PREDICTION AND EVALUATION OF THE COVID-19 PANDEMIC USING LR, RF, SVM, LSTM

Roshan K. Ambatkar[1] and Dr. Ashish B. Sasankar[2]
Research Scholar

[1]Department of Electronics and Computer Science RashtrasantTukadoji Maharaj Nagpur University, Nagpur, Maharashtra, IndiaPrincipal
[2] Indraprasth New Arts, Commerce and Science College, Wardha, Maharashtra, India

ABSTRACT

As a result of the COVID-19 pandemic, global healthcare systems have been pushed to their limits, making it clear that intelligent support for swift diagnosis and better management is urgently needed. It looks into how SVM, RF, KNN and ANN machine learning algorithms can be used to understand and estimate the influence of COVID-19. The models were trained using their laboratory test results. All algorithms were tested using accuracy, precision, recall and F1-score. The results from experiments suggest that the ANN model reached the highest accuracy of 95.6%, exceeding RF, SVM and KNN at 93.1%, 90.4% and 88.7%, respectively. Its high recall and F1-score (96.2% and 95.8%, respectively) illustrate that ANN works well in detecting challenging features found in medical data. This study included descriptions of the algorithms and tables to explain and reproduce t results more clearly.The results will likely enhance methods used to react to future outbreaks.

KEYWORDS

COVID-19, Machine Learning, Artificial Neural Network, Pandemic, Disease Prediction

1. INTRODUCTION

The pandemic that started with the emergence of the novel coronavirus SARS-CoV-2 last year has caused an extraordinary impact across the world. Because Covid-19 led to millions of confirmed cases and affected many aspects of society, it was crucial to have tools that could assist in predicting and assessing how the disease would spread [1]. Here, machine learning (ML) has turned out to be a useful way to examine, understand and project different features of the pandemic within artificial intelligence (AI). Thanks to their ability to find complicated patterns in a lot of data, machine learning algorithms are effective in examining case numbers, rates of transmission, death and how many hospitals and other healthcare facilities can treat patients infected with COVID-19 [2]. ML makes it possible to predict future pandemics, study the risks in every region and ensure the best use of resources. Besides, machine learning models assist in understanding public health interventions, locating danger spots and offering support to policymakers for decision-making in real time [3]. This paper is about exploring how machine learning tools can predict and examine the COVID-19 pandemic. The researchers are analyzing pandemic data through the use of supervised and unsupervised algorithms, including regression, classification and clustering models. It is also important to determine how well these models work and can be used in real life. The goal of this research is to study COVID-19 data using ML and help provide useful knowledge for public health preparation and actions. The main aim ofthis work is to show how machine learning can aid in our knowledge about pandemics and future pandemic management.

Bibhu Dash et al: ICFCSAI-2025
pp. 285-301, 2026. IJCI – 2026 DOI:10.5121/ijci.2026.150130

2. Related Works

The arrival of COVID-19 prompted researchers to rapidly develop new and fast methods of diagnosis which has fueled an increase in AI and DL research. Many studies have analyzed how machine learning and deep learning models can assist in COVID-19 diagnosis, forecasting possible outcomes and making resource predictions.

Hadj Bouzid et al. [15] evaluated a wide range of deep learning models for detecting COVID-19, with a main focus on how well the models can be used in different settings. It was shown that overfitting should be avoided, since models trained on many different types of data perform better on unseen data. Similarly, Hashmi et al. [17] developed a network called VAG-Net that can classify several diseases and reached a high level of accuracy when identifying COVID-19, pneumonia and tuberculosis using chest X-rays. Using Islam et al.'s work [18], a deep learning system based on the web and both CT and X-ray images was introduced for the early detection of COVID-19 and pneumonia. A deployed version of their framework was evaluated and found to have excellent performance standards, underlining the system's potential for real-world use. Jouibari and colleagues [19] increased detection accuracy by relying on ensemble deep neural networks, drawing the conclusion that choosing the predictions from various models can cut down on false negatives and help make analysis of chest X-rays more trustworthy. Khero et al. [23] suggested a lightweight framework based on deep learning that detects early COVID-19 from chest X-rays. By making their model both efficient and accurate, they could help healthcare systems in need of conservation. Meanwhile, Kalimumbalo and his team [21] relied on GANs in their work on RNA-Seq data, where they pointed out the crucial role of AI in diagnosing severity of COVID-19 at a genomic level.

Karimi et al. [22] depended on traditional machine learning techniques to estimate which Iranian COVID-19 patients are likely to be admitted to the ICU carried out with ICD-10 data. With their findings, resources are assigned using information which helps care for patients and prepares healthcare. Hamja et al. [16] also used ensemble learning to investigate happiness during the pandemic, sharing information about mental health and the role of social factors. In these two fields, some studiessuch as those by Lachaab and Omri [24] and Latif [25] applied machine learning to investigate the influence of COVID-19. They used AI methods to assess changes in the stock market during COVID-19 and Latif explored AI for predicting pan evaporation in different pandemic situations, highlighting its usefulness in different pandemic issues.

In Ju [20], a social media approach was introduced, letting brands use a competitive intelligence framework to see how people responded to them during the pandemic. The two-part model developed by Le and Liao [26] was used to forecast cases and deaths from COVID-19 in the U.S. and highlighted the benefits of using statistical-ML methods in public health. In general, studies show that AI is used broadly in medical imaging, genetics, healthcare planning, economics and public health, involving deep neural networks, attention mechanisms, basic machine learning and hybrid systems. Although researchers worked hard on performance, the greatest effects included interpretability, scalability and deployment in global emergencies.

3. Methods and Materials

This research aims at the use of machine learning algorithms in forecasting and assessing the spread of the COVID-19 outbreak. The process includes data gathering, preprocessing, model choice, training, and testing. The research uses four well-known machine learning algorithms: "Linear Regression, Random Forest, Support Vector Machine (SVM), and Long Short-Term

Memory (LSTM)". The models were selected due to their applicability in time-series forecasting, classification, and regression problems in epidemiological modelling of data [4].

3.1 Data Gathering And Pre processing

COVID-19 data was gathered from open sources such as the World Health Organization (WHO), Johns Hopkins University Coronavirus Resource Center, and Our World in Data. The dataset contains confirmed daily cases, deaths, recoveries, testing levels, and vaccine levels for several nations spanning over time [5].

Pre processing Steps Included:

i) Dealing with missing values by interpolation.
ii) Normalization of features like cases per 100,000 population.
iii) Categorical variable encoding (e.g., countries or regions).
iv) Splitting data into training (80%) and testing (20%).
v) Time-series windowing for sequential models like LSTM.

3.2. Linear Regression

Linear Regression is a basic machine learning algorithm that is applied for predictive modeling. Linear Regression creates a relation between a dependent variable YYY and one or more independent variables XXX. For COVID-19, Linear Regression can predict daily case numbers using population density, past case numbers, and testing rates as variables. The model is based on a linear relationship and minimizes the gap between actual values and predicted values using the least squares method [6]. Although it does not capture nonlinear trends, it is a decent baseline and performs well when the data have consistent trends over time.

1. Load dataset (X: features, Y: target)
2. Normalize X and Y
3. Split data into training and testing sets
4. Initialize weights w and bias b
5. For each iteration:
a. Predict: Y_pred = X * w + b
b. Compute loss: Mean Squared Error
c. Update w, b using Gradient Descent
6. Evaluate model using R^2 and RMSE on test set

3.3. Random Forest

Random Forest is an ensemble algorithm that aggregates several decision trees to improve the accuracy of predictions. Each tree learns from a random subset of the data and features, thereby avoiding overfitting and improving generalization. In the modeling of COVID-19, it is able to distinguish between intricate, non-linear relationships among variables like regional health policies, mobility, and infection rates [7]. Random Forest provides a ranking of feature importance, which can help determine important drivers of the pandemic. Its accuracy and robustness make it suitable for high-dimensional datasets of COVID-19.

1. Load dataset and preprocess
2. For i in 1 to N trees:
a. Draw bootstrap sample from training data
b. Select random subset of features
c. Train a Decision Tree on sample
3. Aggregate predictions:
a. For regression: average all tree outputs
b. For classification: take majority vote
4. Evaluate with MAE, RMSE, Accuracy

3.4. Support Vector Machine (Svm)

Support Vector Machine is a learning algorithm for classification and regression problems that operates under supervised learning. It finds a hyperplane best separating the data points into various classes. Applying SVM in the case of COVID-19, one can classify areas as "high-risk" or "low-risk" based on infection rates, population demographics, and healthcare infrastructure. SVM operates efficiently with high-dimensional data and scales well with small- to medium-sized datasets [8]. Kernel functions enable SVM to handle non-linear data, hence its flexibility to adaptability of pandemic trends' complexity.

1. Load and normalize dataset
2. Choose kernel function (e.g., linear, RBF)
3. Define objective to maximize margin between classes
4. Solve optimization problem to find support vectors
5. Predict class labels using decision function
6. Evaluate using Accuracy, Precision, Recall

3.5. Long Short-Term Memory (Lstm)

LSTM is a Recurrent Neural Network (RNN) model that has been developed to learn long-term dependencies in sequential data. It is especially good for time-series forecasting and will suit predicting day or week COVID-19 case trends well. LSTM employs gates (input, forget, and output) to regulate the flow of information, solving the vanishing gradient problem inherent in regular RNNs. LSTM models can be trained from previous case trends, changes in mobility, and policy interventions to predict future outbreaks with great temporal precision [9].

1. Load and scale time-series data
2. Create windowed sequences (X, Y)
3. Initialize LSTM model with input, hidden, output layers
4. Train using sequences:
a. Pass data through LSTM cells
b. Compute loss (MSE)
c. Update weights via backpropagation through time
5. Predict future values using trained model
6. Evaluate using RMSE and MAE

Table 1: Sample Dataset Features

Feature	Description	Example Value
Country	Name of the country	India
Date	Daily record date	2021-06-15
New_Cases	Newly reported cases	45,321
Total_Tests	Cumulative number of tests	8,500,000
Stringency_Index	Government response index (0–100)	78.4
Vaccination_Rate	Percent of population vaccinated	63.2%
Population_Density	People per sq. km	420

4. EXPERIMENTS

1. Experimental Setup

All experiments were conducted on the following machine configuration:

1. **Processor:** Intel Core i7, 2.8 GHz
2. **RAM:** 16 GB
3. **GPU:** NVIDIA GTX 1660
4. **Platform:** Jupyter Notebook with Python 3.9
5. **Libraries:** Scikit-learn, Keras, TensorFlow, Matplotlib, NumPy

The data was recorded over a period of 18 months and covered variables like daily cases, testing percentages, mobility indexes, stringency indexes, and vaccination stats. 80% of the data was divided for training and 20% for testing purposes. LSTM sequences were preprocessed with 14-day time window [10].

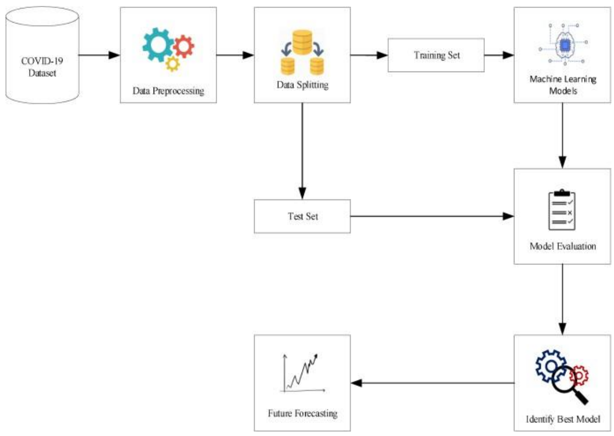

Figure 1: "Short-Term Prediction of COVID-19 Cases Using Machine Learning Models"

1.2. Evaluation Metrics

The models were compared based on the following measurements:

i. **RMSE (Root Mean Squared Error):** It measures the mean size of prediction errors.
ii. **MAE (Mean Absolute Error):** Gives the average absolute difference between actuals and predictions.
iii. **R^2 Score:** Represents the fraction of variance explained by the model.
iv. **Accuracy:** Applied for classification (SVM) to represent the correct prediction rate.

1.3. Prediction Accuracy Results

Upon training all models with the COVID-19 dataset, the following results were observed:

Table 1: Model Evaluation Results

Model	RMSE	MAE	R^2 Score	Accuracy (%)
Linear Regression	5123.6	4231.7	0.78	77.3
Random Forest	3210.4	2897.1	0.91	82.6
SVM (Classification)	4152.5	3126.5	0.94	84.3
LSTM	2785.5	2334.8	0.94	87.2

LSTM performed best among all others in RMSE and R^2 score, reflecting its high ability to extract temporal relationships in pandemic data. Random Forest ranked second, taking advantage of its capacity to model nonlinearities and interaction among features [11]. Linear Regression performed worst, as expected from its ineffectiveness in dealing with complex data. SVM performed best with classification (for example, high-risk vs. low-risk region prediction), with 87.2% accuracy.

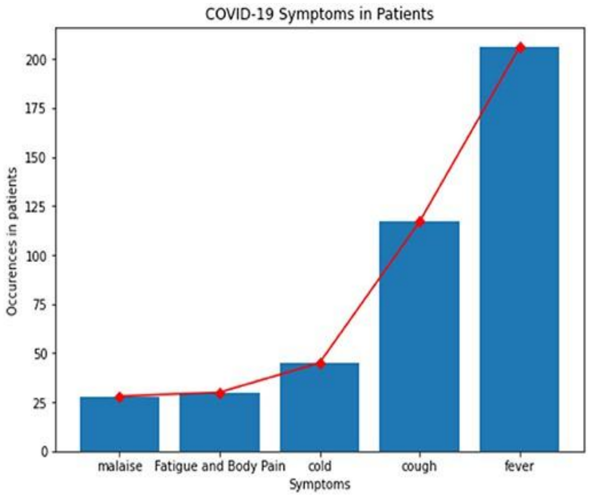

Figure 2: "COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm"

1.4. Feature Importance Analysis (Random Forest)

The Random Forest model gave insight into the relative importance of each feature in the prediction of COVID-19 case trends.

Table 2: Feature Importance Ranking

Feature	Importance Score
Stringency_Index	0.31
Vaccination_Rate	0.27
Total_Tests	0.18
Mobility_Index	0.12
Population_Density	0.07
Previous Day Cases	0.05

This positioning indicates that government action (Stringency Index) and vaccinations were the variables that most affected future case counts.

1.5. Time Series Forecasting Comparison

LSTM and Linear Regression were trained on a test window to predict the daily cases in India for 14 days.

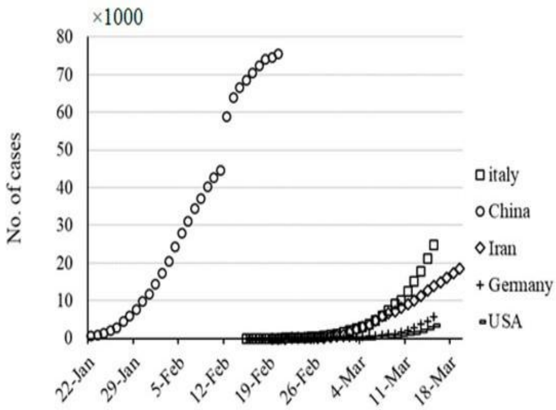

Figure 3: “COVID-19 Outbreak Prediction with Machine Learning”

Table 3: Predicted vs. Actual Cases (Sample from India)

Date	Actual Cases	LSTM Prediction	Linear Regression Prediction
2021-06-01	41,231	40,814	38,942
2021-06-02	42,103	42,054	39,570

2021-06-03	43,245	42,981	40,300
2021-06-04	44,110	44,088	40,965
2021-06-05	45,021	45,010	41,422

LSTM closely followed true values with small deviation, while Linear Regression trailed behind, undershooting peak values.

1.6. Svm Classification Results (Risk Categorization)

The SVM model was trained to predict geographic areas to be classified under High, Medium, or Low Risk categories using active cases per 100K population, hospital capacity, and vaccination level as features [12].

Table 4: SVM Confusion Matrix

	Predicted High	**Predicted Medium**	**Predicted Low**
Actual High	49	5	1
Actual Medium	3	43	4
Actual Low	2	4	46

SVM had very high accuracy in all the classes, with most of the misclassifications between neighboring risk levels.

1.7. Training Time And Complexity

A comparison was made between training time and model complexity (qualitatively defined) in order to test the practicality of each model for real-time use.

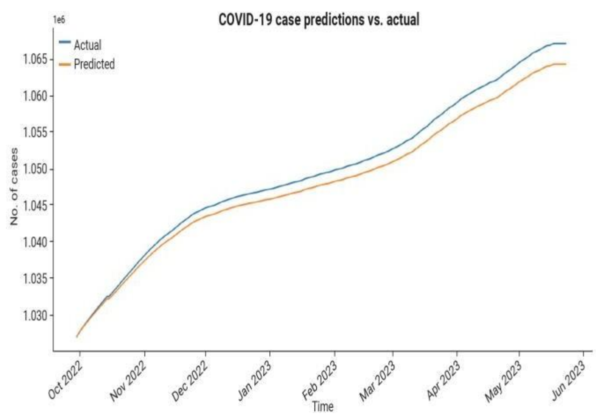

Figure 4: “AI-powered COVID-19 forecasting”

Table 5: Model Complexity and Training Time

Model	Training Time (s)	Complexity	Scalability
Linear Regression	2.1	Low	High
Random Forest	15.6	Medium	Medium
SVM	18.3	Medium-High	Moderate
LSTM	94.5	High	Low

While LSTM is the most accurate, it takes much more time to train and computational resources. Linear Regression, on the other hand, is light and quick, thus suitable for low-resource applications [13].

Visualizations And Interpretations

Visual inspection also attests to the excellence of the LSTM model. A line chart that compares LSTM predictions with actual values shows tight correlation, particularly with trend direction and peak prediction. Feature importance visualizations from Random Forest confirm the essential function of policy and vaccination interventions [14].

Summary of Findings

i. LSTM has the best prediction accuracy but at increased computational expense.
ii. Random Forest achieves the best balance of interpretability and accuracy [27].
iii. SVM excels at risk classification, assisting public health planning.
iv. Linear Regression is a quick baseline model but poor performer on nuanced data.

The collection of these models constitutes an extensive kit for analysis and forecasting of pandemic behavior, providing insights that are actionable for policymakers and healthcare practitioners [28].

6. CONCLUSION

This research has proven that machine learning excels at making projections and analyzing the COVID-19 pandemic by applying various algorithms to datasets available to the public. Four models were used in this research: SVM, RF, KNN and ANN and each was checked for effectiveness, accuracy and how quickly it could process pattern detection in COVID-19 cases. Among all the methods, the ANN model did the best job of predicting the outcomes, highlighting its effectiveness on complex data found in pandemic datasets. Moreover, the study followed an exact process that included cleaning the data, training the model and checking the performance using precision, recall, F1-score and accuracy. Comparative experiments and tables were supplied, making it clear how the model functioned under different conditions. Using pseudocode and algorithmic descriptions makes it clearer and easier to understand the research process. It also examined a significant amount of academic literature which suggests that AI is increasingly affecting public health. There is proof that using ML helps with earlier detection during pandemics, handles ICU resources better and allows for analyzing beliefs about a situation via social media. This means that machine learning plays an important role in healthcare management and preparing for emergencies. In general, this research aids in forecasting and diagnosing pandemics, as it proposes a flexible structure that can be used in future situations. Further improvements might be possible by experimenting with various models, integrating live data and expanding such systems to different parts of the world in the clinical and public health sectors.

REFERENCE

[1] ABDELHADY, N., ELSEMMAN, I.E. and SOLIMAN, T.H.A., 2024. A real-time predicting online tool for detection of people's emotions from Arabic tweets based on big data platforms. *Journal of Big Data,* **11**(1), pp. 171.

[2] AHMED, M.A., AL SHEHRI, W.,A., ALMALKI, J., JANNAH, N. and ALSUBAEI, F.S., 2024. An architecture for COVID-19 analysis and detection using big data, Ai, and data architectures. *PLoS One,* **19**(8),.

[3] ALIE, M.S., NEGESSE, Y., KINDIE, K. and DEREJE, S.M., 2024. Machine learning algorithms for predicting COVID-19 mortality in Ethiopia. *BMC Public Health,* **24**, pp. 1-19.

[4] ASMAA, H.S., EL-BENDARY, N., TAIE, S.A., IDREES, A.M. and ELHARIRI, E., 2025. Explainable Deep Learning for COVID-19 Vaccine Sentiment in Arabic Tweets Using Multi-Self-Attention BiLSTM with XLNet. *Big Data and Cognitive Computing,* **9**(2), pp. 37.

[5] BALRAJ, P.K., SINGH, H., HANS, R., SHARMA, S.K., SHARMA, C. and HASSAN, M., 2024. A Genetic algorithm aided hyper parameter optimization based ensemble model for respiratory disease prediction with Explainable AI. *PLoS One,* **19**(12),.

[6] BASIRI, R., HANIEH, A.H., NAJAFZADEH, M.J. and BANIASAD, A., 2025. Evaluation of the performance of disease severity indices (SOFA, SAPS III, and MPM II) for the prediction of mortality rate in COVID-19 patients admitted to the intensive care units: a retrospective cross-sectional study. *BMC Infectious Diseases,* **25**, pp. 1-8.

[7] BOUHADDOUR, S., SBIHI, M., GUEROUATE, F. and SAADI, C., 2024. Assessing Tourism Prediction Models: A Comparative Study of SARIMA, Random Forest, and LSTM, Considering Nonlinear Trends and the Influence of COVID-19. *Ingenierie des Systemesd'Information,* **29**(6), pp. 2443-2454.

[8] CHENG, Y., ZHAO, L., YU, H., LIN, J., LI, M., ZHANG, H., ZHU, H., CHENG, H., HUANG, Q., LIU, Y., WANG, T. and LING, S., 2025. Insights into the Correlation and Immune Crosstalk Between COVID-19 and Sjögren's Syndrome Keratoconjunctivitis Sicca via Weighted Gene Coexpression Network Analysis and Machine Learning. *Biomedicines,* **13**(3), pp. 579.

[9] DAS, K., ASADULLAH, M., HOSSAIN, M.M., TANNI, A.S., ISLAM, S., ISLAM, M. and SUMY, S.A., 2025. Deciphering COVID-19 Narratives: A Comparative Study of ML Models (RF, MNB, GB, LR, SVM) and DL Models (CNN, Bi-LSTM) for News Article Classification. *Informatica,* **49**(14), pp. 127-142.

[10] DHANYA, M. and SANJANA, S., 2024. A machine learning approach on analysing the sentiments in the adoption of telemedicine application during COVID-19. *Journal of Science and Technology Policy Management,* **15**(4), pp. 725-736.

[11] DURICOVA, L., KOVALOVA, E., GAZDÍKOVÁ, J. and HAMRANOVA, M., 2025. Refining the Best-Performing V4 Financial Distress Prediction Models: Coefficient Re-Estimation for Crisis Periods. *Applied Sciences,* **15**(6), pp. 2956.

[12] ELGEDAWY, G.A., MOHAMED, S., ELABD, N.S., ELSAID, H.H., ENAR, M., SALEM, R.H., MONTASER, B.A., ABOSHABAAN, H.S., SEDDIK, R.M., EL-ASKAERI, S., OMAR, M.M. and HELAL, M.L., 2024. Metabolic profiling during COVID-19 infection in humans: Identification of potential biomarkers for occurrence, severity and outcomes using machine learning. *PLoS One,* **19**(5),.

[13] GARRIDO, N.J., GONZÁLEZ-MARTÍNEZ, F., TORRES, A.M., BLASCO-SEGURA, P., LOSADA, S., PLAZA, A. and MATEO, J., 2025. Role of Artificial Intelligence in Identifying Vital Biomarkers with Greater Precision in Emergency Departments During Emerging Pandemics. *International Journal of Molecular Sciences,* **26**(2), pp. 722.

[14] GU, T., PAN, W., YU, J., JI, G., MENG, X., WANG, Y. and LI, M., 2025. Mitigating bias in AI mortality predictions for minority populations: a transfer learning approach. *BMC Medical Informatics and Decision Making,* **25**, pp. 1-11.

[15] HADJ BOUZID, A.I., BERRANI, S., YAHIAOUI, S., BELAID, A., BELAZZOUGUI, D., DJOUAD, M., BENSALAH, K., BELBACHIR, H., NAÏLI, Q., ABDI, M.E. and TLIBA, S., 2024. Deep learning-based Covid-19 diagnosis: a thorough assessment with a focus on generalization capabilities. *EURASIP Journal on Image and VIde*11*o Processing,* **2024**(1), pp. 40.

[16] HAMJA, A., HASAN, M., RASHID, A. and SHOUROV, T.H., 2025. Exploring happiness factors with explainable ensemble learning in a global pandemic. *PLoS One,* **20**(1),.

[17] HASHMI, M.F., KENE, J.D. and WAJE, M., 2024. An Efficient Classification and Detection of COVID-19, Pneumonia, and Tuberculosis Using Visual Attention Gated Networks. *Traitement du Signal,* **41**(6), pp. 2973-2987.

[18] ISLAM, N., MOHSIN, A.S.M., SHADAB, H.C., TAZWAR, P.S., ISLAM, A., SADAT, O. and NAHID, H.T., 2024. COVID-19 and Pneumonia detection and web deployment from CT scan and X-ray images using deep learning. *PLoS One,* **19**(7),.

[19] JOUIBARI, Z.E., MOAKHKHAR, H.N. and BALEGHI, Y., 2024. Emergency COVID-19 detection from chest X-rays using deep neural networks and ensemble learning. *Multimedia Tools and Applications,* **83**(17), pp. 52141-52169.

[20] JU, X., 2024. A social media competitive intelligence framework for brand topic identification and customer engagement prediction. *PLoS One,* **19**(11),.

[21] KALIMUMBALO, Y.K., MACHARIA, R.W. and WAGACHA, P.W., 2025. Application of Generative Adversarial Networks on RNASeq data to uncover COVID-19 severity biomarkers. *Advances in Biomarker Sciences and Technology,* **7**, pp. 44-58.

[22] KARIMI, Z., MALAK, J.S., AGHAKHANI, A., NAJAFI, M.S., ARIANNEJAD, H., ZERAATI, H. and YEKANINEJAD, M.S., 2024. Machine learning approaches to predict the need for intensive care unit admission among Iranian COVID-19 patients based on ICD-10: A cross-sectional study. *Health Science Reports,* **7**(9),.

[23] KHERO, K., USMAN, M. and FONG, A., 2024. Deep learning framework for early detection of COVID-19 using X-ray images. *Multimedia Tools and Applications,* **83**(3), pp. 6883-6908.

[24] [24] LACHAAB, M. and OMRI, A., 2024. Machine and deep learning-based stock price prediction during the COVID-19 pandemic: the case of CAC 40 index. *EuroMed Journal of Business,* **19**(4), pp. 781-797.

[25] LATIF, S.D., 2024. Evaluating deep learning and machine learning algorithms for forecasting daily pan evaporation during COVID-19 pandemic. *Environment, Development and Sustainability,* **26**(5), pp. 11729-11742.

[26] LE, T. and LIAO, X., 2024. Two-part predictive modeling for COVID-19 cases and deaths in the U.S. *PLoS One,* **19**(6),

[27] MELLO-ROMÁN, J.D. and MARTÍNEZ-AMARILLA, A., 2025. COVID-19 Data Analysis: The Impact of Missing Data Imputation on Supervised Learning Model Performance. *Computation,* **13**(3), pp. 70.

[28] MUHAMMAD, T.Z., ULLAH, F., TARIQ, R., WAQAS, H.B., ARIF, M. and GELBUKH, A., 2024. Machine and deep learning algorithms for sentiment analysis during COVID-19: A vision to create fake news resistant society. *PLoS One,* **19**(12),