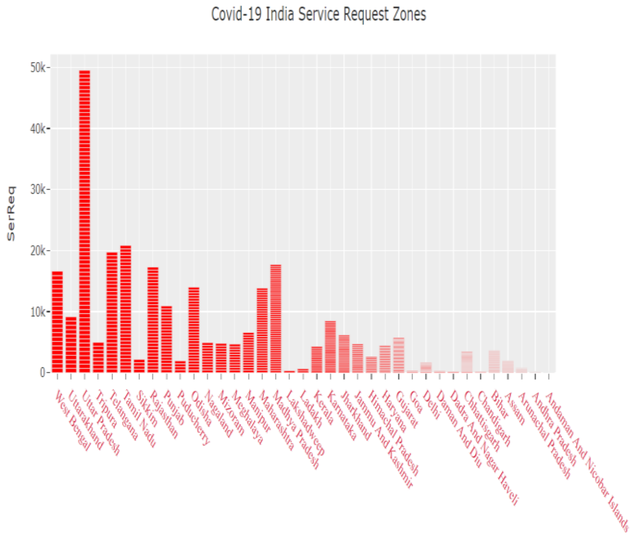

Fig 1: Impact of COVID-19 on States vs. Service Requests

The study's outcomes highlight the probable of machine learning in augmenting traditional epidemiological approaches and offer a robust foundation for dynamic zoning strategies in reaction to the COVID-19 pandemic.

This methodology integrates a multi-stage approach, beginning with the assembly and refinement of diverse datasets, followed by feature engineering and exploratory analysis to comprehend the intricacies of the data and the proposed framework is depicted in Figure 2. Through the application of machine learning models, specifically K-means clustering for regional categorization and decision tree classification for zone prediction, the study systematically extracts insights into COVID-19 severity across Indian regions. [5] By effectively amalgamating data-driven insights and machine learning, this methodology offers a promising framework for proactive policy formulation and strategic resource allocation, empowering authorities to respond to the challenges posed by the COVID-19 crisis.

A) Data Collection and Preprocessing:

We initiated this study by gathering a comprehensive dataset related to the COVID-19 pandemic in India. This dataset encompassed a wide array of information, including daily case counts, testing rates, hospital bed capacities, demographic data, geographical locations, and other relevant socio-economic indicators. The sources for this data included government health departments, reputable international health organizations, and publicly available repositories. The collected data was then subjected to rigorous cleaning and integration processes. This involved handling missing values, removing duplicates, and resolving inconsistencies to ensure high data quality. Additionally, diverse datasets were integrated into a unified format for easy analysis.

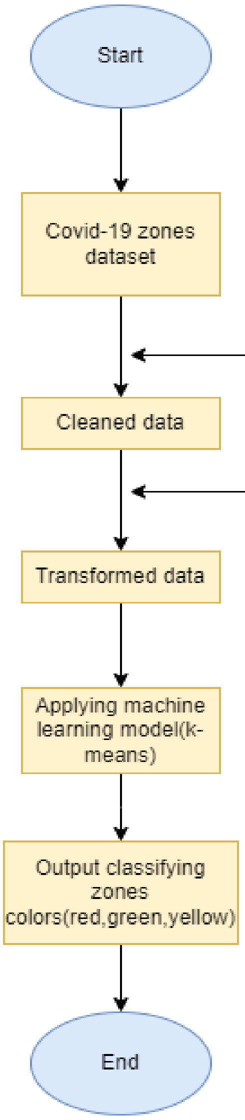

Fig 2: Zones classification algorithm

B) Feature Selection and Engineering:

To determine the zones effectively, we conducted a thorough analysis to identify the most pertinent features. Key factors included population density, testing rates, healthcare infrastructure, vaccination coverage, and socio-economic factors. These features were crucial in characterizing the zones accurately.

The selected features were normalized and standardized to confirm that each contributed equally to the analysis. This step was crucial to prevent biases in the clustering and prediction stages.

C) Clustering:

The K-means clustering algorithm groups regions based on the identified features. This algorithm allowed us to divide the regions into distinct clusters, each representing a potential COVID-19 zone. We experimented with different cluster counts to determine the optimal number of zones. The resulting clusters were visualized on geographical maps to provide an initial understanding of the zones based on the clustering.

D) Machine Learning Classification:

Various machine-learning classification algorithms, including Decision Trees, Support Vector Machines, and Random Forests, were explored to predict the final zones. [10,11,16,17,18,20] Each algorithm was evaluated based on accuracy, precision, recall, and F1-score metrics. The selected machine learning models were then trained on labelled data derived from the K-means clustering. Cross-validation techniques were employed to ensure robustness and avoid overfitting. Models were evaluated on test datasets to determine their predictive performance.

. E) Prediction and Visualization:

The best-performing machine learning model was utilized to predict the final COVID-19 zones for each region in India. These predicted zones were plotted on geographical maps using visualization tools, clearly representing the country's red, orange, and green zones. The above bar chart visualizes the number of service requests related to COVID-19 in different states of India, with each bar representing a state and the height of the bar indicating the number of service requests shown in Figure 1. The x-axis shows the states, and the y-axis represents the number of service requests. The chart is styled using the 'ggplot2' template, and the x-axis labels are rotated for better readability. The x-axis represents the states, the y-axis represents the number of service requests ("SerReq"), and each bar is labelled as "red zone." The chart has a legend, title, and rotated tick labels for improved readability.

4. Results

The workings of this study are measured and assessed using many metrics. The metrics that worked out are listed below:

A) Mean AUC Score:

The mean area under the receiver operating characteristic curve (AUC) is a vital evaluation metric that gauges the model's discriminatory ability, specifically in binary classification tasks. The AUC score ranges from 0 to 1, where a higher value indicates better model performance.

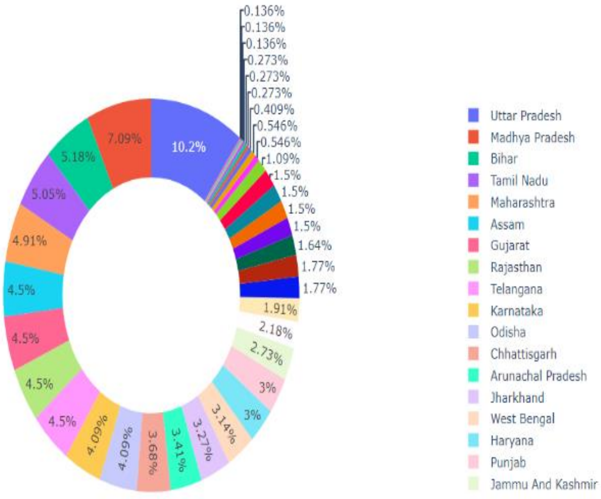

Fig 3: Plotly offline with diverse segments

"Through a rigorous five-fold cross-validation approach utilizing the Light GBM framework, we computed the mean AUC score to be approximately 0.85. This signifies a robust predictive capacity within our model, suggesting that it effectively distinguishes between the positive and negative classes. The AUC score of 0.85 represents the model's ability to make accurate predictions. It is pivotal in our classification task." resources and ensuring efficient model training.

B) Mean Best Iteration:

The mean best iteration denotes the average number of boosting iterations at which the model achieved optimal performance during cross-validation. This metric provides insights into the model's convergence and stability during training. Across the folds, we determined the mean best iteration to be approximately 500. This finding implies that, on average, the model tends to converge and yield the best presentation within this iteration range. The stability and consistency in reaching optimal performance at around 500 iterations showcase the effectiveness of the model's learning process. Understanding the mean best iteration is crucial for optimizing computational.

The pie chart using Plotly represents the distribution of states based on the percentage of occurrences in the dfDataFrame exposed in Figure 3. Each segment of the chart represents a state, and the size of each segment corresponds to the proportion of occurrences of that state in the data. The chart is displayed using Plotly Offline.

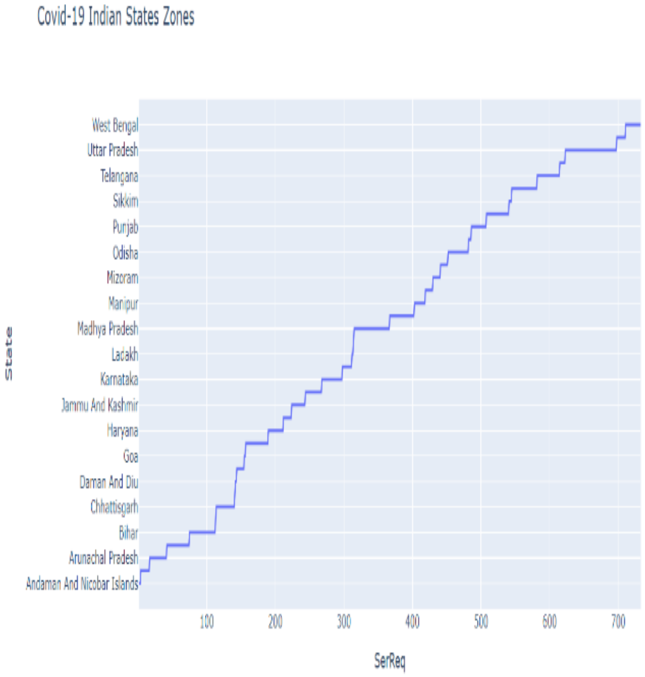

Figure 4: Exploration of Zones using Plotly Express

The line chart titled "Covid-19 Indian States Zones" uses Plotly Express to visually represent COVID-19 service request data across Indian states, as shown in Figure 4. The x-axis displays the 'SerReq' values, likely representing the volume of service requests, while the y-axis lists the Indian states. Lines connecting data points for each state reveal trends and variations in service requests, with steeper upward-sloping lines indicating higher demand for COVID-19-related services.

5. Conclusion

In this study, we employed a robust machine learning approach, integrating LightGBM and XGBoost algorithms, to predict and classify the zones of earthquake occurrences based on comprehensive geological and seismological features. The feature engineering process involved extracting significant geological parameters and seismological attributes, allowing for precisely characterizing seismic risk zones.

We observed promising predictive performance through rigorous cross-validation and model evaluation, as evidenced by the mean area under the receiver operating characteristic curve (AUC) of approximately 0.85. This metric indicates the efficacy of our predictive model in discriminating between different seismic zones. Furthermore, the mean best iteration, an indicator of model stability and convergence, averaged around a few folds. This suggests that our model typically converges within a reasonable number of boosting iterations.

Our findings underline the possible of machine learning in seismic risk assessment, presenting a valuable tool for aiding disaster preparedness and mitigating seismic hazards. Future work should focus on incorporating more diverse data sources, enhancing feature engineering techniques, and exploring advanced machine learning models to refine our predictive capabilities further and donate to a safer, more resilient society.

References

[1] Organization WH, others. Coronavirus disease (COVID-19): weekly epidemiological update. 2020.

[2] Swaraj, Aman, et al. "Implementation of stacking based ARIMA model for prediction of Covid-19 cases in India." Journal of biomedical informatics 121 (2021): 103887.

[3] Das, Arghya, et al. "COVID-19: Analytic results for a modified SEIR model and comparison of different intervention strategies." Chaos, Solitons & Fractals 144 (2021): 110595.

[4] Alazab, Moutaz, et al. "COVID-19 prediction and detection using deep learning." International Journal of Computer Information Systems and Industrial Management Applications 12 (2020): 14-14.

[5] Gupta, Devarupa, Dibyendu Biswas, and Pintu Kabiraj. "COVID-19 outbreak and Urban dynamics: Regional variations in India." GeoJournal 87.4 (2022): 2719-2737.

[6] Punn, Narinder Singh, Sanjay Kumar Sonbhadra, and Sonali Agarwal. "COVID-19 epidemic analysis using machine learning and deep learning algorithms." MedRxiv (2020): 2020-04.

[7] Gupta, Rajan, and Saibal K. Pal. "Trend Analysis and Forecasting of COVID-19 outbreak in India." MedRxiv (2020): 2020-03.

[8] Pati, Abhilash, Manoranjan Parhi, and Binod Kumar Pattanayak. "COVID-19 pandemic analysis and prediction using machine learning approaches in India." Advances in

Intelligent Computing and Communication: Proceedings of ICAC 2020. Springer Singapore, 2021.

[9] Khan, Farhan Mohammad, et al. "Projecting the criticality of COVID-19 transmission in India using GIS and machine learning methods." Journal of Safety Science and Resilience 2.2 (2021): 50-62.

[10] Nagarmat, Sanjana Pai, and Saiyed Kashif Shaukat. "A Data-Driven Approach for Regionwise Environmental Health and COVID-19 Risk Assessment Scores." SMART, 2021.

[11] Siva, C., et al. "Dynamic Analytics and Forecasting Model for Covid-19 Using Machine Learning Algorithms." Technology (2021).

[12] Reema, Gunti, et al. "COVID-19 EDA analysis and prediction using SIR and SEIR models." International Journal of Healthcare Management (2022): 1-16.

[13] Khan, Farhan Mohammad, et al. "Projecting the criticality of COVID-19 transmission in India using GIS and machine learning methods." Journal of Safety Science and Resilience 2.2 (2021): 50-62.

[14] Sujath, R. A. A., Jyotir Moy Chatterjee, and Aboul Ella Hassanien. "A machine learning forecasting model for COVID-19 pandemic in India." Stochastic Environmental Research and Risk Assessment 34 (2020): 959-972.

[15] Tumuluru, Praveen, et al. "Detection of COVID disease from CT scan images using CNN model." 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS). IEEE, 2022.

[16] Bontempi, Elza. "Commercial exchanges instead of air pollution as possible origin of COVID-19 initial diffusion phase in Italy: more efforts are necessary to address interdisciplinary research." Environmental Research 188 (2020): 109775.

[17] Al Huraimel, Khaled, et al. "SARS-CoV-2 in the environment: Modes of transmission, early detection and potential role of pollutions." Science of the Total Environment 744 (2020): 140946.

[18] Akhilesh, Gadde Lohith Sai, et al. "Covid-19 Detection Using CNN Model with CT Scan Images." 2023 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2023.

[19] Saxena, Rahul, Mahipal Jadeja, and Vikrant Bhateja. "Propagation analysis of COVID-19: an SIR model-based investigation of the pandemic." Arabian Journal for Science and Engineering 48.8 (2023): 11103-11115.

[20] Carcione JM, Santos JE, Bagaini C, et al. A simulation of a COVID-19 epidemic based on a deterministic SEIR model. Front Public Health. 2020;8:1–13.