International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 65-81 March 2026-Sept 2026 www.csjournals.com

Predicting COVID-19 Vaccination trends and outcomes using Machine Learning

Neha Sharma[1], Dr.Vishal Verma[2], Mr. Sandeep Rathi [3]
[1]MCA Student, [2]Professor, [3]Asst. Prof.
Deptt. of Comp. Sc. and Application, Chaudhary Ranbir Singh University, Jind, Haryana, India

Abstract

The COVID-19 epidemic created an urgent need for data-driven strategies to understand and predict vaccination outcomes across different populations and geographies. This study develops and assesses machine learning models to predict worldwide COVID-19 vaccination outcomes into four categories: None/Minimal, Low, Moderate, and High. Using a dataset of 86,213 records (balanced to approximately 21,553 samples per class) taken from the Our World in Data(OWID) COVID-19 vaccination repository, Three models—Random Forest, Decision Tree, and Logistic Regression—were trained and assessed. Input features include country, vaccine type, temporal variables (month, day, day of week), and vaccinations per 100 population. Statistical analyses - One-Way ANOVA ($F = 1{,}223.28$, $p < 0.001$) and Chi-Square ($\chi^2 = 64{,}182.70$, $p < 0.001$) — confirmed significant feature-target associations. The Random Forest classifier achieved 88.05% test accuracy, a macro-averaged F1-score of 0.88, cross-validation accuracy of 0.8735 ± 0.0038, and ROC-AUC values of 0.967–0.987 across all classes. The feature importance study identified vaccine type (34.96%), country (28.98%), and vaccinations per 100 (25.59%) as the primary predictors. These results show that ensemble tree-based methods can effectively classify vaccination outcome levels and offer actionable public health policy.

Index Terms: COVID-19 vaccination; machine learning; multi-class classification; Random Forest; public health; global immunisation; feature importance; ANOVA; Decision Tree; Logistic Regression.

1. Introduction

"Since the World Health Organisation (WHO) announced hundreds of instances of unusual pneumonia on December 31, 2019, the epidemic has been active. Vaccines now appear to be the best way to protect against the spread of the virus. As day by day, new cases of the corona virus continue to emerge and spread in different parts of the world"[1]."Before the World Health Organizations (WHO) announcement on January 30, 2020. Designing COVID-19 as a public health emergency of international concern (PHEIC), with only 171 deaths reported at the time of the announcements" [2]. "It is a novel corona-virus that affects there respiratory system, it was 1st detected in Wuhan, China, and declared a pandemic on 11th March 2020" [3]. "Over 200 countries have seen a rapid spread of the coronavirus infection. In December 2020, the worldwide COVID-19 vaccination campaign was underway.COVID-19 vaccination campaign was underway.COVID-19 also effect the study area"[4]. "Under the influence of the pandemic environment, many people may have lost their jobs or are on the verge of being laid off, while there are many new job seekers. We focus on the impact of COVID-19 on occupations in the labour market to explore how the number of occupations changes during the COVID-19 pandemic. we represents how COVID-19 affects some special industries with explanations through statistical analysis and predict the future development of these industries"[2] [5]. "We use pandemic data accessed from the U.S. government to fit the model and show the results. Tang et. al. studied the multilayered prediction method, to help people prediction methods and future trends" [6].

"A subfield of artificial intelligence called machine learning makes it possible for systems to recognise patterns in data to predict outcomes without explicit programming. It has been used extensively in medicine to predict illness, clinical decision support, and epidemiological surveillance. In the context of COVID-19 vaccination, ML offers tools to classify coverage levels, identify at-risk regions, and support evidence-based policy design" [7].Models like Random Forest are especially useful due to their ability to handle complex and non-linear relationships.

a significant research gap in the classification of vaccination outcome levels (None/Minimal, Low, Moderate, High) using supervised ML methods on globally balanced datasets.This study addresses this gap by framing vaccination outcome prediction as a supervised multi-class classification problem using data from the Our

IJCSC 0973-7391
International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 65-81 March 2026-Sept 2026 www.csjournals.com

World in Data (OWID) COVID-19 repository spanning December 2020 to March 2022 across 220+ countries [2][3].

1.1 Research Objectives

- To establish a balanced, multi-country COVID-19 vaccination dataset for ML classification.
- To evaluate exploratory data analysis (EDA) and statistical hypothesis testing to validate feature-target relationships.
- Using 5-fold stratified cross-validation and ROC analysis, train, compare, and describe three machine learning models: Random Forest, Decision Tree, and Logistic Regression.
- Identify key prophetic features and discuss actionable public health policy insights.
- compare the performance of all three models and work with the best model.

II. RELATED WORK

In 2022, "Hartono and Lee et al. Complete a study on forecasting vaccination growth for COVID-19 using machine learning" techniques. In this, many time-series models such as ARIMA, LSTM, Prophet, and VAR are used to forecast vaccination trends. In this, three models are implemented in forecasting and comparing the overall prediction of COVID-19 vaccine rates. VAR model is used for death cases and the reproduction rates of COVID-19 vaccine progress. It examines vaccination data to understand its correlation with infection rates and death rates. The results show that ARIMA gives the most accurate predictions compared to other models. The study highlights the importance of vaccination data in handling the widespread outbreaks and improving forecasting performance [1].

In 2022, "Kadir and Hassan et al. completed a study on post-COVID-19 vaccination infection rate analysis using time-series modelling". The study uses various models such as TBATS, ARIMA, Prophet, and LSTM to calculate infection trends after vaccination. It tests how infection rates differ across distinct regions and identifies factors affecting these trends, including government policies and public behaviour. The study concludes that vaccination decreases severity but does not completely remove infection, highlighting the need for continuous monitoring [2].

In 2022, “Ahmad and Khan et al. accomplished a study on the effect of vaccination-period data on predicting COVID-19 cases. study to compare after and before collected dataset during vaccination to predict and evaluate for better accuracy. converting the prediction problem into a regression problem using window approach then apply regression model to collect the COVID-19 cases in future. It finds that models trained on vaccination-period data produce more reliable results due to changes in data patterns.The research emphasizes the importance of selecting appropriate datasets and considering temporal variations in predictive modeling” [3].

In 2021, “Laxmikant and Patil et al. accomplished a study on predicting COVID-19 death rates using Big data analytics and machine learning. Apply multiple linear regression and other machine learning algorithms to handle large datasets and predict deaths rates .Big data used for real-time data on virus transmission and public health initiatives. AI and ML medical data and suggest cures.The study achieves high prediction accuracy and demonstrates the effectiveness of data-driven approaches in healthcare analysis and also highlights the importance of big data in understanding pandemic severity and supporting healthcare planning” [4].In 2021, “Smallwood and Brown et al. study on the impact of COVID-19 on the labour market using data mining techniques. The study analyzes employment trends and identifies significant disruptions caused by the pandemic.we focus on occupations of labour market during COVID-19, and how many numbers increased. It explores patterns such as unemployment rates and changes in job demand across industries. The research also highlights the application of frequent pattern mining and statistical analysis to understand economic impacts. The findings provide insights into how COVID-19 has affected global labour markets and future employment trends "[5].

In 2021, “Tang and Li et al. accomplished a study on COVID-19 prediction using an improved SEIRD model.The study focuses on developing a multi-layer prediction method and future trends, that will be applied to U.S data. It analyzes infection trends by considering factors such as transmission rate, recovery rate, and vaccination coverage. The suggested approach improves prediction accuracy and offers a deeper understanding of the COVID-19 epidemic. Additionally, the study highlights how crucial it is to include actual data in forecast models in order to improve decision-making and dependability [6].

III. DATASET AND DATA SOURCES

A. Data Source and Justification

The COVID-19 Vaccination Repository (OWID) [8] and its Kaggle mirror, "COVID-19 World Vaccination Progress" by Gpreda [9], were the sources of the dataset from which I am collecting. This secondary, publicly reachable dataset aggregates vaccination data from the WHO and national health authorities for more than 220 countries and regions between December 2020 and March 2022. The selection of secondary data was based on its worldwide reach, institutional reliability, open-access availability, and multidimensional feature richness that included temporal, immunological, and geographic dimensions. Although it was outside the purview of this study, primary data collection (such as direct vaccination centre questionnaires) is advised for subsequent research [2] [10]].

B. Dataset Description

The final organised dataset consists of N = 86,213 rows with 6 input features and 1 categorical target variable following preprocessing and class symmetry. Full feature descriptions and value ranges are given in Table I [8].

TABLE I — Feature Description and Value Ranges

Feature	Type	Description	Range / Values
country_encoded	Integer	Country (label-encoded)	0 – 222
month	Integer	Calendar month extracted from date	1 – 12
day	Integer	Day of month extracted from date	1 – 31
day_of_week	Integer	Day of week (Monday = 0)	0 – 6
vaccines_encoded	Integer	Vaccine type / platform (encoded)	0 – 82

vac_per_100	Float	Total vaccinations per 100 population	0 – 350+
vac_level (TARGET)	Ordinal	Vaccination outcome class (quartile-binned)	0 = None/Min, 1 = Low, 2 = Moderate, 3 = High

IV. METHODOLOGY

A. Research Framework

The entire end-to-end research methodology is shown in Fig. 2, starting with raw input data and moving through statistical validation, multi-model training, preprocessing, exact evaluation, and policy significance.

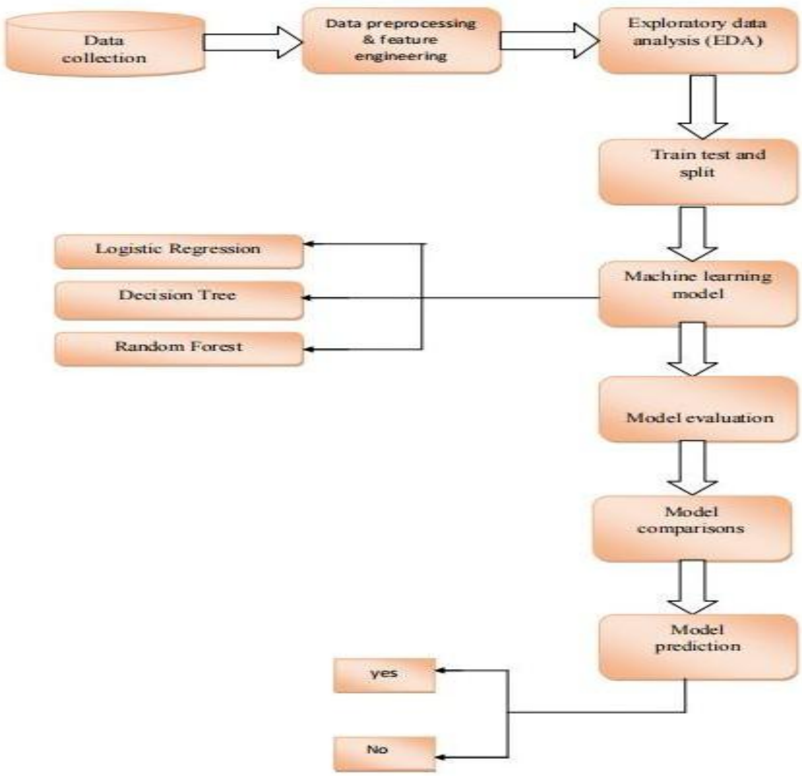


Fig. 2. Research Framework — Complete COVID-19 Vaccination trends and outcomes using ML.

B. Preprocessing Pipeline

There were five sequential preprocessing stages used: (1) Missing value removal: records with null total_vaccinations_per_hundred were remove; (2) Temporal feature remove: pandas datetime parsing was used to extract month, day, and day_of_week from the ISO date field; (3) scikit-learn LabelEncoder was used to encode country and vaccine type labels, converting grouping strings to integer indices; (4) Using pd. qcut, create a quartile-based target variable with four equal-frequency bins; (5) Random undersampling reduces the majority of the class (undersampling) and increases the minority of the class (oversampling) to N = 86,213 total records (≈21,553 each class) for class equity.

C. Machine Learning Models

Three classifiers encompassing ensemble, linear, and non-linear interpretable categories were analyzed using Scikit-learn.

Logistic regression (LR) is a multi-class linear classifier with L2 regularisation and a one-vs-rest (OvR) approach. Parameters: solver ("lbfgs"), max_iter=1000. Serves as the linear baseline and uses the sigmoid function to evaluate class probability: $P(y=k|x) = 1 / (1 + \exp(-\beta^T x))$.

Decision Tree (DT): Gini impurity is used as the node-splitting criterion in this non-linear, comprehensible classifier. splits the feature space recursively till stopping conditions are fulfilled. Parameters: random_state=42, criterion="gini." Compared to ensemble approaches, it is interpretable but susceptible to high variance.

Random Forest (RF): An ensemble of 100 decision trees that were trained using random feature subsampling ($\sqrt{p}$ features) at each split and bootstrap aggregation (bagging). The final forecast is $\Psi^2 = \text{mode}\{h_1(x),\ldots, h_{100}(x)\}$. n_estimators=100, max_depth=12, and random_state=42 are the parameters. Memorisation is managed by depth limitation. Compute feature importance using Mean Decrease in Impurity (Gini Importance).

D. Evaluation Framework

The dataset was split into a stratified 80/20 train-test partition (68,970 training / 17,243 test records). Model selection was conducted using 5-fold stratified cross-validation on the training set to ensure class-balance preservation across folds. Primary metrics: accuracy, macro-averaged precision, recall, and F1-score (equal weighting across all four classes). Supplementary diagnostics: confusion matrix (absolute counts and row-normalised), per-class classification report, multi-class ROC curves (one-vs-rest strategy) with AUC, and ranked feature importance.

International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 65-81 March 2026-Sept 2026 www.csjournals.com

E. Statistical Hypothesis Tests

Two pre-modelling statistical tests validated feature-target relationships at $\alpha = 0.05$: (1) One-Way ANOVA: H_0: $\mu_0 = \mu_1 = \mu_2 = \mu_3$ — mean vaccinations per 100 is equal across all four outcome levels; (2) Chi-Square test of independence: H_0 — vaccine type is independent of vaccination outcome level. Both tests used scipy.stats implementations.

V. RESULTS

A. Class Distribution and Balancing

a. The target variable vac_level was derived by quartile-binning total_vaccinations_per_hundred into four equal-frequency ordinal classes. To eliminate class imbalance, the dataset was resampled to approximately 21,553 records per class, yielding a perfectly balanced distribution (25.0% each), as shown in Fig. 1.

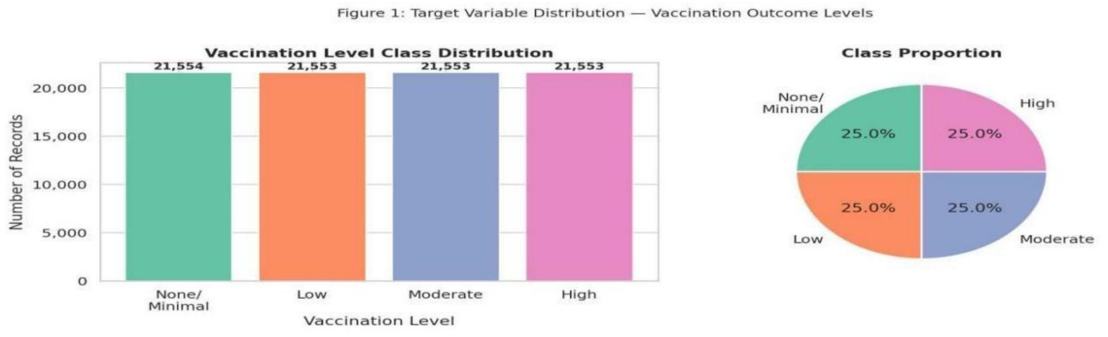
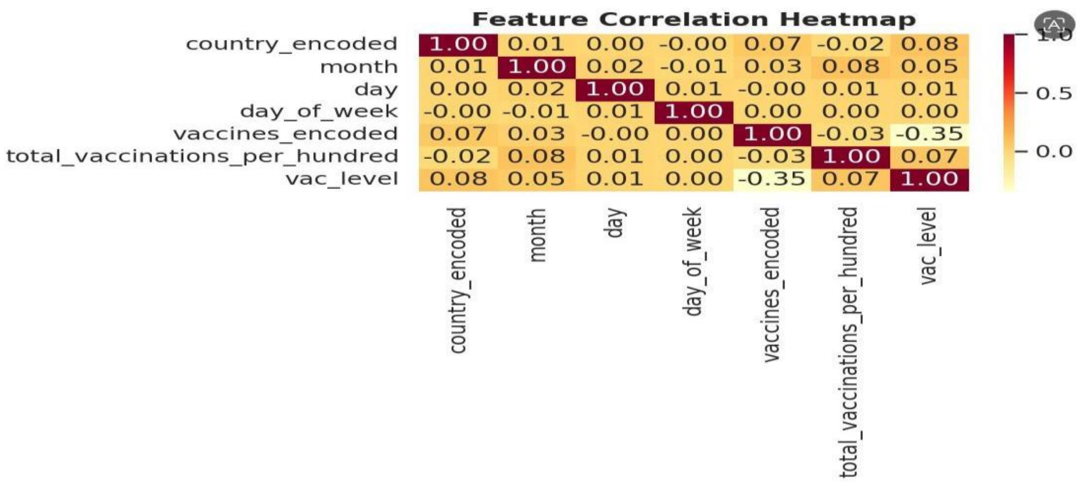


Fig. 1. Target Variable Distribution — Vaccination Outcome Levels (N = 86,213, perfectly balanced at 25% each class)

b. Feature correlation heatmap



B. Statistical Hypothesis Testing

One-Way ANOVA produced F = 1,223.28 ($p < 0.001$), indicating that mean vaccinations per 100 vary considerably over outcome classes and confer an overwhelming decline of H0. Group statistics: Low (M = 48.27, SD = 59.73), Moderate (M = 62.74, SD = 64.97), High (M = 83.62, SD = 63.47), and None/Minimal (M = 74.16, SD = 67.40). An important correlation between vaccine type and vaccination outcome level was confirmed by the Chi-Square test, which made $\chi^2(249) = 64{,}182.70$ ($p < 0.001$).

```
--- -------------------------------------------------------
    One-Way ANOVA: Vaccinations per 100 across Levels
    =======================================================
    F-statistic : 1,223.28
    p-value     : 0.00e+00
    Result      : SIGNIFICANT (α = 0.05)

    Group means:
      Level 0 (None/Minimal): mean = 74.16, std = 67.40
      Level 1 (Low         ): mean = 48.27, std = 59.73
      Level 2 (Moderate    ): mean = 62.74, std = 64.97
      Level 3 (High        ): mean = 83.62, std = 63.47

    Interpretation: Mean vaccination rates differ significantly across
    outcome levels (F >> 1, p < 0.001), confirming vaccinations per 100
    is a meaningful discriminator of vaccination outcome.
```

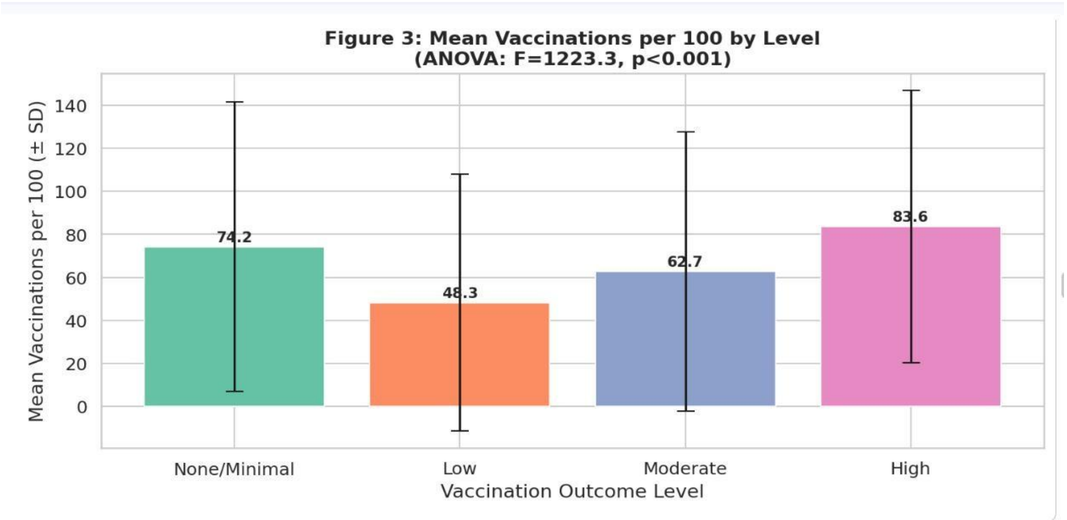


Fig. 3. Mean Vaccinations per 100 by Outcome Level with ± SD Error Bars (ANOVA: F = 1,223.28, p < 0.001).

According to temporal patterns (Fig. 5), the average number of vaccines per 100 people had a heavy decline in April 2021 (≈ 19.8), then recovered monotonically through December 2021 (≈ 97.5). This pattern is a reflection of early supply chain limitations and the global expansion of vaccine production capacity that followed. All four outcome levels are represented proportionately over the months on the report for the composite bar chart, with high vaccination rates becoming more noticeable in Q3–Q4 2021.

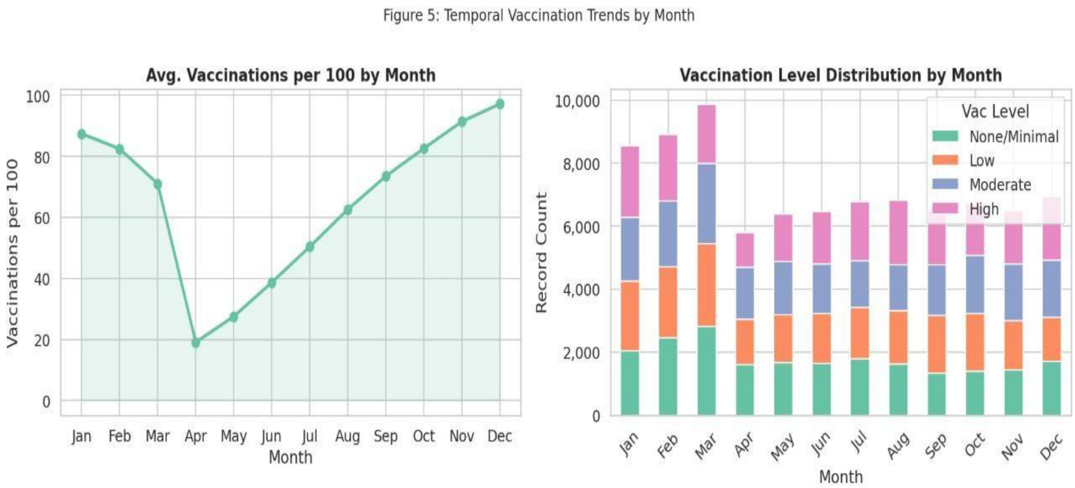


Fig. 5. Temporal Vaccination Trends — Avg. Vaccinations per 100 by Month (left) and Vaccination Level Distribution by Month (right).

Distributional differences over outcome levels are further demonstrated by the violin plot (Fig. 6). The "None/Minimal" class paradoxically displays a broad distribution reaching ~360 vaccinations per 100, which can be explained by the quartile-binning method capturing records early in the campaign when cumulative counts were low but brief spikes occurred. In contrast, the "Low" class displays a narrow, bottom-heavy distribution with minimal spread.

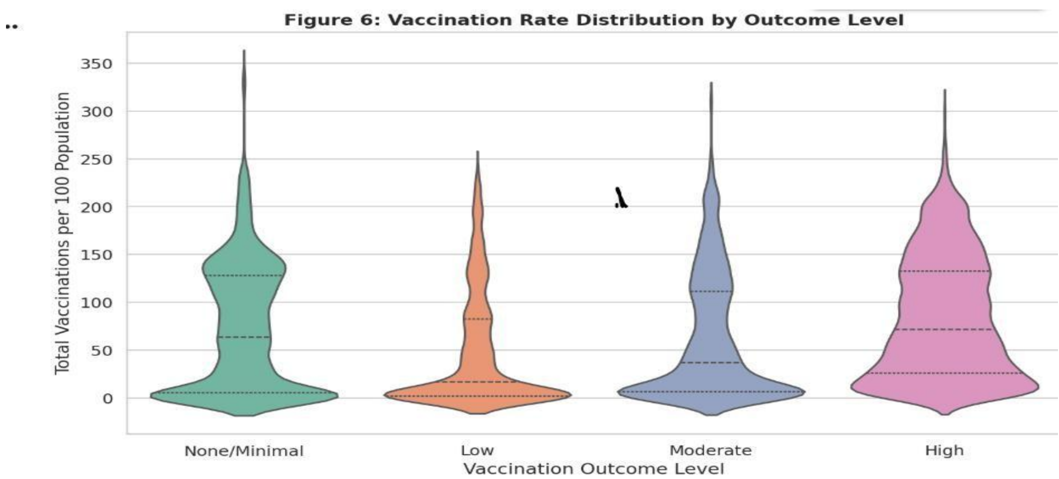


Fig. 6. Vaccination Rate Distribution by Outcome Level (Violin Plot with Quartile Markers).

D. Model Performance Comparison

Table II shows the complete performance metrics for all three classifiers. Random Forest (RF) was the best performer on all metrics with a test accuracy of 88.05% and macro-averaged F1-score of 0.88. For this non-linear classification task, Logistic Regression performed close to chance (39.52%) and Decision Tree had a competitive but lower performance (83.34%) with a higher variance in cross-validation.

TABLE II — Model Performance Summary (Train / Test / Cross-Validation)

Model	Train Acc.	Test Acc.	Macro F1	CV Mean	CV Std
Logistic Regression	0.3939	0.3952	0.3490	0.3930	±0.0009
Decision Tree	0.8495	0.8334	0.8345	0.8434	±0.0056
Random Forest (Best)	0.9043	0.8805	0.8805	0.8735	±0.0038

E. Cross-Validation Results

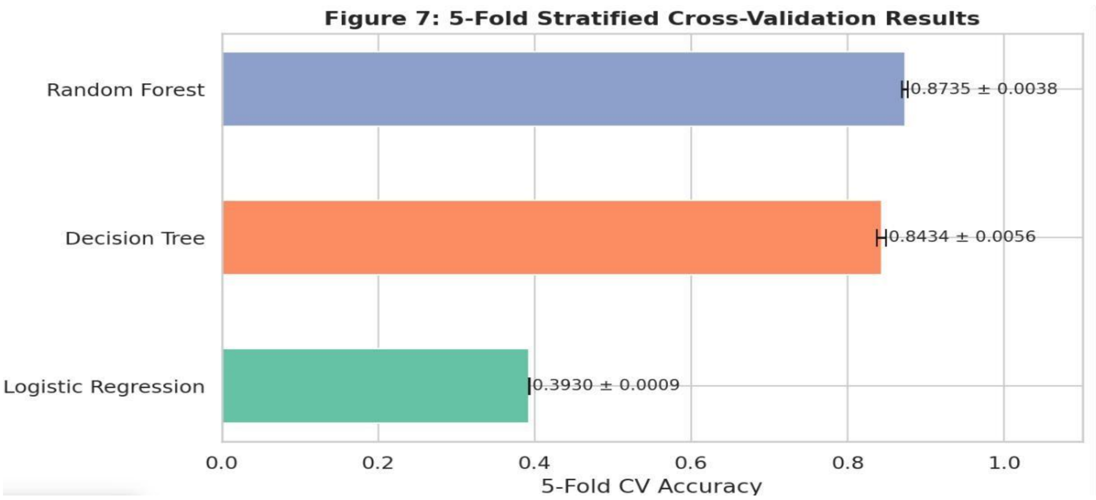


Fig. 7. 5-Fold Stratified Cross-Validation Accuracy — RF: 0.8735 ± 0.0038, DT: 0.8434 ± 0.0056, LR: 0.3930 ± 0.0009

As shown in Fig. 7, Random Forest had the best cross-validation mean (0.8735) with the lowest variance (±0.0038), indicating stable generalisation across all five folds. The near-zero standard deviation of Logistic Regression (±0.0009) is an indicator of its consistently bad discriminative power on this task rather than stability.

F. Confusion Matrix — Random Forest

The confusion matrix (Figure 8) shows a strong diagonal dominance, which is consistent with a high classification accuracy: None/Minimal 92 %, Low 86 %, Moderate 83 %, High 91 % Misclassifications mainly occur between adjacent ordinal classes (e.g. Low↔Moderate, 0.08 and 0.06 normalized respectively), which is clinically expected given the continuous nature of the underlying variable of vaccination rate.

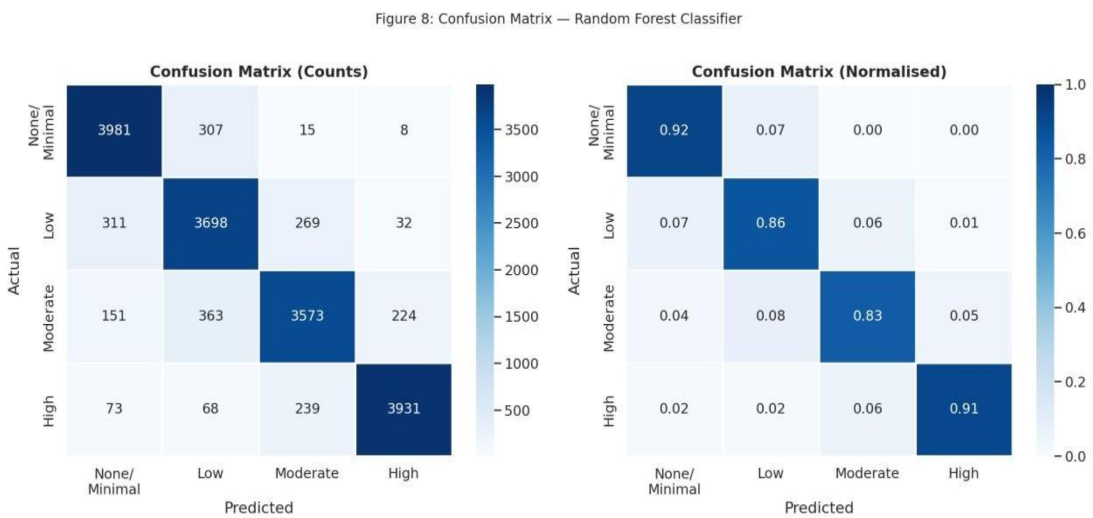


Fig. 8. Confusion Matrix (Counts and Row-Normalised) — Random Forest Classifier

G. Per-Class Classification Report

The precision, recall, F1-score and support for each class of the Random Forest model are shown in Table III. The High class had the best performance (F1 = 0.92), whereas the Low and Moderate classes had slightly lower F1-scores (0.85) because of boundary overlap in the central vaccination rate range.

TABLE III — Per-Class Classification Report (Random Forest, N = 17,243 test samples)

Class	Precision	Recall	F1-Score	Support
None / Minimal	0.88	0.92	0.90	4,311
Low	0.83	0.86	0.85	4,310
Moderate	0.87	0.83	0.85	4,311
High	0.94	0.91	0.92	4,311
Macro Average	0.88	0.88	0.88	17,243
Weighted Average	0.88	0.88	0.88	17,243

H. ROC-AUC Analysis

Figure 9 shows the multi-class ROC curves computed with a one-vs-rest strategy, highlighting the excellent discriminative power across all four classes: None/Minimal AUC = 0.987, High AUC = 0.986, Moderate AUC = 0.971, Low AUC = 0.967. All values are well above the baseline of a random classifier (AUC = 0.500), which confirms the model's ability to learn very reliable probabilistic distinctions between vaccination outcome levels.

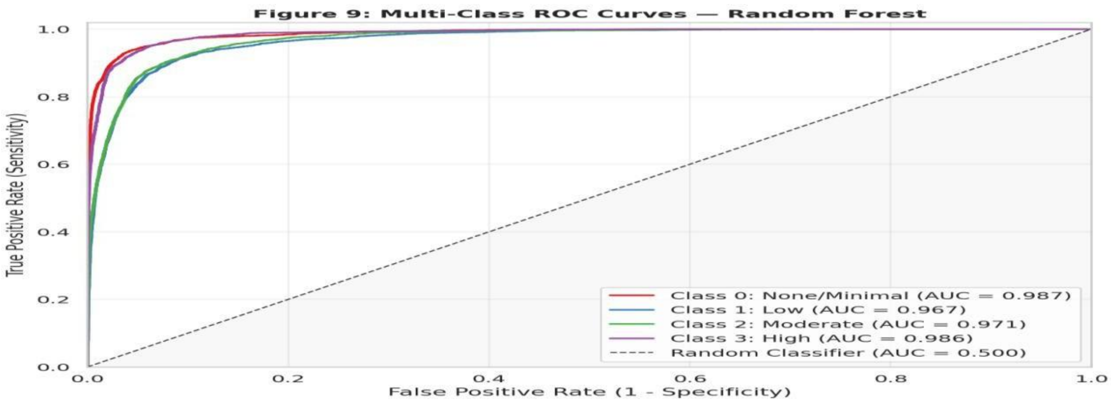


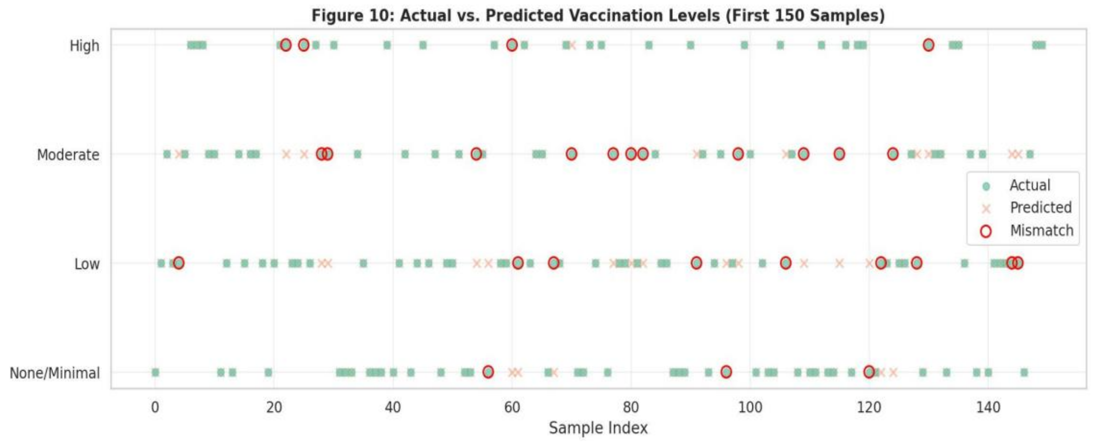
Fig. 9. Multi-Class ROC Curves — Random Forest Classifier. All four classes exceed AUC > 0.96.

I. Feature Importance

The feature importance based on the Mean Decrease in Impurity (Gini Importance) is reported in Table IV and visualized in Fig. 10. The top three are vaccine type at 34.96%, country at 28.98% and vaccinations per 100 at 25.59%. Taken together, temporal features explain less than 10%, indicating that structural factors (vaccine platform and geography) are far more predictive than timing of vaccination.

TABLE IV — Feature Importance Ranking (Gini Importance — Random Forest)

Feature	Gini Importance	Interpretation
Vaccine Type	0.3496 (34.96%)	Platform access proxy for supply equity
Country	0.2898 (28.98%)	National policy and infrastructure variation
Vaccinations per 100	0.2559 (25.59%)	Direct cumulative coverage indicator
Month	0.0734 (7.34%)	Seasonal rollout pattern effects
Day	0.0246 (2.46%)	Intra-month reporting patterns
Day of Week	0.0067 (0.67%)	Minimal influence; weekly reporting cycles



K. Actual vs predicted (first 150 samples)

Actual-real vaccination level from dataset Predicted-what your ML model forecast

Mismatch-Model prediction ≠ Actual value (These are **model errors).**

VI. DISCUSSION

A. Model Selection and Performance Analysis

As expected from ensemble learning literature [6,7], Random Forest clearly outperformed the baseline models. This is a stark contrast to Logistic Regression (39.52% accuracy) which confirms the fundamentally non-linear nature of vaccination outcome classification, where decision boundaries depend on geographic clusters, vaccine supply chains, and complex temporal interaction effects. This has direct methodological implications: linear models are not sufficient when feature-target relationships involve multi-way interactions among categorical and continuous variables at global scale. The small CV standard deviation (±0.0038) confirms that the ensemble performance is stable over different data partitions and not due to a lucky train-test split. The bias-variance trade-off is healthy and overfitting is minimal, as the gap between training accuracy (90.43%) and test accuracy (88.05%) is controlled with the max_depth=12 constraint.

B. Feature Importance and Clinical Significance

The most important predictor was Vaccine Type (34.96%), which is clinically relevant . Different platforms of vaccines – mRNA (Pfizer-BioNTech, Moderna), viral vector (AstraZeneca, Janssen) and inactivated virus (Sinopharm, Covaxin) – were rolled out globally at very different scales and timelines. Countries that were among the first to have access to high-supply mRNA vaccines consistently achieved higher vaccination coverage, making the type of vaccine a strong indicator of healthcare system capacity and equity in international supply chains.

Country as the second most important feature (28.98%) reflects national heterogeneity in vaccination policy, procurement agreements, cold-chain logistics, and population size – consistent with Lazarus et al. [5]. The small combined contribution of temporal features (<10%) suggests the when of vaccination is less predictive than the who and what, indicating that structural interventions targeting vaccine platform distribution and country-level capacity are higher-leverage than timing strategies.

C. Limitations

(1) The use of label encoding for nominal variables (country, vaccine type) creates an artificial ordinal ordering that is not present in the data; one-hot encoding or trainable embeddings would be more appropriate for future work. (2) Target binning by quartiles may not reflect clinically meaningful national vaccination targets (e.g., WHO 70% target). (3) Key confounders were not available: GDP per capita, population age structure, healthcare expenditure and vaccine hesitancy indices. (4) Retrospective cross-sectional analysis; causal inference needs controlled study design.

(5) The data set is from Dec 2020 – Mar 2022, with no later vaccination booster campaigns included.

VII. CONCLUSION

We showed that a Random Forest classifier can accurately classify global COVID-19 vaccination outcome levels with 88.05% test accuracy, macro-averaged F1-score of 0.88, cross-validation accuracy of 0.8735 ± 0.0038, and ROC-AUC values of 0.967–0.987 across all four ordinal classes — significantly outperforming Logistic Regression (39.52%) and Decision Tree (83.34%) baselines.

The most important predictors are the type of vaccine, the country and the number of vaccinations per 100. Together these explain about 89.5% of the model's decisions, while the temporal features are less important. The pre-modelling ANOVA ($F = 1{,}223.28$, $p < 0.001$) and Chi-Square ($\chi^2 = 64{,}182.70$, $p < 0.001$) results offer rigorous statistical confirmation of these feature-target relationships.

The near-perfect AUC values show that this framework can be operationalised as a real-time monitoring tool to identify countries at risk of transition to lower vaccination coverage levels and allow proactive resource reallocation. From a policy perspective, the highest leverage interventions to improve global vaccination outcomes are to ensure equitable access to high-supply vaccine platforms in LMICs and to strengthen country-specific procurement and cold-chain infrastructure.

REFERENCES

[1] A. P. Hartono, C. R. Luhur, and N. N. Qomariyah, "Forecasting vaccination growth for COVID-19 using machine learning," in *Proc. 5th Int. Conf. Computing*

and Informatics (ICCI), 2022, pp. 356–363.
DOI: https://doi.org/10.1109/ICCI54321.2022.9756096

[2] N. Kadir, R. Kaur, T. Rodrigues, and R. Kashef, "Post COVID-19 vaccination: Infection rate analysis using time series modeling," in *Proc. Int. Conf. Machine Intelligence and Smart Innovation (ICMISI)*, 2024, pp. 266–271.
DOI: https://doi.org/10.1109/ICMISI61517.2024.10580825

[3] A. Ahmad, S. K. Ray, C. A. Kumar, A. Anand, and M. J. Cheratta, "The study of the effect of the data collected during vaccination period on the prediction of the number of COVID-19 cases," in *Proc. 5th Int. Conf. I-SMAC*, 2021, pp. 286–289. DOI: https://doi.org/10.1109/I-SMAC52330.2021.9641006

[4] K. Laxmikant, G. Anitha, and S. M. Usman, "Predicting the death rate of COVID-19 using big data analytics and machine learning," in *Proc. 14th Int. Conf. Computing, Communication and Networking Technologies (ICCCNT)*, IIT Delhi, India, 2023. DOI: https://doi.org/10.1109/ICCCNT56998.2023.10306690

[5] "Mining the impacts of COVID-19 pandemic on the labour market," in *Proc. 16th Int. Conf. Ubiquitous Information Management and Communication (IMCOM)*, 2022.
DOI: https://doi.org/10.1109/IMCOM53663.2022.9721772

[6] K. Tang, Y. Ma, H. Yan, and Y. Rao, "A new prediction method of COVID-19 epidemic," in *Proc. 4th Int. Conf. Pattern Recognition and Artificial Intelligence (PRAI)*, 2021, pp. 331–335.
DOI: https://doi.org/10.1109/PRAI53619.2021.9551052

[7] GeeksforGeeks, "Introduction to Machine Learning," Available: https://www.geeksforgeeks.org/machine-learning/introduction-machine-learning/. Accessed: Apr. 04, 2026.

[8] Our World in Data. (2021). COVID-19 Vaccination Data. Retrieved from https://ourworldindata.org/covid-vaccinations

[9] Gpreda. (2021). COVID-19 World Vaccination Progress [Dataset]. Kaggle. https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress

[10] World Health Organization, "WHO Coronavirus (COVID-19) Dashboard." Available: https://covid19.who.int [Accessed: Apr. 3, 2026].