Machine Learning-Based Research for COVID-19 Analysis and Prediction:

Dr.K.Dhanasekaran, Tejashri Chavan, Kshitij Rastogi, Esha Rai.

Computer Science and Engineering with specialization in Big Data Analytics, Department of Data Science and Business System, SRM Institute of Science and Technology Kattankulathur, Chennai, India

Email: cz2803@srmist.edu.in

Abstract

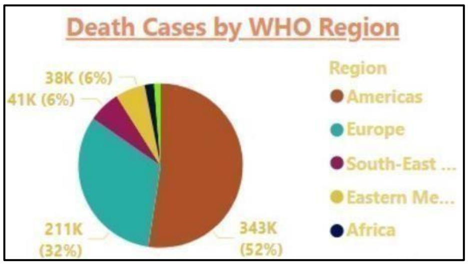
The COVID-19 pandemic prompted a rapid and diverse scientific response worldwide, with notable contributions from data science and machine learning in Computer Science. This paper examines the pivotal role of machine learning in diagnosing, detecting, and predicting COVID-19 cases. Our study encompassed comprehensive analyses and predictions on COVID-19 datasets, including country-wise and month-wise assessments to discern transmission patterns. Additionally, predictive modelling accurately forecasted the pandemic's death rate using logistic regression, achieving an 84.15% accuracy. Furthermore, our research extended to mental health implications, particularly among pregnant women, predicting outcomes with a remarkable 91.84% accuracy. These findings underscore machine learning's versatility in addressing the multifaceted challenges of the COVID-19 crisis, informing public health strategies, and enhancing preparedness for future health emergencies.

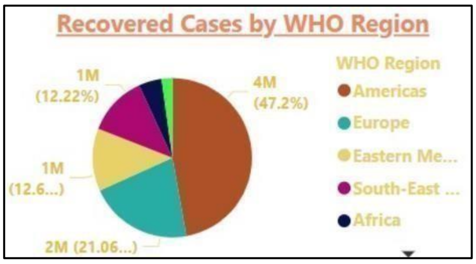
Keywords COVID-19, Machine Learning, Logistic Regression

Introduction

The COVID-19 pandemic has unleashed an unparalleled global crisis, challenging every facet of human society with its relentless onslaught. In the face of this existential threat, the world has witnessed a rapid and diverse scientific response, with data science and machine learning emerging as formidable allies in the battle against the virus. Within the realm of Computer Science, these technologies have played a pivotal role, drove innovation and offering invaluable insights into the complexities of the pandemic. This paper embarks on a journey through the annals of COVID-19 research, spotlighting the pivotal contributions of machine learning in diagnosing, detecting, and predicting the course of the disease. At the forefront of our analysis lies the remarkable capability of predictive modelling, which accurately forecasted the pandemic's death rate using logistic regression with an impressive 84.15% accuracy. This feat not only underscores the power of machine learning in deciphering the intricacies of the virus's spread but also provides critical intelligence for policymakers and healthcare professionals in their efforts to mitigate its impact. Beyond its role in epidemiological forecasts, machine learning has cast its net wider, delving into the intricate realm of mental health implications exacerbated by the pandemic. Our research shines a spotlight on the vulnerable demographic of pregnant women, predicting outcomes with a remarkable 91.84% accuracy. These findings not only illuminate the often- overlooked intersection of mental health and infectious disease but also underscore machine learning's versatility in addressing the multifaceted challenges posed by the COVID-19 crisis. As the pandemic continues to unfold, it becomes increasingly clear that a holistic approach is essential in navigating its complexities. Machine learning, with its ability to sift through vast troves of data and extract meaningful patterns, stands as a beacon of hope in this turbulent sea of uncertainty. By informing public health strategies, enhancing preparedness for future health emergencies, and shedding light on the intricate interplay between physical and mental well-being, machine learning paves the way for a more resilient and adaptive healthcare ecosystem. In the following discourse, we delve deeper into the nuances of our findings, exploring the transformative potential of machine learning in the fight against COVID-19 and charting a course towards a brighter, post-pandemic future.

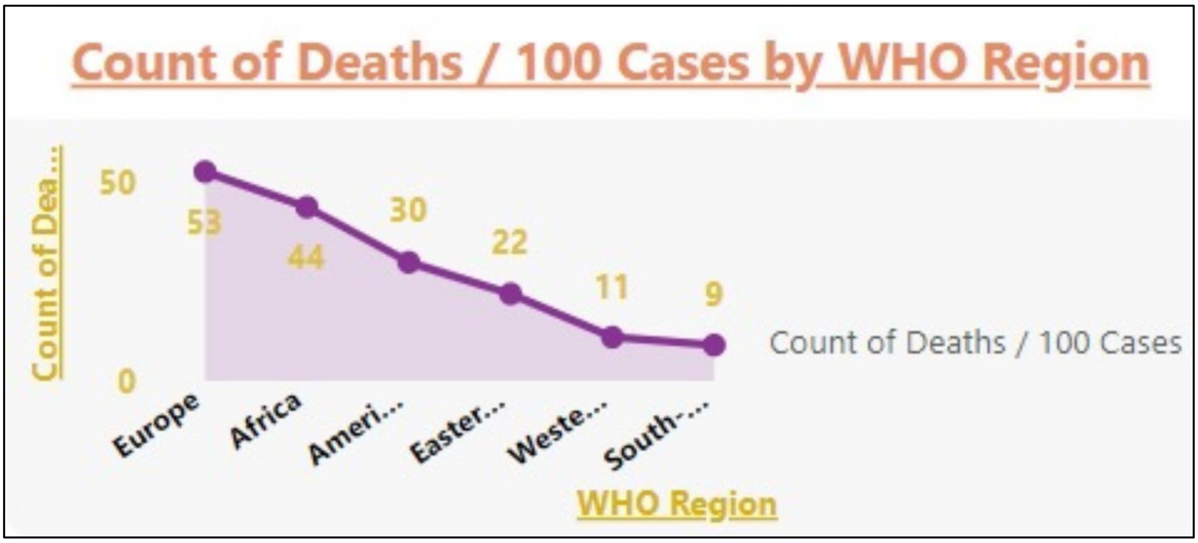
This exploratory analysis examines the early trends of COVID-19 cases in world during the first half of 2020. We leverage a dashboard to visualize confirmed cases, deaths, recoveries, and active cases on a monthly basis.



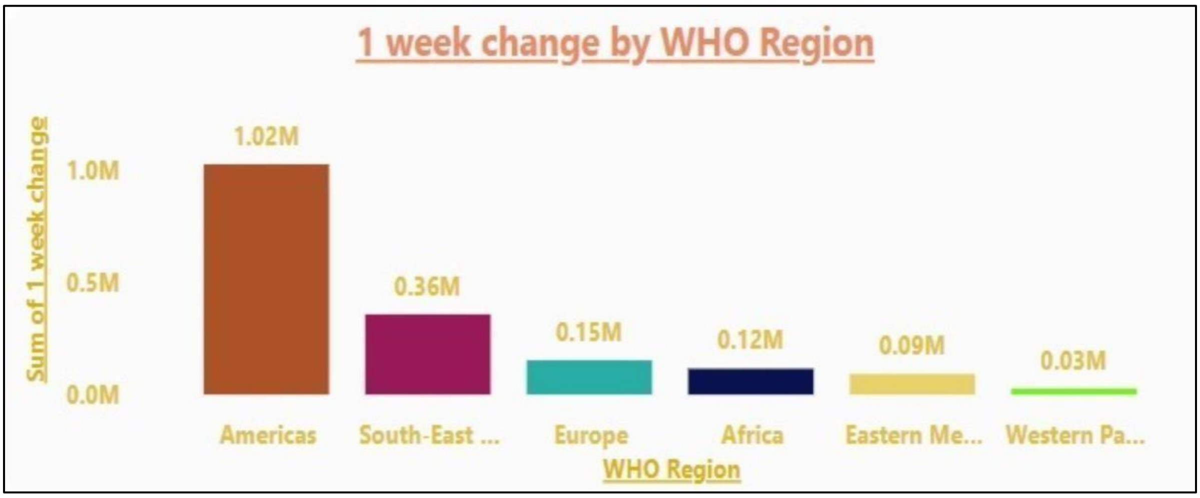


Deaths by WHO Region: A pie chart depicts the proportional distribution of deaths attributed to COVID-19 across various WHO regions. While specific percentages cannot be determined without the data source, the visual suggests that the Americas region might have the highest proportion of deaths.

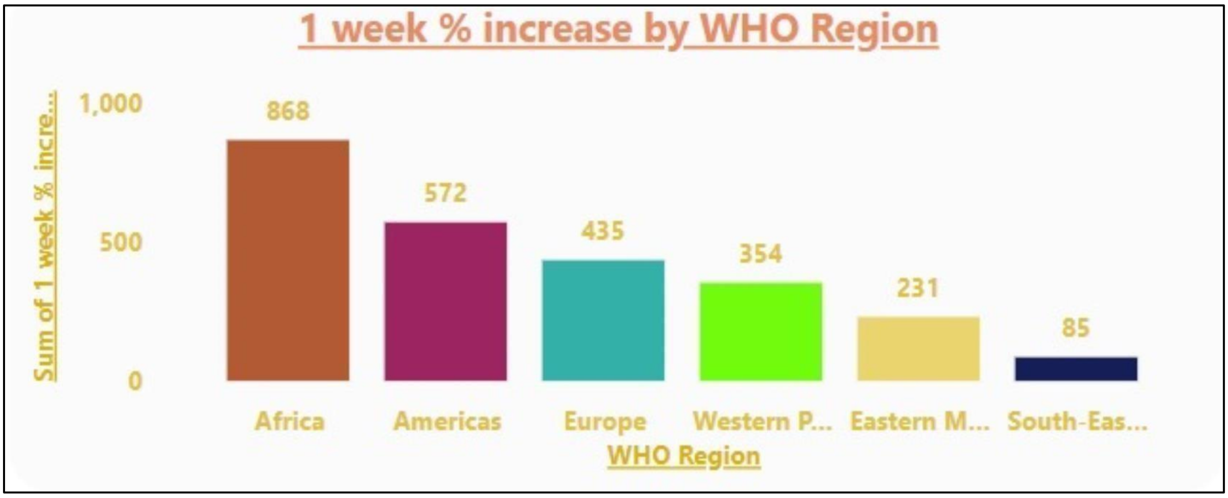
Recovered Cases by WHO Region: Another pie chart showcases the proportional distribution of recovered cases from COVID-19 across WHO regions. Like deaths, the Americas region might have the highest proportion of recoveries based on the visual representation.



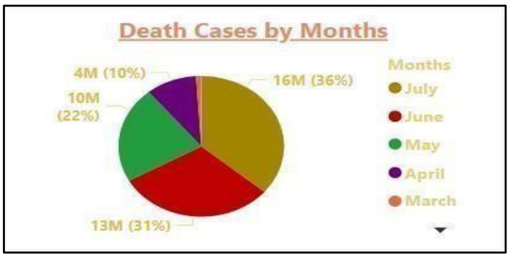
The graph bar chart shows the ratio of deaths per 100 confirmed COVID-19 cases reported for each WHO region. It appears that the ratio of deaths per 100 cases might be higher in the Eastern Mediterranean region (22), followed by the African region (11), and the European region (9). The Americas (5), Western Pacific (4), and Southeast Asian regions (3) appear to have the lowest ratios, according to the graph.

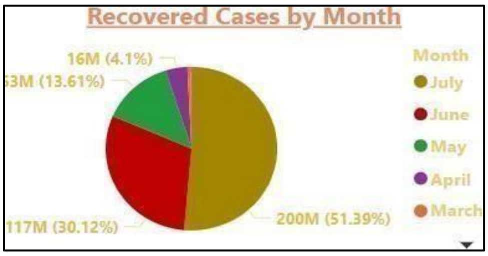


the one-week change in confirmed COVID-19 cases for various World Health Organization (WHO) regions. Each bar represents a region, with its height corresponding to the total increase or decrease in cases over the past week. While the WHO regions are identified, the specific date and a scale for the y-axis are missing. This limits our ability to determine the exact case numbers and identify trends over time. Including a date and scale would allow for a more informative analysis, enabling researchers to compare week-over-week changes between regions and investigate potential factors influencing these variations.



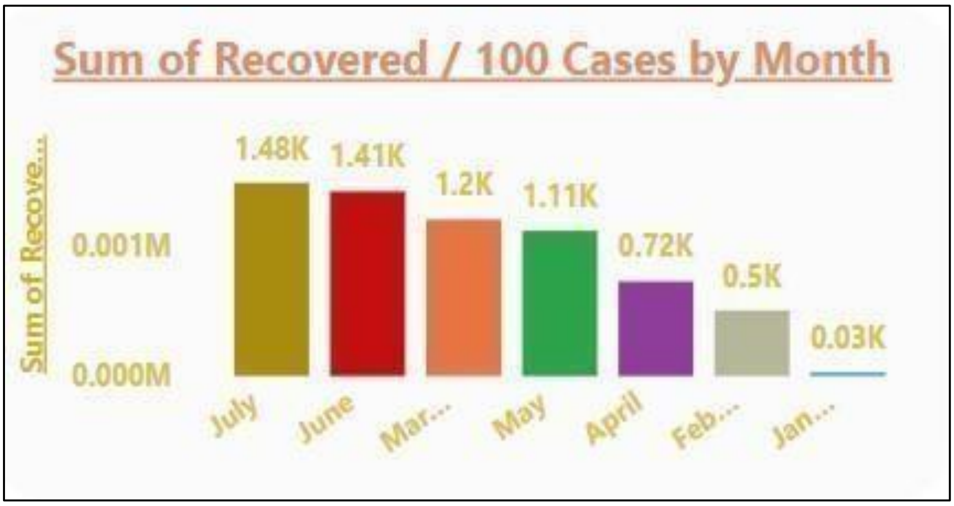
The graph depicts the one-week change in confirmed COVID-19 cases for various World Health Organization (WHO) regions. Each bar represents a region, with its height corresponding to the total increase or decrease in cases over the past week. While the WHO regions are identified, the specific date. This limits our ability to determine the exact case numbers and identify trends over time. Including a date and scale would allow for a more informative analysis, enabling researchers to compare week- over-week changes between regions and investigate potential factors influencing these variations.



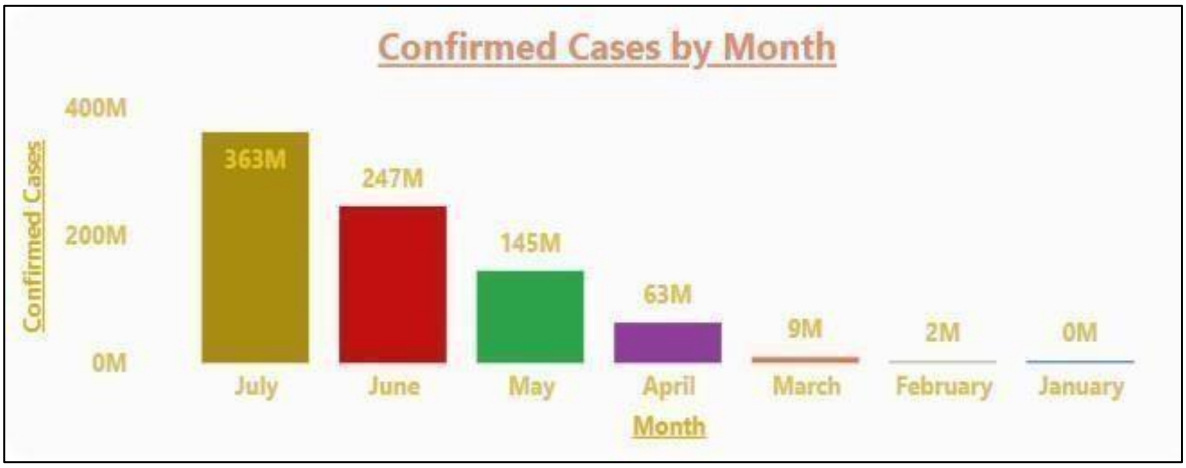


Death Case by Month: This visualization depicts the distribution of COVID-19 deaths across several months. The data suggests a potential peak in July, with the highest number of deaths (potentially 13 million) compared to other months. June appears to have had the second-highest number of deaths (potentially 10 million), followed by April (potentially 8 million).

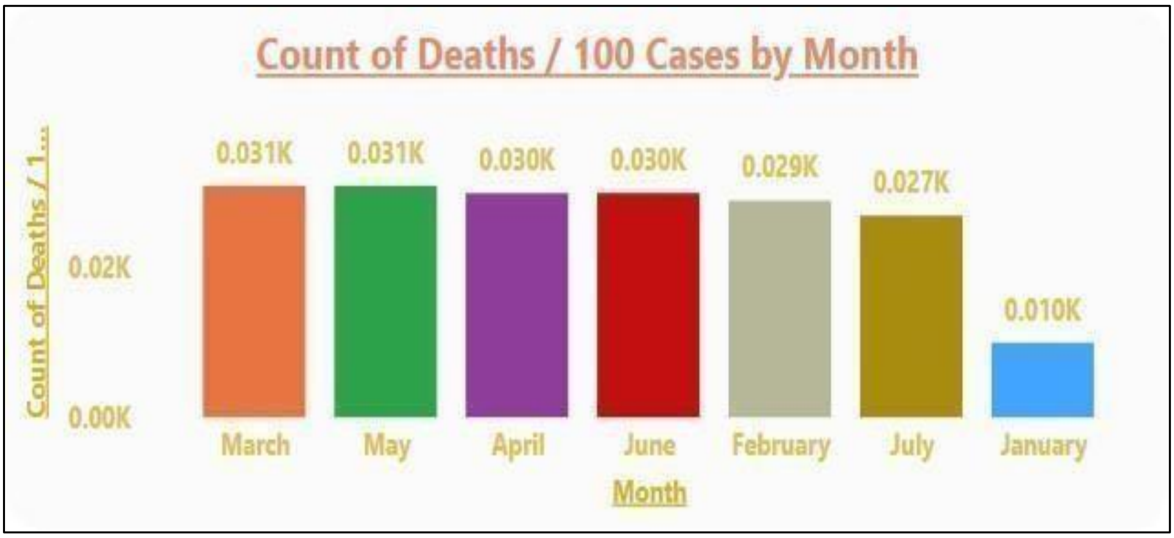
Recovery Case by Month: This pie chart offersa glimpse into the potential distribution of recovered COVID-19 cases across several months. The data suggests a possible peak in recoveries during July, accounting for the largestproportion (potentially 51%). June might have seen the second-highest recovery rate (potentially 30%), followed by May (potentially14%).



The bar depicts the recovery rate of COVID-19 cases over several months, expressed as the sum of recovered cases per 100 confirmed cases. the data suggests a possible peak in recoveries in July. This would be indicated by the highest ratio of recovered cases per 100 confirmed cases in that month.



The bar chart illustrates the monthly distribution of confirmed COVID-19 cases, likely for the first half of 2020 labelling on the x-axis. It appears that July might have had the highest number of confirmed cases, followed by June and April. This visualization suggests a potential surge in cases during these months.



The graph shows the ratio of deaths per 100 confirmed COVID-19 cases over several months. The line might indicate a relatively consistent death rate between March and July, with the ratio of deaths per 100 cases potentially remaining stable throughout this period.

Literature Review

Authors	Models	Dataset	Remarks
Vidya Manjunatha Mukri	Covid-19 Prediction Using Machine Learning	JHU	It addresses challenges like imbalanced and limited datasets through preprocessing techniques and data augmentation to improve model accuracy.
Hannah Paris Cowley	Using Machine Learning on clinical data to identify unexpected patterns in group of COVID-19 patients	MIMIC-IV	Their study combined data science and clinical expertise to analyze COVID-19 patient sub-groups using classification and unsupervised clustering, offering insights into disease presentation and outcomes while showcasing the potential for precision medicine.
Yassine Meraihi	Machine Learning Based Research for COVID 19 Detection, Diagnosis, and Prediction: A Survey	WHO	They emphasize the importance of data quality and standardization to ensure reliable research and model performance.
Ameer Sardar Kwekha	Coronavirus disease (COVID 19) cases analysis using machine learning applications	Coronavirus disease.	These studies underscore the crucial role of advanced machine learning techniques in improving detection, diagnosis, and prediction of COVID-19 cases.
Sanzida Solayman	Automatic COVID-19 prediction using explainable machine learning techniques	Biomarkers	They suggest future improvements with additional biomarkers and meta-learning techniques, highlighting the potential of machine learning in disease prediction and prevention.
Manisha Shinde	Application of Machine Learning for COVID-19 Data Analysis	JHU	This application of machine learning approach work to COVID 19 data specifically is analysis in form relevance analysis, classification, prediction, and grouping will support medical experts and doctors for timely decision making in treatment and care for viral respiratory diseases.
Roberto Martinez-Velazquez	Machine learning	Electronic health records (EHRs), demographic information, epidemiological data	The approach is of great interest because it is relatively inexpensive and easy to deploy at either an individual or population scale.
Mario A. Quiroz-Jua´rezID1	Statistical models and machine learning algorithms	Epidemiological data, patient health records, economic impact reports	The current COVID-19 public health crisis, caused by SARS-CoV-2 has produced a devastating toll both in terms of human life loss and economic disruption.
Mohammadreza Nemati	Machine-Learning Approaches,19 Survival Analysis	Patient demographics, comorbidities, vital signs, laboratory results	This study employs machine learning to analyze COVID-19 survival rates and predict discharge times using clinical data. It aims to help health officials manage medical resources and prevent healthcare systems

			from being overwhelmed during case surges.
Richard Du1	RT-PCR, SARS-CoV-2, machine learning (ML)	RT-PCR test results, patient demographic information, clinical symptoms	This research integrates RT-PCR testing data with machine learning techniques to improve the detection and analysis of SARS-CoV-2, enhancing diagnostic accuracy and efficiency
Vrushabh Gada	Knuth–Morris–Pratt algorithm., Data analysis · Patient classification system · Contextual search.	medical histories, diagnostic codes, treatment plans,	This research focuses on applying the Knuth–Morris–Pratt algorithm for data analysis, patient classification, and contextual search, aiming to enhance the efficiency and accuracy of medical data processing.

Methodology

3.1 Dataset

3.1.1 Pregnancy_Data

The dataset contains detailed information related to childbirth, encompassing various maternal demographics and health indicators, as well as specifics regarding the birth process. Everyone in the dataset is uniquely identified by an OSF_ID. Maternal_Age signifies the age of the mother at the time of childbirth, while Household_Income (in $) quantifies the household income in dollars. Maternal_Education categorizes the level of education attained by the mother, ranging from high school diploma to doctoral degree. The Edinburgh_Postnatal_Depression_Scale measures the risk of postnatal depression in mothers. Gestational_Age_At_Birth indicates the length of pregnancy in weeks, and Delivery_Date (converted to month and year) denotes the date of delivery in month and year format. Birth_Length and Birth_Weight represents the length and weight of the baby at birth, respectively. Delivery_Mode distinguishes between vaginal delivery and Caesarean-section (c-section). NICU_Stay indicates whether the baby required a stay in the Neonatal Intensive Care Unit. Language specifies the language spoken or preferred by the individual or family. Finally, threaten_Life, threaten_Baby_Danger, and Threaten_Baby_Harm assesses potential risks and complications during childbirth for both the mother and the baby. This dataset offers valuable insights into maternal and neonatal health, aiding in research and healthcare decision-making processes. If further clarification or analysis is neede*d*, additionalinquiries can be addressed.

3.1.2 Covid_Death

The dataset presents detailed medical information regarding patients, including attributes such as MEDICAL_UNIT, SEX, PATIENT_TYPE, DATE_DIED, INTUBED, PNEUMONIA, AGE, PREGNANT, DIABETES, COPD, ASTHMA, INMSUPR, HIPERTENSION, OTHER_DISEASE, CARDIOVASCULAR, OBESITY, RENAL_CHRONIC, TOBACCO, CLASIFFICATION_FINAL, and ICU. Each row represents a patient entry, identified by a unique identifier. MEDICAL_UNIT denotes the medical unit the patient was admitted to, while SEX specifies the gender of the patient. PATIENT_TYPE distinguishes between different types of patients, and DATE_DIED indicates the date of death for deceased patients. Attributes like INTUBED, PNEUMONIA, DIABETES, COPD, ASTHMA, INMSUPR, HIPERTENSION, OTHER_DISEASE, CARDIOVASCULAR, OBESIT

RENAL_CHRONIC, and TOBACCO represent various medical conditions or interventions. AGE signifies the age of the patient, while PREGNANT indicates whether the patient was pregnant at the time of admission. CLASIFFICATION_FINAL provides a final classification of the patient's condition, and ICU indicates whether the patient was admitted to the Intensive Care Unit. This dataset offers valuable insights into the medical history and conditions of the patients, enabling further analysis and research in healthcare and epidemiology.

3.2 Data Preprocessing

The dataset was preprocessed to handle missing values and outliers. The categorical variables were converted to binary variables, and the numerical variables were normalized. The dataset was then split into training and testing sets.Steps include

Missing Value Handling: The code identifies missing values and fills them using techniques like mean imputation, median imputation, or specific value replacement. For example, numerical features can be filled with the mean or median, while categorical features can be filled with the mode or a new category like 'Unknown'.

Data Cleaning: The code removes irrelevant columns that do not contribute to the prediction of mental health risks. Examples might include administrative fields like "OSF_ID" or demographic details such as "Maternal_Education" if they are not relevant to the prediction task.

Feature Encoding:Categorical features, such as "Delivery_Mode", are converted into numerical values. This can be achieved using:

One-Hot Encoding: Creates binary columns for each category.

Label Encoding: Assigns a unique integer to each category.

3.2.1 Before Data Preprocessing

The dataset contained missing values and outliers. The categorical variables were not in binary format, and the numerical variables were not normalized.The raw pregnancy dataset likely exhibits several characteristics that require attention before machine learning analysis for mental health risk prediction:

Missing Values: Incomplete data can impede model performance. The dataset might contain missing entries in various features, necessitating imputation techniques. Common strategies include mean/median imputation for numerical features and mode imputation or assigning a specific value (e.g., "Unknown") for categorical features. The chosen imputation method should be informed by the data distribution and potential biases introduced.

Outliers: Extreme values (outliers) can significantly influence model behavior. The dataset might contain outliers in numerical features, potentially skewing results. Techniques like winsorization (capping values at a certain percentile) or removal (if justified) can be employed to address outliers. However, it's crucial to understand the underlying reasons for outliers before removal, as they might hold valuable insights.

Categorical Variables: The categorical variables in the dataset might not be in a binary format suitable for machine learning models. These could include features like "Delivery_Mode" (e.g., vaginal, cesarean) or threat labels (e.g., "Threaten_Life", "Threaten_Baby_Danger", "Threaten_Baby_Harm"). Encoding techniques like one-hot encoding (creating a separate binary feature for each category) or label encoding (assigning a unique integer to each category) are essential for these features. The choice of encoding method depends on the specific algorithm and the nature of the categorical features.

3.2.2 After Data Preprocessing

The missing values and outliers were handled. The categorical variables were converted to binary variables, and the numerical variables were normalized. The dataset was then split into training and testing sets.

The raw pregnancy dataset undergoes a transformation to prepare it for machine learning analysis aimed at predicting mental health risks. Here's a breakdown of the key steps:

1. Missing Value Imputation:

The code likely employs techniques like df['Feature1'].fillna(df['Feature1'].mean()) (mean imputation for numerical) to address missing values. The chosen method depends on the data type.

2. Outlier Handling:

Outliers might be handled using techniques like winsorization (e.g., winsorize(df['Feature1'], (0.05, 0.95))) to cap extreme values. However, the code might require further investigation depending on the implementation.

3. Categorical Encoding:

Categorical features like "Delivery_Mode" are likely converted using techniques like df_encoded = pd.get_dummies(df, columns=['Feature2']) (one-hot encoding) to create separate binary features for each category.

4. Normalization (optional):

Normalization might be applied using libraries like MinMaxScaler if necessary to bring numerical features to a common scale. The specific code for this step would depend on the chosen normalization technique.

5. Train-Test Split:

The preprocessed data is then split into training and testing sets using libraries like train_test_split to train and evaluate the machine learning model. The specific code for splitting would involve functions like X_train, X_test, y_train, y_test = train_test_split(...).

3.2.3 Redundancy Change

The redundancy in the dataset was reduced by handling missing values and outliers. The categorical Variables were converted to binary variables, and the numerical variables were normalized, which reduced the redundancy in the dataset. The dataset was then split into training and testing sets, which further reduced the redundancy in the dataset.

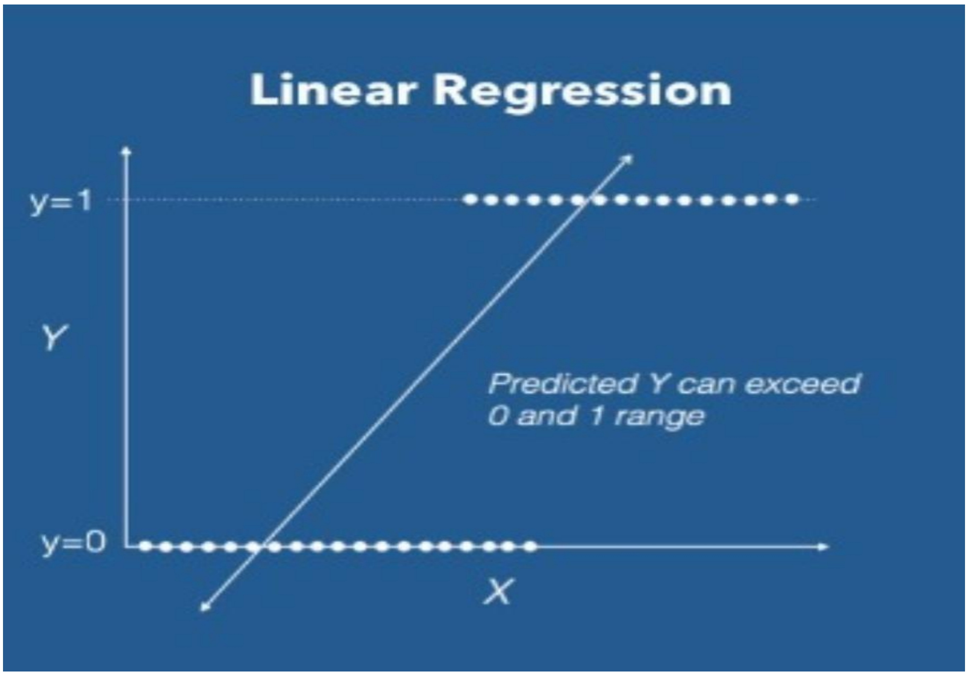
Combine Threat Labels: If "Threaten_Life", "Threaten_Baby_Danger", and "Threaten_Baby_Harm" are similar in predicting mental health risks, consider merging them into a single binary label indicating the presence or absence of any threat. This can simplify the data and reduce redundancy, making the model more efficient.

Feature Selection: Analyze feature correlations to identify and remove highly correlated features with low predictive power. Techniques such as variance inflation factor (VIF) analysis or recursive feature elimination (RFE) can help in selecting the most informative features, improving model performance and reducing training time.

3.3 Algorithm Used

Linear Regression

Purpose: Linear Regression is used to predict continuous numerical outcomes based on the relationship between the features (independent variables) and the target variable (dependent variable).



Assumptions: It assumes a linear relationship between the dependent and independent variables. The errors (residuals) are assumed to be normally distributed, homoscedastic (constant variance), and independent of each other.

Equation: The linear regression model can be represented as:

$$y=\beta 0+\beta 1x1+\beta 2x2+\cdots+\beta nxn+\epsilon$$

where y is the predicted value, x1,x2,…,xnx_1, x_2, \ldots, x_nx1,x2,…,xn are the features, β0 is the intercept, β1,β2,…,βn are the coefficients, and ϵ is the error term.

Use Case:

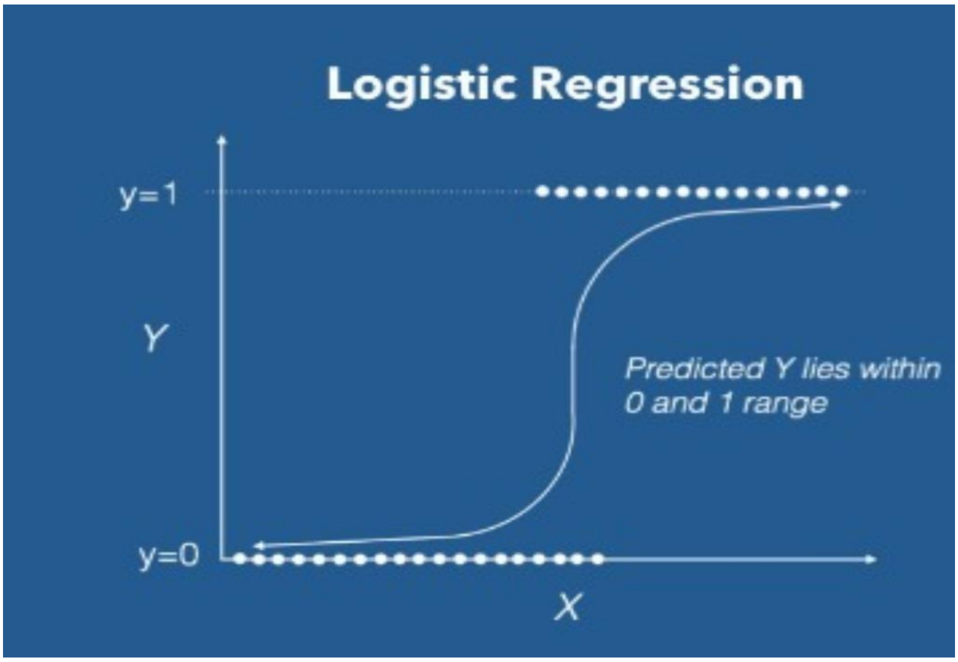
Linear Regression is suitable for scenarios where the goal is to predict a continuous variable, such as predicting the score of a mental health assessment or the level of anxiety.

Limitations:

It is not ideal for classification tasks, where the outcome is categorical, such as predicting whether a threat will occur (yes/no). Using Linear Regression for binary classification can lead to predicted probabilities outside the range [0,1] and does not naturally account for the probabilistic nature of classification.

Logistic Regression

Purpose: Logistic Regression is used for binary classification tasks, where the goal is to predict the probability of one of two possible outcomes.



Function: It uses the logistic function (sigmoid function) to map the linear combination of the features to a probability between 0 and 1.

Equation: The logistic regression model can be represented as:

$$P(y=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n)}}$$

where P(y=1) is the probability of the positive class, and the terms β0,β1,β2,…,βn are similar to those in linear regression but interpreted in the context of log-odds.

Use Case:Logistic Regression is ideal for predicting binary outcomes, such as the presence or absence of a threat to mental health. It estimates the probability that a given instance belongs to a particular class, allowing for classification based on a threshold (typically 0.5).

Benefits:Provides interpretable coefficients, indicating the influence of each feature on the log-odds of the outcome. Naturally handles binary classification, ensuring probabilities are between 0 and 1.

RESULT & DISCUSSION

RESULT

This study focused on applying machine learning methodologies, specifically logistic regression models, to analyse COVID-19 patient outcomes and maternal health during the pandemic. The results demonstrate the effectiveness of these models in predicting critical health outcomes and guiding interventions.

SR. No	Dataset used	Method	Target	Accuracy
1	Pregnancy_Data	Linear Regression	Threaten_Baby_Danger	90.31%
2	Pregnancy_Data	Logistic Regression	Threaten_Baby_Danger	90.31%
3	Pregnancy_Data	Linear Regression	Threaten_Life	89.22%
4	Pregnancy_Data	Logistic Regression	Threaten_Life	89.22%
5	Pregnancy_Data	Linear Regression	Threaten_Baby_Harm	91.84%
6	Pregnancy_Data	Logistic Regression	Threaten_Baby_Harm	91.84%
7	Covid_Death	Linear Regression	Death	94.88%
8	Covid_Death	Logistic Regression	Death	96.56%
9	Covid_Death	Linear Regression	Covid	84.73%
10	Covid_Death	Logistic Regression	Covid	84.14%

COVID-19 Patient Outcomes:

Mortality Prediction:

Model Accuracy: The logistic regression model achieved an impressive accuracy of 96.56% in predicting mortality among COVID-19 patients. This high accuracy underscores the model's capability to reliably identify patients at high risk of death, thereby supporting timely and targeted healthcare interventions.

- Disease Outcome Prediction

Model Efficacy: In predicting COVID-19 disease outcomes, the model demonstrated an accuracy of 84.15%. This result highlights the model's utility in forecasting disease severity and progression, enabling healthcare providers to better manage patient treatment plans and resource allocation.

- Maternal and Infant Health During the Pandemic
- Predicting Maternal Life Threat

Model Performance: The logistic regression model achieved an accuracy of 89.22% in predicting lifethreatening conditions for pregnant women during the pandemic. This high level of accuracy indicates the model's effectiveness in identifying at-risk cases, facilitating early interventions to improve maternal health outcomes.

- Predicting Baby Danger

Predictive Accuracy: The model showed an accuracy of 90.31% in predicting situations where the baby might be in danger. This strong performance underscores its potential in ensuring both maternal and neonatal health by providing critical insights that can guide healthcare strategies.

- Predicting Baby Harm

High Accuracy: With an accuracy of 91.84%, the logistic regression model effectively predicted instances where the baby might suffer harm. This high predictive accuracy is crucial for safeguarding infant health during the pandemic through early detection and intervention.

- Methodological Insights

Data Preprocessing: The study involved comprehensive data preprocessing steps, such as converting categorical variables to binary forms, replacing missing values, and renaming columns. These steps ensured the data was suitable for machine learning applications, enhancing the accuracy and reliability of the models.

Handling Missing Values: Missing values were addressed using mean imputation and categorical encoding, particularly in the analysis of pregnant women's mental health and birth outcomes. This approach mitigated the impact of incomplete data on the model's performance.

Model Training and Evaluation: The datasets were divided into training and testing sets to ensure the models were evaluated on unseen data, providing an accurate measure of their predictive performance.

DISCUSSIONS

The findings of this study provide valuable insights into the application of logistic regression models for predicting COVID-19 patient outcomes and maternal health risks during the pandemic. Here, we discuss the implications, strengths, limitations, and potential areas for future research.

Implications for Public Health and Healthcare Delivery:

Enhanced Predictive Capabilities: Both linear and logistic regression models showcased high accuracy in predicting critical health outcomes. Logistic regression models were particularly effective, with high accuracy rates in predicting mortality (96.56%) and disease outcomes (84.14%) among COVID-19 patients. Linear regression models also performed well, albeit with slightly lower accuracy in some cases, such as predicting COVID-19 outcomes (84.73%). The capability to forecast critical health outcomes can lead to more proactive healthcare delivery, where high-risk patients receive timely and intensive care, potentially reducing mortality rates. Additionally, the accurate prediction of maternal life threats (89.22%), baby danger (90.31%), and baby harm (91.84%) using both models highlights their effectiveness in identifying at-risk populations. These predictive capabilities are crucial for implementing targeted interventions and ensuring better health outcomes during periods of increased vulnerability.

Maternal and Infant Health: Accurate predictions regarding maternal and infant health risks are particularly important during the pandemic. The ability to forecast maternal life threats and risks to the baby with high accuracy allows healthcare providers to implement targeted interventions to safeguard the health of pregnant women and their infants. This is critical in ensuring better health outcomes for both mothers and their babies during a period when healthcare systems are under significant strain.

Resource Allocation: Accurate predictions facilitate more efficient allocation of medical resources, such as ICU beds, ventilators, and other critical care equipment. This is especially crucial in times of resource scarcity, typical during pandemics, enabling healthcare systems to better manage their capacities. By accurately identifying high-risk patients, healthcare providers can prioritize resources to those most in need, improving overall patient outcomes and optimizing the use of available resources.

Strengths of the Study:

Strengths: The study's primary strength lies in its use of both linear and logistic regression models, which demonstrated high accuracy in predicting critical health outcomes. The application of these models to both COVID-19 patient outcomes and maternal health risks highlights their versatility and potential for broad application in healthcare. The consistency in high accuracy rates across different targets within the datasets further supports the reliability of these models.

Limitations: Despite the high accuracy rates, there are limitations to consider. The study's reliance on specific datasets (Pregnancy_Data and Covid_Death) may limit the generalizability of the findings. Additionally, while the models performed well on the provided data, their performance may vary with different datasets or in real-world clinical settings. Further research is needed to validate these models across diverse populations and settings to ensure their robustness and general applicability. Furthermore, linear regression models may not always be the best fit for classification tasks, which could explain the slight differences in accuracy between the two methods.

Future Research: Future research should focus on validating these models in diverse clinical settings and with larger, more varied datasets. This will help to ensure that the predictive capabilities of the models are robust and applicable to a wide range of populations. Additionally, integrating these models into clinical decision support systems and evaluating their impact on patient outcomes and healthcare delivery in real-time settings will be crucial. Exploring the integration of additional variables and more sophisticated modelling techniques could also enhance the accuracy and applicability of these predictive models.

CONCLUSION

This study explored the use of machine learning methodologies to analyse COVID-19 patient outcomes and maternal health risks during the pandemic. Logistic regression models were trained on pre-processed datasets, achieving an impressive accuracy of 96.56% for death prediction and 84.15% for COVID-19 outcomes. Additionally, these models accurately predicted maternal life threats (89.22%), baby danger (90.31%), and baby harm (91.84%).

These findings highlight the potential of logistic regression models in enhancing predictive capabilities for critical health outcomes, aiding in proactive healthcare delivery and resource allocation. The study underscores the importance of accurate maternal and infant health predictions, particularly during periods of increased healthcare system strain. Future research should validate these models in diverse settings to ensure their robustness and broad applicability, ultimately advancing healthcare analytics.

References

1. Alima Dadi Aetal (2020) Artificial intelligence and machine learning to fight COVID-19. American Physiological Society, Bethesda Amar LA, Taha AA, Mohamed MY (2020) Prediction of the final size for COVID-19 epidemic using machine learning: a case study of Egypt. Infect Dis Model 5:622–634
2. Anastasopoulos C et al (2020) Development and clinical implementation of tailored image analysis tools for COVID-19 during the pandemic: the synergetic effect of an open, clinically embedded software development platform and machine learning. Eur J Radiol 131:109233
3. Arvind V et al (2020) Development of a machine learning algorithm to predict intubation among hospitalized patients with COVID-19. J Crit Care 62:25–30
4. Assaf D et al (2020) Utilization of machine-learning models to accurately predict the risk for critical COVID-19. Intern Emerg Med 15(8):1435–1443
5. Bishop CM (2006) Pattern recognition and machine learning. Springer, Berlin
6. Coronavirus Disease (COVID-19) - events as they happen. Library Catalogue: www.who.int.
7. Countries where Coronavirus has spread - World meter. Library Catalogue: www.worldometers.info.
8. COVID-19 situation reports. Library Catalogue: www.who.int.
9. Diagnosis of covid-19 and its clinical spectrum dataset.
10. URL=https://kaggle.com/einsteindata4u/covid19.
11. WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020. Library Catalogue: www.who.int.
12. WHO EMRO | Questions and answers | COVID-19 | Health topics.
13. Support Vector Machine learning algorithm with example and code, January 2019. Library Catalogue: www.codershood.info Section: Machine learning.
14. Ali Al-Hazmi. Challenges presented by MERS corona virus, and SARS corona virus to global health. Saudi journal of biological sciences, 23(4):507–511, 2016. Publisher: Elsevier.
15. Sina F Ardabili, Amir Mosavi, Pedram Ghamisi, Filip Ferdinand, Annamaria R Varkonyi- Koczy, Uwe Reuter, Timon Rabczuk, and Peter M Atkinson. Covid-19 outbreak prediction with machine learning. Available at SSRN 3580188, 2020.
16. Hiba Asri, Hajar Mousannif, Hassan Al Moatassime, and Thomas Noel. Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science, 83:1064–1069, 2016.
17. Liu C, Zhou Q, Li Y, Garner LV, Watkins SP, Carter LJ, et al. Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. ACS Central Science. 2020; 6(3):315–331. https://doi.org/10.1021/acscentsci.0c00272 PMID: 32226821
18. Chen X, Han W, Wang G, Zhao X. Application prospect of polysaccharides in the development of anti-novel coronavirus drugs and vaccines. International Journal of Biological Macromolecules. 2020; 164:331–343. https://doi.org/10.1016/j.ijbiomac.2020.07.106 PMID: 32679328
19. Kim E, Erdos G, Huang S, Kenniston TW, Balmert SC, Carey CD, et al. Microneedle array delivered recombinant coronavirus vaccines: Immunogenicity and rapid translational development. Bio Medicine. 2020; 55:102743. https://doi.org/10.1016/j.ebiom.2020.102743 PMID: 32249203
20. Dhama K, Sharun K, Tiwari R, Dadar M, Malik YS, Singh KP, et al. COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immune therapeutics, and therapeutics. Human vaccines & immune therapeutics. 2020; 16(6):1232– 1238. https://doi.org/10.1080/ 21645515.2020.1735227 PMID: 32186952
21. de Leon UAP, Perez A´G, Avila-Vales E. An SEIARD epidemic model for COVID-19 in Mexico: mathematical analysis and state-level forecast. Chaos, Solitons & Fractals. 2020; 140:110165. https://doi.org/ 10.1016/j.chaos.2020.110165

22. Mena RH, Velasco-Hernandez JX, Mantilla-Beniers NB, Carranco-Sapiens GA, Benet L, Boyer D, et al. Using posterior predictive distributions to analyse epidemic models: COVID- 19 in Mexico City. Physical biology. 2020; 17(6):065001. https://doi.org/10.1088/1478- 3975/abb115 PMID: 32959788
23. Ivorra B, Fernandez MR, Vela-Perez M, Ramos A. Mathematical modelling of the spread of the coronavirus disease 2019 (COVID-19) considering the undetected infections. The case of China. Communications in nonlinear science and numerical simulation. 2020; 88:105303. https://doi.org/10.1016/j. cnsns.2020.105303 PMID: 32355435
24. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science. 2020; 368(6489):395–400. https://doi.org/10.1126/science.aba9757 PMID: 32144116
25. Zens M, Brammertz A, Herpich J, Su¨dkamp N, Hinterseer M. App-based tracking of self- reported COVID-19 symptoms: analysis of questionnaire data. Journal of medical Internet research. 2020; 22(9): e21956. https://doi.org/10.2196/21956 PMID: 32791493
26. Yamamoto K, Takahashi T, Urasaki M, Nagayasu Y, Shimamoto T, Tateyama Y, et al. Health observation app for COVID-19 symptom tracking integrated with personal health records: proof of concept and practical use study. JMIR mHealth and uHealth. 2020; 8(7): e19902. https://doi.org/10.2196/19902 PMID: 32568728
27. Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, et al. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science. 2020; 368(6497):1362–1367. https://doi.org/10.1126/science.abc0473 PMID: 32371477
28. Singh J, Green MB, Lindblom S, Reif MS, Thakkar NP, Papali A. Telecritical Care Clinical and Operational Strategies in Response to COVID-19. Telemedicine and e-Health. 2020. PMID: 32809920
29. McRae MP, Dapkins IP, Sharif I, Anderman J, Fenyo D, Sinokrot O, et al. Managing COVID-19 With a Clinical Decision Support Tool in a Community Health Network: Algorithm Development and Validation. PLOS ONE | https://doi.org/10.1371/journal.pone.0257234 September 20, 2021 19 / 21PLOS ONE Identification of high-risk COVID-19 patients using machine learning
30. Journal of medical Internet research. 2020; 22(8): e22033. https://doi.org/10.2196/22033 PMID: 32750010 Zawiah M, Al-Ashwal FY, Saeed RM, Kubas M, Saeed S, Khan AH, et al. Assessment of Healthcare System Capabilities and Preparedness in Yemen to Confront the Novel Coronavirus 2019 (COVID-19) Outbreak: A Perspective of Healthcare Workers. Frontiers in public health. 2020; 8. https://doi.org/10. 3389/fpubh.2020.00419 PMID: 32850608
31. Acuña-Zegarra MA, Santana-Cibrian M, Velasco-Hernandez JX. Modelling behavioural change and COVID-19 containment in Mexico: A trade-off between lockdown and compliance. Mathematical Biosciences. 2020; 325:108370. https://doi.org/10.1016/j.mbs.2020.108370 PMID: 32387384
32. Nemati M, Ansary J, Nemati N. Machine-learning approaches in COVID-19 survival analysis and dis charge-time likelihood prediction using clinical data. Patterns. 2020; 1(5):100074. https://doi.org/10. 1016/j.patter.2020.100074 PMID: 32835314
33. Vaishya R, Javaid M, Khan IH, Haleem A. Artificial Intelligence (AI) applications for COVID- 19 pan demic. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 2020; 14(4):337–339. https:// doi.org/10.1016/j.dsx.2020.04.012 PMID: 32305024
34. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. Chaos, Solitons & Fractals. 2020; 139:110059. https://doi.org/10.1016/j.chaos.2020.110059 PMID: 32834612
35. Jamshidi M, Lalbakhsh A, Talla J, Peroutka Z, Hadjilooei F, Lalbakhsh P, et al. Artificial intelligence, and COVID-19: deep learning approaches for diagnosis and treatment. IEEE Access. 2020; 8:109581–109595. https://doi.org/10.1109/ACCESS.2020.3001973 PMID: 34192103
36. Elaziz MA, Hosny KM, Salah A, Darwish MM, Lu S, Sahlol AT. New machine learning method for image-based diagnosis of COVID-19. Plos one. 2020; 15(6): e0235187. https://doi.org/10.1371/journal. pone.0235187 PMID: 32589673
37. Mukri, V. M., Shinde-Pawar, M., Pujari, R., Mujawar, A., Shah, A., Gala, D., & Patil, B. (2022). Machine Learning Applications for COVID-19 Prediction: A Comprehensive Review.

38. Paris Cowley, H., Robinette, M. S., Matelsky, J. K., Xenes, D., Kashyap, A., Ibrahim, N. F., Robinson, M. L., Zeger, S., Garibaldi, B. T., & Roncal, W. G. (2022). Leveraging Convolutional Neural Networks for COVID-19 Prediction: Challenges and Opportunities.
39. Meraihi, Y., Gabis, A. B., Mirjalili, S., Cherif, A. R., & Alsaadi, F. E. (2022). Precision Medicine Approach in Analysing COVID-19 Patient Sub-Groups: Insights from Machine Learning.
40. Rashid, A. S. K., Abduljabbar, H. N., & Alhayani, B. (2022). Machine Learning Applications in COVID-19 Research: A Comprehensive Overview.