doi:10.20944/preprints202206.0115.v1

Article

Explainable Machine Learning Approaches to Assess the COVID-19 Vaccination Uptake: Social, Political, and Economic Aspects

Jishan Ahmed[a], Robert Green[a], Md. Alauddin[c], Md. Hasnat Jaman[b] and Goutam Saha[d,*]

[a] Department of Computer Science, Bowling Green State University, Bowling Green-43402, USA
[b] Department of Geology and Mining, University of Barishal, Barishal-8200, Bangladesh
[c] Department of Economics, Jahangirnagar University, Dhaka-1000, Bangladesh
[d] Department of Mathematics, University of Dhaka, Dhaka-1000, Bangladesh
*Corresponding Author: gsahamath@du.ac.bd

Abstract: COVID-19 vaccine hesitancy is considered responsible for the lower rate of acceptance of vaccines in many parts of the world. However, sources of this hesitancy are rooted in many social, political, and economic factors. This paper strives to find the most important variables in predicting the COVID-19 vaccination uptake. We introduce an explainable machine learning (ML) framework to understand the COVID-19 vaccination uptake around the world. To predict vaccination uptake, we have trained a random forest (RF) regression model using a number of sociodemographic and socioeconomic data. The traditional decision tree (DT) regression model is also implemented as the baseline model. We found that the RF model performed better than the DT model since RF is more robust to handle nonlinearity and multi-collinearity. Also, we have presented feature importance based on impurity measure, permutation, and Shapley values to provide the most significant unbiased features. It is found that electrification coverage and Gross Domestic Product are the strongest predictors for higher vaccination uptake, whereas the Fragile state index (FI) contributed to lower vaccination uptake. These findings suggest addressing issues that are found responsible for lower vaccination uptake to combat any future public health crisis.

Keywords: Explainable machine learning; COVID-19; Vaccination uptake; Shapley values; Feature importance

1. Introduction

SARS-CoV-2 virus causes a contiguous disease called Coronavirus disease (COVID-19). From December 12, 2019, in Wuhan, China, to May 1, 2022, the disease has infected 517,860,190 people worldwide and killed 6,278,347 people [1, 2]. Ahmed at el. [3] have taken a statistical approach based on data of 70 cities/provinces to find the role of environmental and socio-economic factors on the spreading of COVID-19. This research found that the negative binomial model is more suitable than the Poisson model for their analysis as the negative binomial provides the best fit to the data. According to this article, the biggest reason for the rapid spread of COVID-19 is population density where temperature and humidity do not play any significant role. The research also found that GDP and $PM_{2.5}$ played a role in reducing proliferation, while PM_{10} and test numbers played a significant role in the spread of COVID-19. Personal hygiene is a must to prevent COVID-19 infection which is very difficult in practice. That is why vaccination is now the most effective way to reduce infections and complications. Although Pfizer launched its first human trial in October 2020, this vaccine was approved for general people from February 21, 2021. So far, WHO has approved six vaccines: Pfizer, Moderna, Oxford, Synovac, Johnson, and Sputnik. As of March 09, 2022, more than 2 billion people worldwide have been vaccinated. Although some developed countries have started giving booster doses, 3 billion people have not yet received the first dose of the vaccine. There are several reasons for

 doi:10.20944/preprints202206.0115.v1

this, one of which is vaccine hesitancy which suggests a lower vaccine acceptance rate among people.

Vaccine acceptance shows classified appearance among age range: vaccination rate is lower among young adults ($18yr–24yr$) than adults ($> 40yr$) [4-10]. However, not only does age range impact the vaccine acceptance rate, but also different gender group show different behavior towards vaccine acceptability, namely that female and non-binary genders have a lower probability of being vaccinated [4, 10-17]. Some social and religious views found a negative impact on vaccine acceptance. Lack of being touched with modern technology and campaign to realize personal and social benefits of being vaccinated [18-20]. Cultural barriers, ethnic tradition, and beliefs were found to have adverse effects on vaccination intention [4, 5, 21-24]. Believing that COVID-19 is harmless, some individuals feel hesitant of being vaccinated [25]. Fears about side effects, infections, and being ill, or allergic to vaccination have a negative impact on vaccine acceptance [4, 5, 11, 13, 17, 22, 26-33]. Inadequate reliability and trust in manufacturers pave the belief that vaccination can be risky [23, 25, 26, 30, 33-35]. As a part of human behavior, the copying tendency sometimes goes against vaccine acceptance [36]. The willingness of being vaccinated fluctuates in response to seasonal variation, data indicate that willingness decreases in summer weather increases in winter [37]. Disabled people are less hesitant than the general population [38]. Individuals having lower socioeconomic status bear higher vulnerability to vaccination [39]. Lower-income and economic insecurity promote hesitancy [10-12, 21, 26, 40, 41].

Homelessness has a negative correlation with vaccine acceptance [7]. In terms of vaccination ratio, larger houses increase resistance [42]. Newness, improper medical trials, and the price of the vaccine create hesitation among the participants and guardians [12, 21, 23, 29-31, 43, 44]. Several studies found lack of proper knowledge about COVID-19 and medical emergencies affects vaccine acceptability. Higher literacy tends to maintain a positive trend in vaccination status [4, 9, 10, 12, 41]. Participants hesitate to become vaccinated if they do not observe someone hospitalized due to COVID-19 [13, 37, 45]. Doctor-patient communication gaps puts the participants in a situation where they are almost unaware of the susceptibility and severity of vaccination [27, 46]. Place of vaccination has also an effect where hesitant participants preferred to be vaccinated in a doctor's office [46]. Moreover, improper maintenance and administration of vaccination activity pose an antithetic impact on people's minds. Dissatisfaction with the government or policymakers leads the participants either to oppose or to hesitate in vaccination [13]. Conspiracy theories against vaccine efficacy drive the participants against vaccination [41]. Some studies postulate that people having a previous vaccination refusal history possess less intention of being vaccinated [23, 28, 47]. Inaccessibility to vaccination centers also lowers the vaccination rate [30, 31].

2. Literature on AI Technologies and Machine Learning:

With the recent advancement in AI technologies, scientists' use of machine learning (ML) models to understand the health equity and disparities within public and population health has been on the rise. Researchers used XGBoost (extreme gradient boosting) and tree-based models in the past to predict influenza vaccination uptake [48, 49] and childhood immunizations [50]. Grandhi et al. [51] used a logistic regression model to explore the sociodemographic disparities in influenza vaccination. Therefore, we see the use of such ML models to understand the dynamics of the COVID-19 vaccination campaigns worldwide. Hayawi et al. [52] utilized XGBoost, LSTM, and BERT transformer models to understand vaccine hesitancy by analyzing the Twitter dataset. They have used misleading accuracy metrics to evaluate their model in the presence of class imbalance. Mewhirter et al. [53] implemented a gradient boosting (GB) model to analyze online survey data. In the survey response, 56.1% accepted the vaccine and 43.9% were vaccine-hesitant which is a clear indication of class imbalance. They have also compared the performance GB model with logistic regression, decision trees, bagging, and random forests models. They have also reported accuracy metrics to compare models. They claimed that the GB approach outperformed the other four models although logistic regression did pretty well to classify vaccine acceptant and vaccine-hesitant respondents. They mentioned that they have used the built-in variable importance of the XGBoost R package. They also cited the vip R package which is confusing since vip gives permutation and Shapley values-based variable importance measure. Built-in variable importance is unreliable in the presence of correlated features (We will discuss this issue in more detail in section 3.2). They found that the vaccine trust index is the most important predictor to understand vaccine hesitancy. Hafizh et al. [54] used different machine learning techniques to explain the COVID-19 vaccine willingness and hesitancy among residents in Qatar. They performed dominance analysis to assess the relative importance of variables for ML models. They found that the chance of avoiding quarantine and lockdown restrictions motivated mostly to take the vaccine. Dominance analysis was computed by using the R^2 measure. However, ML models used in this study showed very low R^2 scores (< 0.50) although $R^2 > 0.50$ is expected.

Mondal et al. [55] employed logistic regression and neural network models to identify the most significant sociodemographic predictors on COVID-19 vaccine acceptance. Education, ethnicity, and age were reported as significant variables. The strength of their study was the use of the Chi-squared test to validate the ML models. Jayasurya et al. and others [56-58] conducted sentiment analysis to understand the public sentiments and opinions toward the COVID-19 vaccines by using social media data. Cheong et al. [59] utilized the XGBoost regression model to explore the association between sociodemographic factors and vaccination rate across US counties. They have used XGBoost's built-in feature importance along with permutation and SHAP feature importances to find the important predictors. Location, education, ethnicity, income, and household's internet access were reported as the most important sociodemographic variables in their study in predicting vaccination uptake. However, it is not clear how the authors have evaluated the quality of model fit. Therefore, it is impractical to expect unbiased feature importance from a poor fit model. We usually use accuracy to evaluate classification models. Surprisingly, the authors have mentioned accuracy metric throughout the paper to explain their regression model fit. In our work, we will give a more comprehensive ML framework to understand the factors associated with vaccination uptake. Carrieri et al. [60] introduced a very interesting approach to explaining vaccine hesitancy by using ML models. They have used child immunization campaign data which were conducted in Italian municipalities in 2016 for vaccine-preventable diseases (pertussis, measles, Haemophilus influenza type B, Meningococcus, pneumococcus, mumps, and rubella). They found that the level of waste recycling and employment rate were the most significant area-level predictors of communities with higher vaccine hesitancy.

The above mentioned literature has indicated that disparities in vaccination rates are influenced by sociodemographic and socioeconomic factors. Likewise, few studies have explored the influence of sociodemographic and socioeconomic factors on vaccination uptake on a national level. We utilize the sociodemographic, socioeconomic, and child immunization surveillance data to gain insight into the global COVID-19 vaccination campaign. We have implemented an explainable ML framework to provide a more robust and comprehensive approach for a better understanding of the disparities in vaccination uptake worldwide as well as address the limitations of the previous studies. We have also attempted to find the most influential factors which would allow policymakers to design COVID-19 vaccination campaigns as well as future immunization campaigns. In this study, we employed an RF model along with a baseline DT model to predict the vaccination rate. We have also performed the hyperparameter optimization tool to find the best RF model. We have incorporated the SHAP framework in our work to provide global and local explanations of the predictions. More importantly, we used permutation feature importance and SHAP global feature importance as reliable feature importance techniques.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2022 doi:10.20944/preprints202206.0115.v1

The objective of this study is to offer policy implications to prevent future global pandemics and strengthen systems for public health preparedness worldwide by building an explainable ML model.

Data description

Sociodemographic, socioeconomic, and immunization data was collected from the World Bank [61], Worldometer [1], Human Development Data Center [62], and Fragile States Index site [63]. Vaccine safety data was collected from the Wellcome Global Monitor 2020: COVID-19 site [64]. Most recent years were taken into account while collecting data from these data sources. In addition, the percentage of fully vaccinated people against COVID-19 was found from Johns Hopkins University & Medicine [65] and The New York Times [66] databases. The percentages are representative of COVID-19 vaccination data from March 09, 2022.

Our study included 182 countries and the rest of the countries were excluded due to the lack of available data as we see in Fig. 1. Prevalence Rates Asia Data from 46 countries in Asia shows mostly middle to high COVID-19 vaccination rates. The highest and lowest vaccination rates are 97.48% in the United Arab Emirates and 1.3% in Yemen. The vaccination rate is 88.32% in China ensuring a third place in Asia. Overall, 29 countries show a greater than 50%, and 17 countries show a lesser than 50% vaccination rate in Asia. Africa Among the 51 countries of the African continent, only 7 countries show a vaccination rate greater than 50%. These countries are Seychelles, Mauritius, Morocco, Rwanda, Tunisia, Cabo Verde, and Botswana.

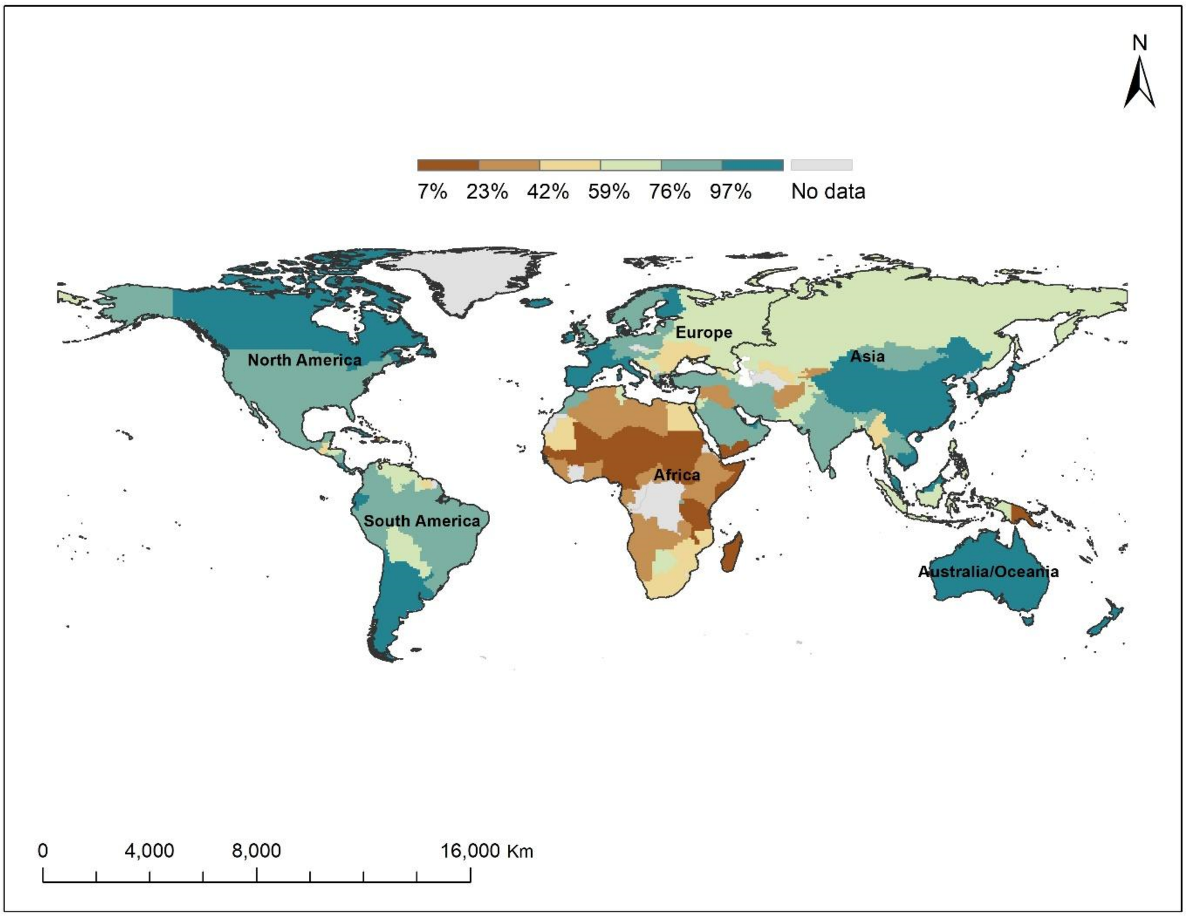


Figure 1. COVID-19 vaccination campaign worldwide.

The highest vaccination was found to be 81.86% in Seychelles whereas only 0.07 percent in Burundi. The COVID-19 vaccination rate is very non-homogeneous in Africa. Australia/Oceania In Australia/Oceania continent, we observed Australia has 80.91%, New

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2022 doi:10.20944/preprints202206.0115.v1

Zeeland has 80.72%, Fiji has 69.66% COVID- 19 vaccination rate but Papua New Guinea has only 2.77% vaccination rate. From the 46 countries of Europe, the data indicates that 34 countries have a higher than 50% vaccination rate and 12 countries have a lower than 50%. The highest rate of vaccination is 92.28% in Malta and the lowest rate is 25.63% in Bosnia and Herzegovina.

We have found vaccination rate data from 23 countries of the North American continent. The majority of countries in North America have intermediate to high rates of vaccination. Cuba has the highest vaccination rate of 87.34% percent while Haiti has the lowest vaccination rate of 0.93%. In North America, 14 countries have a more than 50% vaccination rate and on the other hand, 9 countries exhibit vaccination rates below 50%. From the data of 12 countries of the South American continent, Chile shows the highest COVID-19 vaccination rate which is 91.05%. The other 11 countries are Argentina, Uruguay, Ecuador, Peru, Brazil, Colombia, Venezuela, Bolivia, Paraguay, Guyana, and Suriname shows the vaccination rate of 80.56%, 78.9%, 78.08%, 75.18%, 73.79%, 66.29%, 50.1%, 48.67%, 45.67%, 42.16%, and 40.61% percent respectively.

Variables considered:

Income (IL): Countries, while classified in terms of their income, are found to have significant regional, cultural, and economic differences. These differences may significantly influence the decision of vaccine uptake. Compared with a higher income country (USA) and an upper-middle-income country (Russia), [67] find that people in lower and middle-income countries are more willing to take vaccines. The magnitude and causes of vaccine hesitancy vary significantly across regions with income differences.

Vaccine Safety (VS) and vaccine effectiveness (VE): Concerns about safety reasons for hesitancy [67]. A survey over 15 African countries conducted by the Africa Centres for Disease Control and Prevention, in partnership with the London School of Hygiene and Tropical Medicine, found that 79% of people in Africa would take the vaccine if they deem it safe and effective [68]. The rapid pace of vaccine development is one of the reasons why some people in developed countries think that vaccines are unsafe to take into their bodies [69].

Percentage of people aged 15-64 years (PS): The group between 15-64 years old represents the economically active (Working-age) population. Being healthier and more active, people of this age might be reluctant in getting vaccinated as they might think they are less prone to become ill. Therefore, while considering the age structures of the population, younger age groups are generally observed to be least likely to take COVID-19 vaccines [10, 70-73].

Percentage of people age 60 or over (PO): Several cross-country studies found older age groups are more confident in the safety and importance of vaccines [73, 74]. Therefore, people aged above 60 are expected to be less hesitant in vaccination uptake.

Percentage of the rural population (RP): Vaccine hesitancy in rural areas is a major barrier that should be addressed with a strong emphasis. [75, 76]. People living in rural areas are often lagged behind in terms of education and awareness, which may appear to hinder taking the vaccine. Moreover, rural people have limited access to health care facilities. Rural areas are segregated, and vaccines might not be easily accessible in these areas because of limited coverage. If people living in rural areas are required to travel long distances to take vaccines, they might become hesitant to get vaccinated. Also, urban-biased approaches that focus mainly on raising public awareness in urban areas fail to produce effective results in rural areas. Therefore, such a policy measure will be less successful in motivating people to take vaccine upshot if the population consists of many rural people.

Percentage of Poverty (P): Poverty could be another source of vaccine hesitancy [76]. Poverty presents several challenges and restricts one from availing many life-changing opportunities. A country with a vast majority of poor people has to deal with severe socio-economic problems, including the challenge of motivating people towards any proposed positive change.

 doi:10.20944/preprints202206.0115.v1

Percentage of Immunization (DPT(ID), HepB3(IH), and Measles(IM)): In countries where childhood vaccination for common diseases has been widely accepted, the COVID-19 vaccine can potentially receive vast acceptance, too [67]. Historically, people in some countries exhibited higher acceptance of vaccines. Because people in these countries believe that vaccines can effectively prevent diseases, they are much more likely to accept any new vaccine if it is offered. Therefore, positive attitudes towards vaccination and the consequent success of vaccination programs in the past in some countries indicate that any newly proposed vaccination can be highly accepted in these countries. People in these countries, therefore, are likely to be less hesitant in taking vaccines.

Percentage of people using the Internet (PI): The Internet, facilitating the use of Social media and online conversation, could be another potential source of vaccine hesitancy [77]. Nowadays, information has become much more available through the use of the internet and smartphone. Therefore, it can be seen as a greater source of self-education that can play a role in influencing people's decision to take the vaccine. On the other hand, however, greater internet access can help spread misinformation, too [78].

Percentage of people using electricity (PE): The higher the coverage of electricity in a country, the higher we can expect access to information. This is based on the ground that the availability of electricity enables people to take the privilege of several types of modern technology, such as television, radio, internet, and mobile, all of which can raise public awareness about the benefits of vaccination.

GDP-Per Capita (USD) (GD): GDP per capita measures the economic output of a nation per person, reflecting the economic prosperity of a nation. Countries with higher economic prosperity can allocate more resources to formulating effective vaccine campaigns to combat critical situations, such as the COVID-19 pandemic. In addition, rich countries can spend more on public health and defend any misinformation regarding the vaccine, which may build public confidence in the COVID-19 vaccine.

Education Index (2019) (EI): Educated people are more informed about the benefits of the vaccine. Hence, education can help people to avoid misinformation regarding vaccination. Therefore, the overall education level in a society can be a significant contributing factor toward the decision in taking a vaccine.

Human Development Index (HI): Human Development Index (HI) is a very crucial determinant to evaluate countries' public health infrastructures. Countries with better public health infrastructures are resilient to tackle epidemic or pandemic. Few researchers [79-83]. included the HI variable to assess the vaccination coverage. UNDP also reported that the HI variable is sensitive to COVID-19 impact [84]. Therefore, we included the HI variable in our study to assess its contributions to vaccination uptake.

Fragile state Index 2020 (FI): FI aggregates state failure risk elements such as the prevalence of extensive corruption and criminal behavior, challenges in collecting taxes, voluntary dislocation of mass population, sharp economic decline, group-based inequality, institutionalized discrimination, severe demographic pressures, brain drain, and environmental decay. State fragility can be regarded as another source of vaccine hesitancy. A state's fragility is manifested in the loss of control over physical territories and its failure in making a collective decision and interacting with other nations, inability to provide public services. All of which resulted in an unstable society that is divided and lacks cohesion. Consequently, any government objectives requiring public support are hard to fulfill as any message conveyed to the citizens is mistrusted.

Region: Global vaccine inequity contributed towards disparities in COVID-19 vaccination coverage across different continents [80, 85-87]. Especially low-income and middle-income countries (LMICs) have access to a disproportionate amount of COVID-19 vaccine doses due to vaccine nationalism and vaccine diplomacy [88, 89]. Fair allocation of vaccines among nations is a very important pathway to end the pandemic [90]. As people are now exhausted with this prolonged pandemic, global coordination and solidarity is necessary during this difficult time to achieve the global vaccination target as proposed

 doi:10.20944/preprints202206.0115.v1

by WHO [85]. Region variable is included in our study to understand the impact of vaccine inequity across different continents.

3. . Methodology

3.1. Random forest regression

Random forests (RF) is an ensemble machine learning technique, proposed by Breiman, that combines bagging with random feature selection to decorrelate independent decision trees. RF is considered as a black-box model due to its less interpretability compared to individual decision trees [91].

i. Algorithm:

For b = 1 to B:

Generate a bootstrap sample B_* from the training data. Fit decision trees T_b on a bootstrapped training sample using the following splitting and stopping criteria:

a) Randomly pick q features from p features and allow splits on these q features only.
b) Select the best features among q features
c) Split the node into two leaf nodes.

ii. Return the ensemble of trees $\{T_b\}_1^B$

Now the random forest regression predictor can be written as

$$\hat{f}_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{1}$$

In RF, one-third of the training examples are left out of the bootstrap sample during training. These left-out examples are referred to as out-of-bag (OBB) samples. The utilization of out-of-bag (OBB) samples to achieve an unbiased estimate of the test set error is the most important feature of the RF. More importantly, an OBB error estimate is equivalent to k-fold cross-validation.

3.2. Feature Importance

In RF, the mean decrease in impurity (MDI) is a widely used impurity-based feature importance for regression forests. The mean squared error (MSE) based splitting criterion of the decision trees quantifies the impurity. The splitting criteria of RF optimize the maximum decrease in impurity for each split. In impurity-based feature importance, a split with a maximum decrease in impurity is regarded as an important split and the variable associated with this split is an important variable. Consequently, Impurity decrease measures are averaged across all decision trees to compute the importance of each predictor variable. However, impurity-based importances are biased due to the bias of selecting split variables [92]. Also, they favor the features with high cardinality [92]. More importantly, training set-based impurity importance doesn't reflect the actual performance of the feature on the test set. To avoid such bias, Breiman [93] proposed Permutation-based feature importance. Permutation-based feature importance of RF has been widely accepted by machine learning researchers due to its reliability [92, 94, 95]. Permutation importance for variables, X^p is computed as follows [94]:

i. Compute the MSE for the OBB/test samples

ii. Randomly shuffle the columns the X^p of OBB/test samples to generate the permuted data. Then compute the MSE, $\widehat{MSE}$ for this permuted data.

iii. Feature importance is then computed as

$$VI(X^p) = \frac{1}{n-estimators}\sum_{t}(MSE - \widehat{MSE}) \tag{2}$$

We can consider the permutation importance as a model-agnostic version of the feature importance since it can be used to explain predictions of any model class. We can get the global insight of the model's predictions through permutation importance, but, be that as it may, the permutation feature importance can be misleading due to the presence of correlated features. In such a situation, SHAP feature importance is another good model-agnostic alternative to permutation feature importance.

3.3. Model explainability

Since RF is considered as a black-box model, model-agnostic feature importance ensures the interpretability of the model. Feature importance is computed based on Shapley values by considering the influence of each feature on the model. The basis for Shapley values is a cooperative game theory [96] where the coalition represents the set of explainable input variables and the outcome of the coalition is the model's predictions for these given input variables. In other words, Shapley values are the mean marginal utility of input variables out of all possible combinations of coalitions. The influence of the feature is computed from the shift in the model's expected predictions for an observed vs. unseen feature. Likewise, the Shapley values $\beta_j(rf, z)$ are the distribution of utility among the input variables in z to explain the RF model's predictions. Shapley values also maintain the two significant properties local accuracy (additivity), and consistency (monotonicity) of additive feature attribution methods [97] while assigning feature importance.

3.3.1. Local accuracy (additivity)

$$\widehat{rf} = \beta_0(rf, z) + \sum_{j}^{p} \beta_j(rf, z) \tag{3}$$

where $\beta_0(rf, z) = E_Z[\widehat{rf}(Z)]$ is the expected value of the predictions for the training data, and p represents the total explainable variables associated with the raw explanatory variables.

3.3.2. Consistency (monotonicity)

For any two models rf^1 and rf^2, if

$$rf_z^2(T) - rf_z^2(T \setminus j) \geq rf_z^1(T) - rf_z^1(T \setminus j), \forall T \in Z \tag{4}$$

then

$$\beta_j rf_z^2(T) \geq \beta_j rf_z^1(T) \tag{5}$$

implies that if there is increase or consistency in the marginal influence of an explanatory variable for a shift in models in the presence of other variables, the Shapley value will increase or remain the same. The Shapley values provided following utility allocation satisfy the above two properties simultaneously:

For a given prediction rf, the Shapley value of the variable j is given by

$$\beta_j(rf, z) = \sum_{T \subset T_{full} \setminus \{j\}} \frac{|T|!\,(p - |T| - 1)!}{p!} \times [rf_z^1(T \cup \{j\}) - rf_z^1(T)] \tag{6}$$

This represents the contribution of each feature which is being accumulated in the model by taking the average of all possible sequences (coalitions) of variables. However, the estimation of β_j is NP-hard [98] because it requires taking summation over so many terms. Lundberg et al. [99] introduced a sampling technique to estimate β_j. This method is still computationally expensive [99] due to the exponential complexity of Equation 6. To overcome this challenge, Lundberg et al. [101] proposed TreeExplainer to explain the model-specific tree-based machine learning models. It accommodates the local explanations aligned with local accuracy and consistency properties. It lowers the computational

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2022 doi:10.20944/preprints202206.0115.v1

time complexity from exponential to polynomial time as well by utilizing the intrinsic architecture of tree-based models. TreeExplainer also offers SHAP interaction values as local explanations by using the Shapley interaction index [102]. Shapley interaction index assigns payoff among all pairs of features resulting in a feature attributions matrix where off-diagonal entries represent interaction effects and main effects are placed in the main diagonal. The SHAP interaction values can be written as:

$$\beta_{k,j} = \sum_{T \subset T_{full} \setminus \{j\}} \frac{|T|!\,(p-|T|-2)!}{2(p-1)!} \times \delta_{k,j}(T) \; whenever \; k \neq j \tag{7}$$

Here,

$$\delta_{k,j}(T) = \hat{r}f_z^1(T \cup \{k,j\}) - \hat{r}f_z^1(T \cup \{k\}) - \hat{r}f_z^1(T \cup \{j\}) + \hat{r}f_z^1(T) \tag{8}$$

In Equation 7, the interaction effects are computed by deducting the main effect while individual effects are taken into account. Similar to Shapley value estimation, we take the average over all possible sequences (coalitions) of variables. In TreeExplainer, SHAP summary plots are generated to show us the magnitude and direction of the feature's influence. More importantly, SHAP interaction values allow us to observe the interaction effects locally through SHAP dependence plots. SHAP dependence plots evaluate the influence of each feature on the predictions for every testing observation. We use SHAP summary plots, and SHAP dependence plots in our work by leveraging Shapley values. In this paper, we have also utilized the SHAP force plot to understand the influence of each feature on the predictions by considering Shapley values as 'forces' [99].

3.4. Assessing the quality of fit

To evaluate the performance of the ML models in a regression setting, root mean squared error (RMSE) is the most widely used measure. RMSE is computed as

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{rf}(x_i)^2)} \tag{9}$$

where $\widehat{rf}(x_i)$ represents the $\widehat{rf}$ model's predictions for the i^{th} test observation. RMSE tells us how far the model residuals are from zero on average, i.e. the average distance between the observed values and the predicate values. However, Willmott et al. [103] suggested that RMSE might be misleading to assess the model performance since RMSE is a function of the average error and the distribution of squared errors. Chai et al. [104] recommended using both RMSE and mean absolute error (MAE). In our work, we have reported both metrics. MAE is defined as

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{rf}(x_i)\right) \tag{10}$$

R^2 is another popular measure to evaluate a regression model. R^2 gives us the percent of variance explained by the regression model, for instance, $R^2 = 0.75$ implies that the model can explain 75% of the variation in the outcome. We can computer R^2 as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \widehat{rf}(x_i)\right)}{\sum_{i=1}^{n}\left(y_i - \underline{rf}(x_i)\right)} \tag{11}$$

where $\underline{rf}$ is the mean of the target variable. In this paper, we have reported the results in predicting vaccination uptake using RMSE, MAE, and R^2 evaluation metrics.

4. Experiments

Based on our discussion in section 3, a conceptual framework of the ML pipeline is graphically illustrated in Fig. 2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2022 doi:10.20944/preprints202206.0115.v1

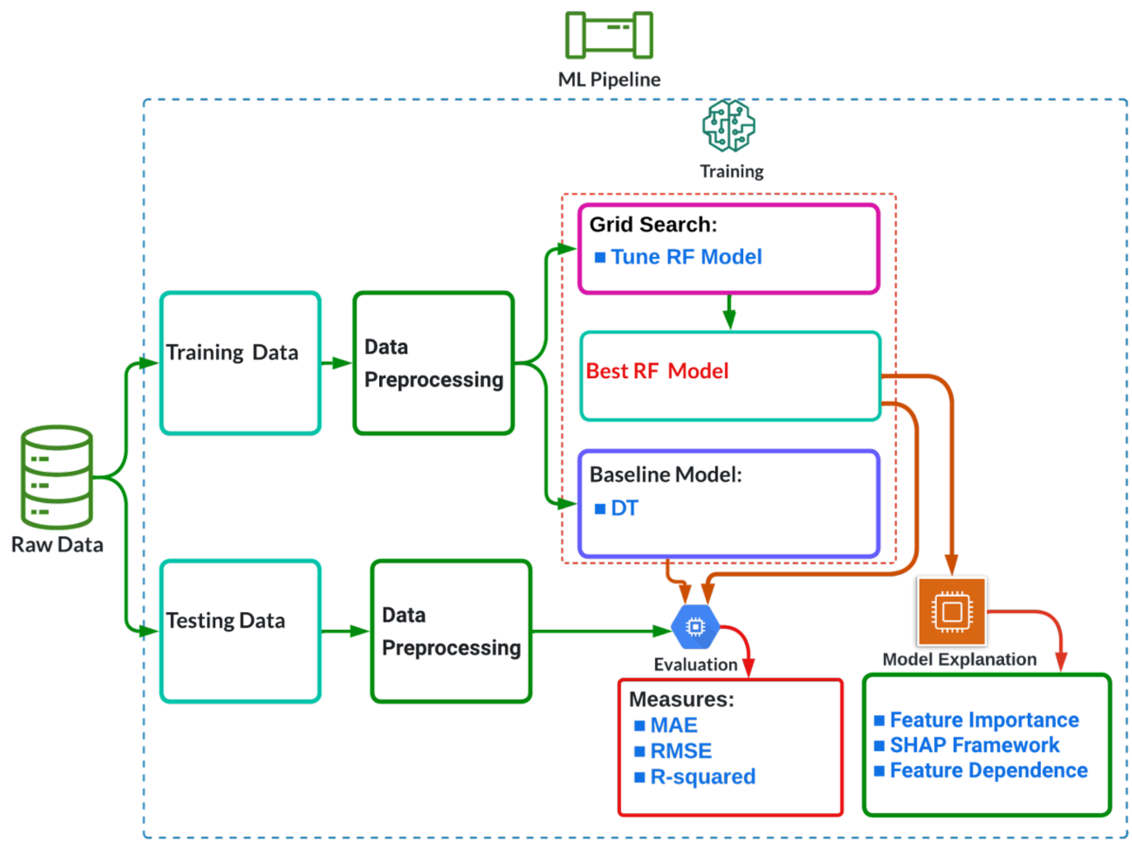


Figure 2. Conceptual architecture of the ML pipeline for vaccination uptake prediction.

Here we provide the following steps of our implementations to facilitate the reproducibility of our results by future researchers and professionals:

- We have performed exploratory data analysis to understand the data distribution and multi-collinearity.
- We have implemented random forests (RF) regression to predict the percentage of fully vaccinated people for a country.
- The predictive performance of the RF regressor is compared with a baseline decision trees (DTs) regressor.
- We have randomly split the dataset into 80% training and 20% testing samples.
- GridSearchCV is performed on the training samples to tune hyperparameters of the RF regressor. GridSearchCV has utilized 5-fold cross-validation to find the best parameters as shown in Fig. 7.
- The performance of regressors were compared in terms of the RMSE, MAE, and R^2 metrics.
- The SHAP framework is established to explain the predictions of the best predictive model by leveraging the Shapley values from coalitional game theory.

4.1. Experimental Setup

All of our implementations are performed using the same multi-core machine. Our data can be accessed from Gitlab to reproduce the results. (Data source Link: https://gitlab.com/Jishan/iml-data) We used a PC with an Intel(R) Core(TM) I7-7700 CPU at 3.60GHz (4 cores) and 16 GB RAM to implement ML models. The operating system for this PC is Windows 10 Enterprise. The main objective of this study is to find the important

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2022 doi:10.20944/preprints202206.0115.v1

factors that influence the global vaccination uptake by utilizing the explainable machine learning tools. We have considered the following tools to build our computational framework in Python:

- RandomForestRegressor provided by the sci-kit learn module in Python
- DecisionTreeRegressor being provided by the sci-kit learn module in Python
- GridSearchCV provided by the sci-kit learn module in Python
- TreeExplainer from the SHAP Python library
- partial_dependence provided by the sci-kit learn module in Python

5. Results and discussion

5.1. Exploratory data analysis

We plot the distribution of the target variable, the percentage of fully vaccinated people in Fig. 3. We see that the distribution of the target variable doesn't follow a normal distribution. However, we don't have to transform our observed non-normal variables since linear regression analysis does not assume normality for either explanatory variables or a target variable [105-109].

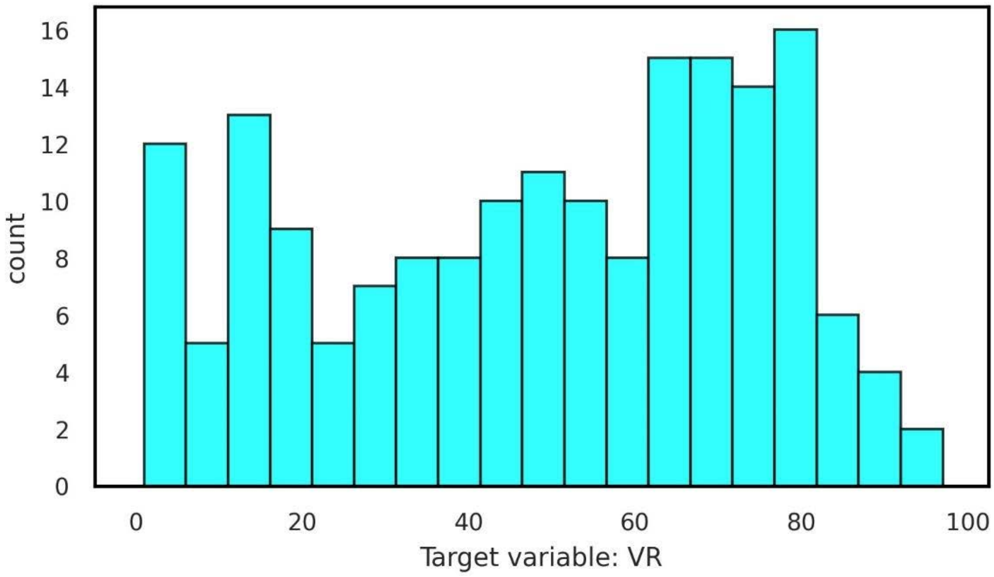


Figure 3. Distribution of the target variable.

Our target and predictor variables can have any distribution since we will use non-parametric tree-based methods in our work. We have shown four countries with high vaccination uptake and four countries with least vaccination uptake Fig. 4. From this heatmap, we see that percentage of the rural population (RP), percentage of internet users (PI), GDP-Per Capita (GD), and fragile state index (FI) may have some influences on the vaccination coverage of a country.

 doi:10.20944/preprints202206.0115.v1

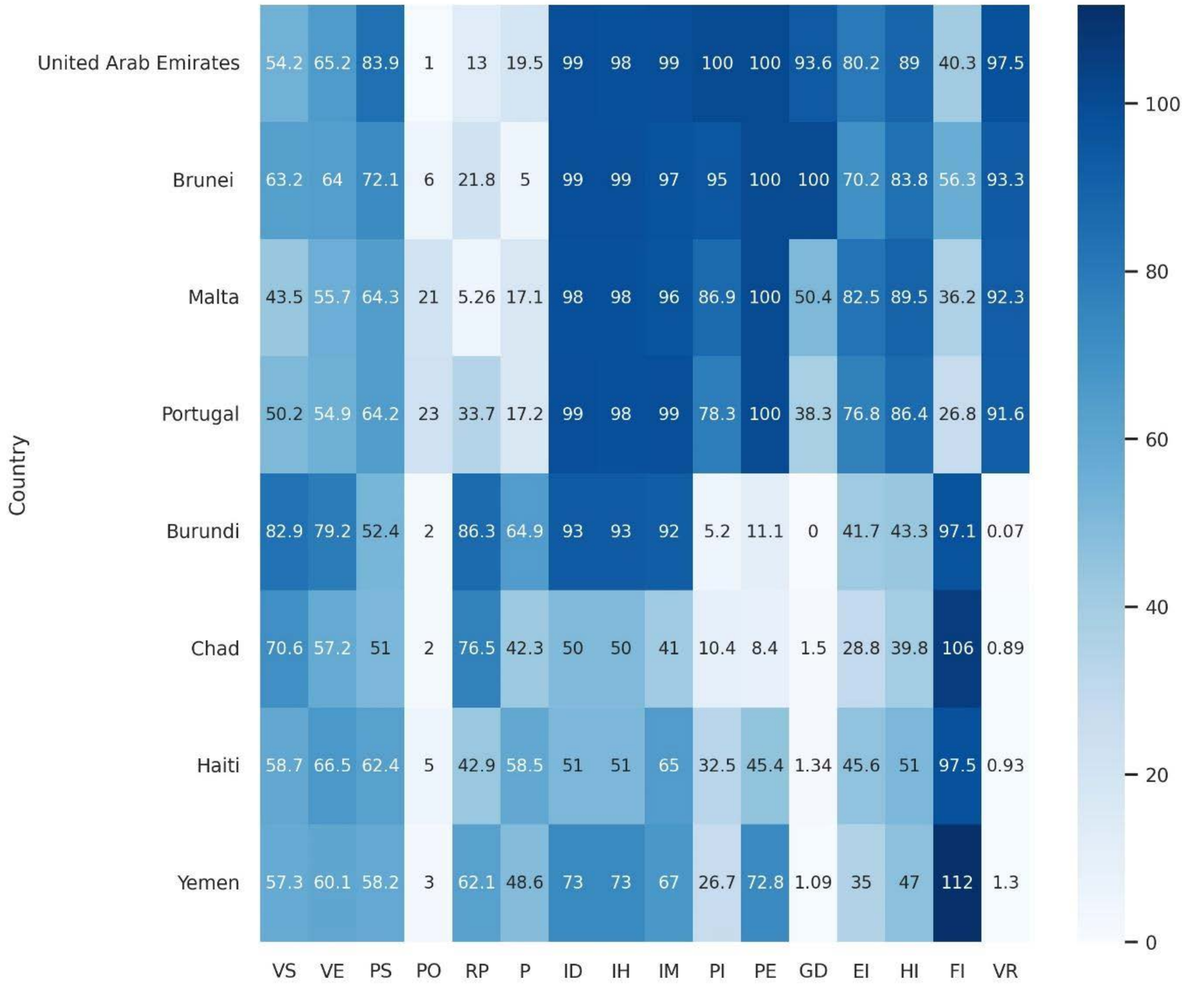


Figure 4. Countries with the highest and least vaccination uptake.

However, we will investigate more to find insight into these variables. We plot a pairwise relationship among these variables in Fig. 5 along with the univariate distribution of each variable in the main diagonal. We clearly see a nonlinear relationship between the pairs of these variables. We also notice that the univariate distribution of each variable differs across continents. We have used nonparametric Spearman's rank-order correlation (Fig. 6) due to the presence of nonlinear relationships in Fig. 5. Spearman's correlation coefficient is an appropriate measure [110, 111] to detect multicollinearity since it is more robust to outliers and it doesn't assume linearity between variables. We see that HepB3 immunization coverage (IH) and human development index (HI) are highly collinear. We might see misleading feature importances from our ML models due to the presence of these collinear features. As a result, we have implemented permutation feature importance and SHAP feature importance in addition to impurity-based feature importances to avoid bias in determining the most influential explanatory variables.

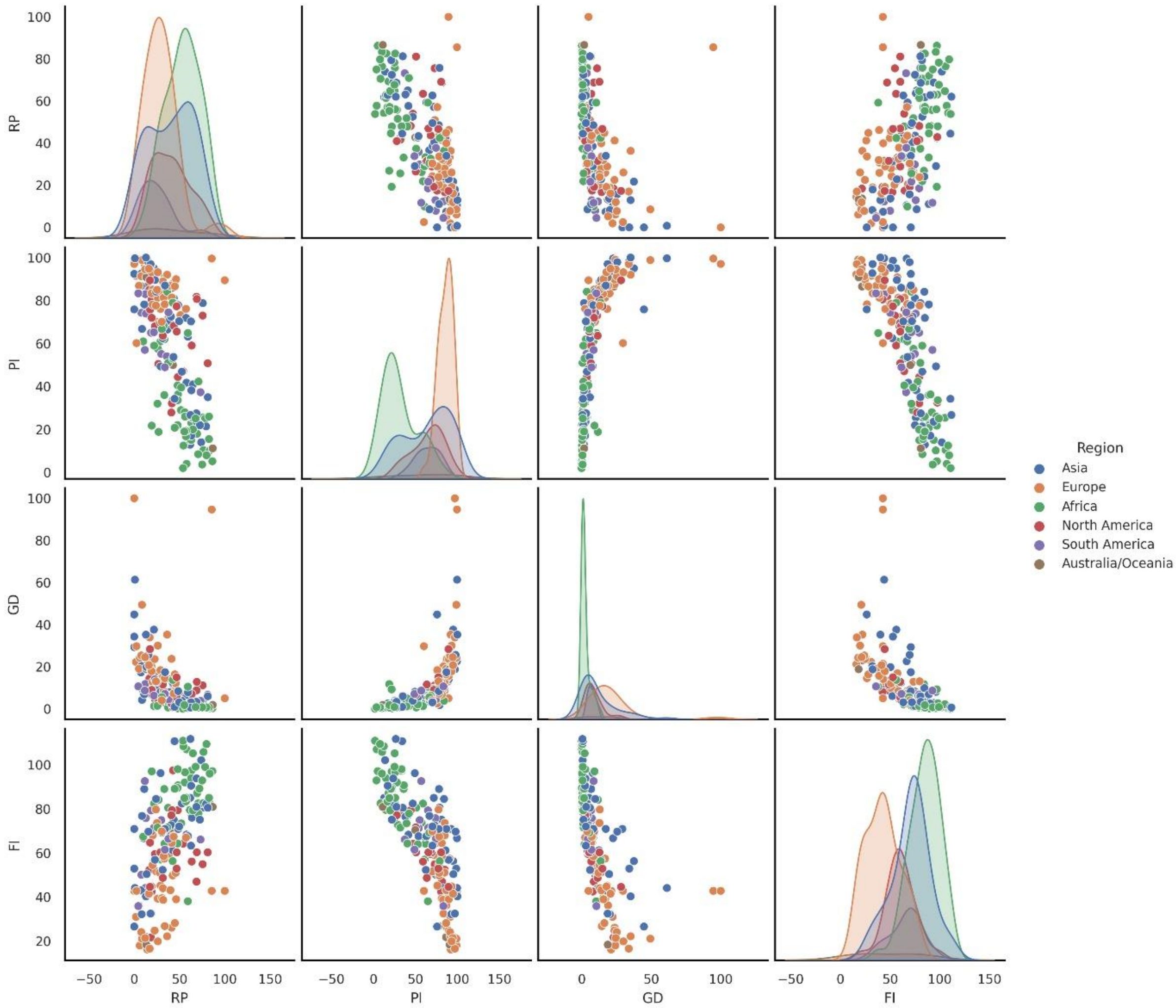


Figure 5. Distribution of predictors across continents.

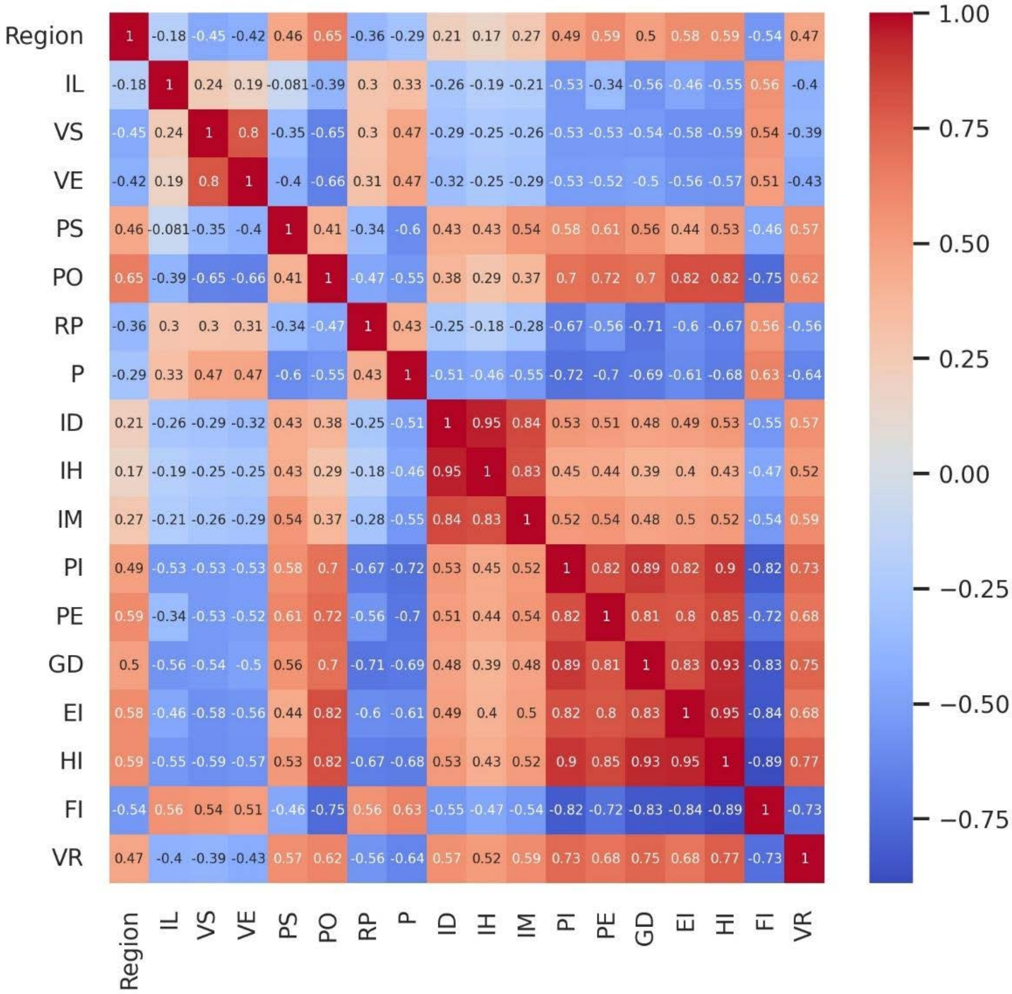


Figure 6: Spearman's rank-order correlation matrix to detect collinear predictors.

5.2. Hyperparameter optimization: GridSearchCV

We have used a grid search provided by GridSearchCV to exhaustively generate 96 candidates from a grid of parameter values as shown in Table 1. We have considered a total of 480 fits resulting from 5 folds for each of 96 candidates to find the optimal hyperparameters of our RF model.

Table 1. Tuning hyperparameter using GridSearchCV.

Hyperparameter	Parameter space	Optimal hyperparameter
maxdepth	{2, 3, 4, 5}	4
criterion	absolute_error, squared_error	absolute_error
n_estimators	{5, 10, 20, 30, 40, 50, 60, 70, 80, 100, 150, 200}	80

From Fig. 7, we observe that RF model with grid search has attained best results when maxdepth is set to 4, n_estimators is set to 80 and absolute_error is chosen as splitting criterion.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2022 doi:10.20944/preprints202206.0115.v1

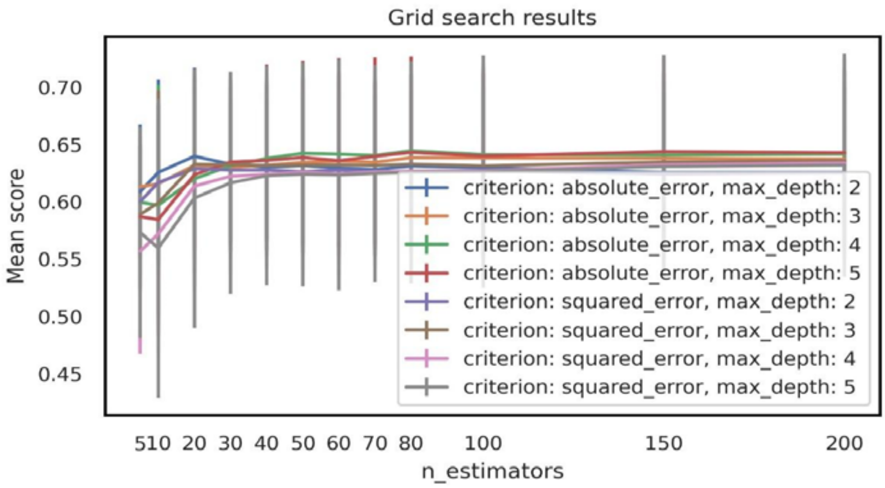


Figure 7. Grid search results using 5-fold CV.

5.3. Model selection and evaluation

We have compared the performance of the RF model with the base DT model using test data. All explanatory tables are included in our model. We have also trained our RF model without HepB3 immunization coverage (IH) and human development index (HI) variables as they are highly collinear. We have achieved the same results even without these variables. However, we did not exclude these two variables from our final RF model since the RF model can handle multicollinearity [112]. We have reported the results in Table 2 using RMSE, MAE, and R^2. We see from Table 2 that the RF model performed well on the test data compared to the DT model. Therefore, we will compute the feature importance using the RF model. Likewise, we will provide the global explanations and local explanations of the predictions of the RF model in section 5.4.

Table 2. Performance of ML models on test data.

Model	MAE	RMSE	R^2
RF	10.88	13.63	0.71
DT	13.86	17.39	0.53

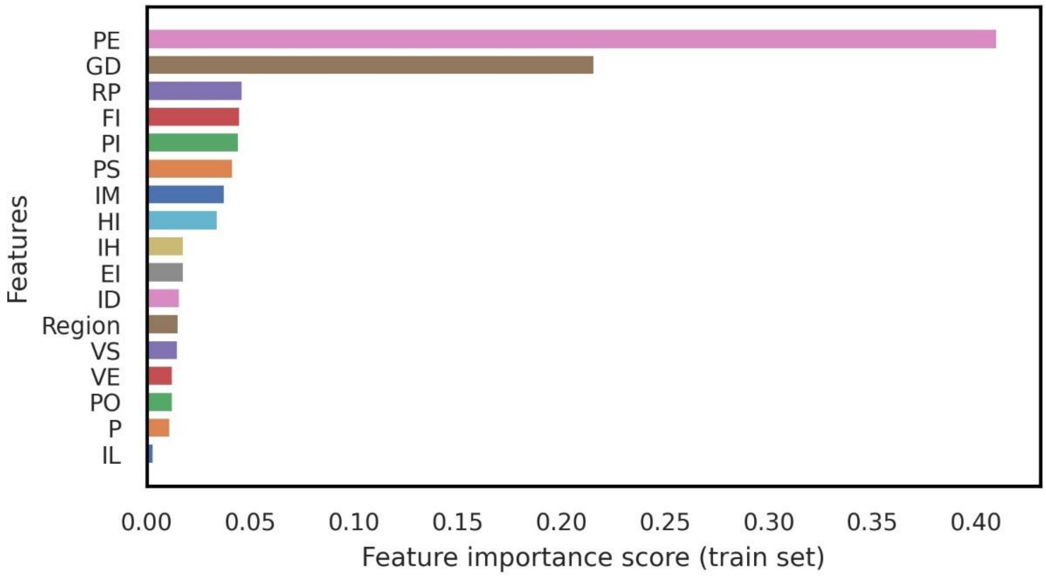
5.4. Explaining predictions

Our aim of this work is to find the most important variables in predicting vaccination uptake using explainable techniques. In this section, we have implemented impurity and permutation-based feature importance to understand the individual feature's contributions on the predictions. Shapley values-based SHAP tree explainer is used as well to provide enhanced global explanations of the feature contributions by incorporating local explanations. We also presented feature dependency analysis and interaction effects to gain more insight into the predictions.

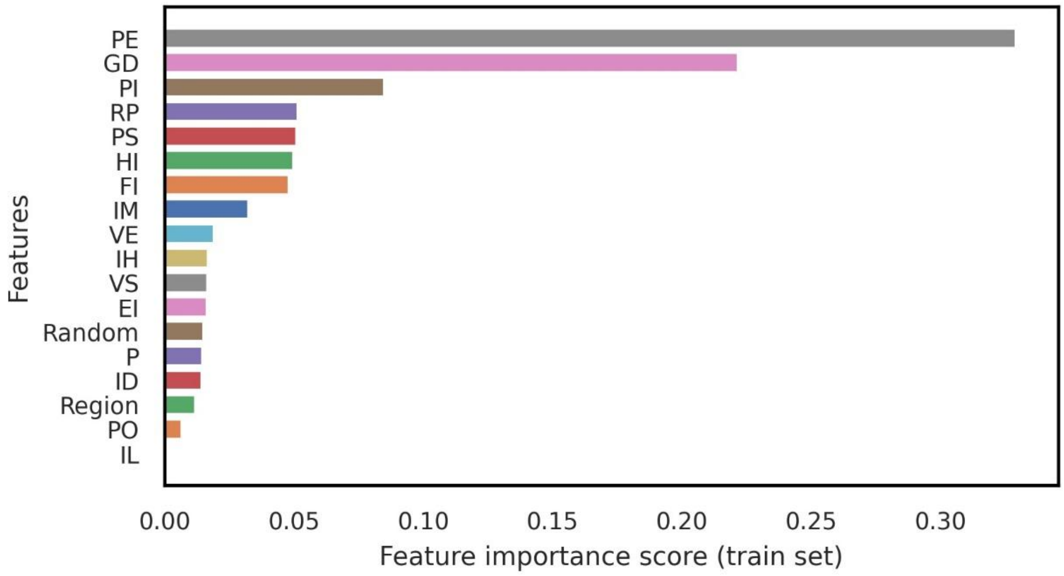
5.4.1. Global feature importance

In this paper, we have attempted to quantify the relevance of the explanatory variables with the response variable. Mean decrease in impurity (MDI) based variable importance values are presented in Fig. 8 to measure the aggregate effect of the variables on the RF model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2022 doi:10.20944/preprints202206.0115.v1



(a) MDI-based variable importance



(b) MDI-based variable importance with non-predictive column

Figure 8. MDI-based variable importance for RF regressor on the training data.

MDI variable importance was computed on the training set. From Fig. 8a, we see that electrification coverage (PE) is the strongest predictor of vaccination uptake. GDP (GD), percentage of the rural population (RP), and fragile state index (FI) are among the top four important predictors. We have also added a feature with a non-predictive column of random numbers to find the nature of the least important features. Fig. 8b shows that the percentage of poverty (p), DPT immunization coverage (ID), Region, percentage of the older population (PO), and Income level (IL) have no influence on the RF model since they are no better than the random feature. We know from section 3.2 that the RF model might give biased feature importance. Therefore, we have computed the feature importance using feature permutation as described in the section. Moreover, permutation importance is a reliable feature importance measure and it also takes into consideration the feature interaction effects. We have computed the permutation importance both on training data and test data.

 doi:10.20944/preprints202206.0115.v1

Fig. 9 displays the permutation importance for each of the 17 predictor variables. Not surprisingly, electrification coverage (PE) is the most relevant predictor. Comparing permutation importance (Fig. 9) with the MDI-based importance in, we see that 2 of the top 4 variables are the same (electrification coverage and GDP). Also, measles immunization coverage (IM) appears to be one of the top four predictors in both training and test set. We notice that the ranking of the features in permutation importance is slightly different from MDI-based feature importance. However, both MDI-based importance and permutation importance don't show how much each feature contributes to a prediction. Therefore, we have implemented SHAP feature importance by aggregating the Shapley values to show the individual feature contributions.

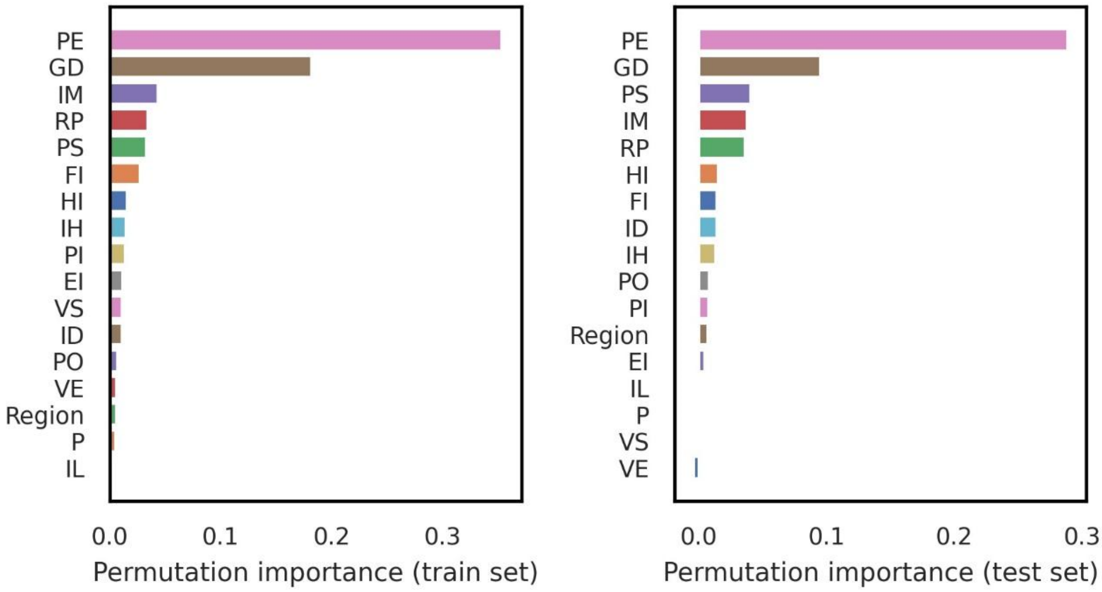
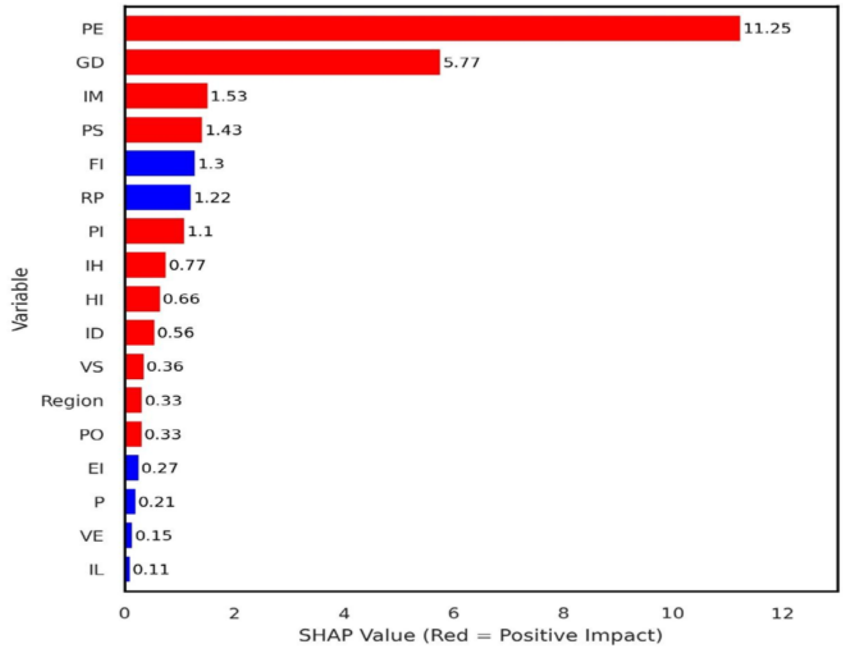


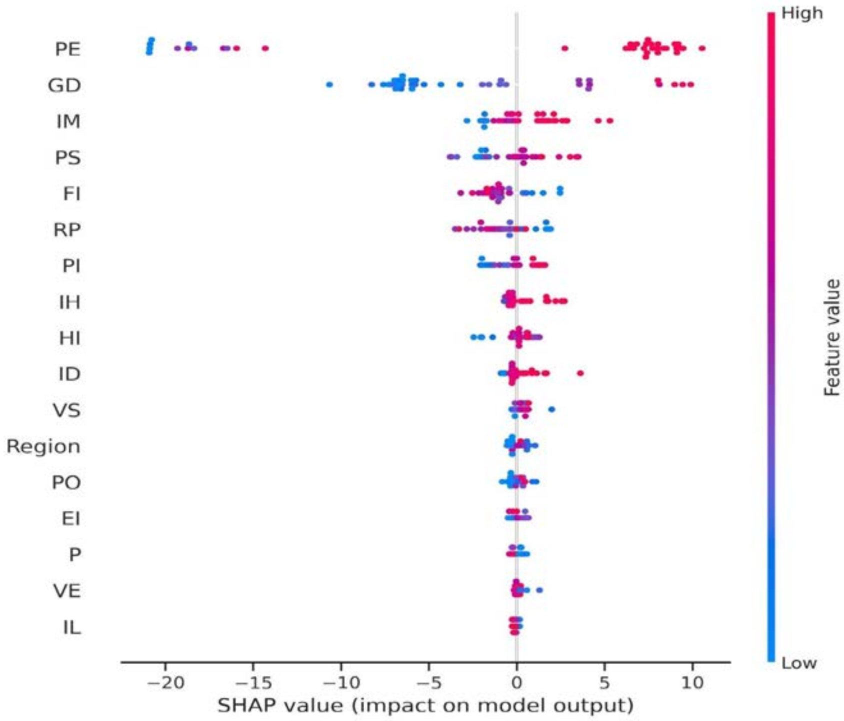
Figure 9. Permutation importance for RF regressor on the testing and training data.

Fig. 10a represents a global variable importance in descending order in terms of predictive power by taking the average of the SHAP value magnitudes across testing dataset. Red horizontal bars represent the positive impact of individual features on the predictions, and blue bars indicate the negative impact. Electrification rate is the most positively contributing variable, followed by the GDP variable. The measles immunization also plays a significant role in vaccination uptake. However, the fragile state index and the higher percentage of the rural population had a negative impact on the vaccination uptake. However, since we now have individualized explanations for all testing examples, we can compute feature importance for all testing examples as well which is shown in Fig. 10b. From the SHAP summary plot (Fig. 10b), we plot the SHAP values of features for all testing data to find the influential features as well as uncover the global view of the model. This beeswarm plot summarizes the distribution of SHAP values for each feature and features are sorted by the magnitudes of the SHAP values. Red color represents a high value feature whereas the blue color represents the low value feature. The placement of the dots on the x-axis represents the contributions of each feature value on the vaccination uptake prediction. In other words, red dots on the right side of zero shows the positive influence of the feature on vaccination coverage whereas red dots on the left represents the negative influence on the predictions. We see that greater access to electricity increases the percentage of fully vaccinated people, and so does the GDP of a country. On the other hand, the fragile state index and the higher percentage of rural population lower the predicted percentage of fully vaccinated people.

 doi:10.20944/preprints202206.0115.v1



(a) The relative importance of the predictors



(b) SHAP summary plot

Figure 10. Global feature importance using SHAP for the testing data. (a) The relative importance of the predictors, (b) SHAP summary plot.

5.4.2. Feature dependence analysis

Fig. 11 shows a single-variable partial dependence plot on the most influential 4 variables. These plots are not smooth due to the use of tree-based RF model [90]. We have integrated both individual (ICE) averaged one (PDP) in our partial dependence plots. We have utilized 20 ICE curves to make our plot readable. In top left PDP, dashed blue line shows us that the partial dependence of the percentage of fully vaccinated people against COVID-19 on the percentage of people with access to electricity in a country is nonlinear,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2022 doi:10.20944/preprints202206.0115.v1

and it increases sharply whenever a country with more than 80% of the population has electricity access. We see that partial dependence of the percentage of fully vaccinated people is monotonic increasing when the GDP of a country increases (top right). The percentage of fully vaccinated people is found to have a very weak partial dependence on the percentage of children of ages 12-23 months who took the Measles vaccine, and the fragile state index that is consistent with the variable importance, permutation importance, and global SHAP importance measures. We see a little influence on the percentage of fully vaccinated people (bottom left) when the percentage of children of ages 12-23 months who took the Measles vaccine is more than 90%. The bottom right plot shows a slight decrease of the percentage of fully vaccinated people for a higher fragile state index. However, we see some aberrations in ICE curves (light orange lines). We do not have many variations in the top left plot, but we have noticed a few exceptions in the top right plot where the percentage of fully vaccinated people slightly increases for an increase in GDP of a country. There are also a few exceptions in the bottom left plot as well where we observe a sudden increase in the percentage of fully vaccinated people whenever the percentage of children of ages 12-23 months who took the Measles vaccine is more than 90%. In the bottom right plot, few ICE curves remain constant for the fragile state index. Hence, ICE plots give us more explanations about the dependence of the prediction on a feature which were not vivid in PDPs. SHAP summary plot (Fig. 10b) provides us an overall summary of the feature contributions whereas the SHAP dependence plot shows the impact of feature value on the predictions. Also, PDP does not give a vertical spread to evaluate the influence of feature interactions on the predictions.

In Fig. 12, we plot the explanatory variables PE, GD, RP, and FI along with their corresponding SHAP values to visualize the interaction effects through the vertical spread. Interaction effects are shown by coloring other automatically selected variables. Fig. 12 clearly reveals that higher vaccination uptake can be achieved for $PE \geq 90$ and $PI \geq 80$. It is also evident from the existing literature [113] that access to information is ensured through electrification coverage and internet access. This validates the existence of interaction between PE and PI variables to predict the higher vaccination uptake. Countries with a GDP of more than $50, 000 and high education index are more likely to have a higher percentage of fully vaccinated people which clearly indicates an interaction effect between education index and GDP. Also, we see that a lower percentage of people aged 15-64 years tends to accelerate the vaccine uptake whenever more than 50% of people live in a rural area of a country. Furthermore, Fig. 12 demonstrates an inverse relationship between the vaccination uptake and the Fragile state index. A higher vaccination uptake goal can be achieved when $FI < 55$ and the $GDP > 30000$. This result establishes that a vulnerable country based on a Fragile state index along with lower GDP contributed to a very low vaccine uptake which is not a surprise at all. The other features are not shown in Fig. 12 due to the lack of dependency on vaccination uptake.

 doi:10.20944/preprints202206.0115.v1

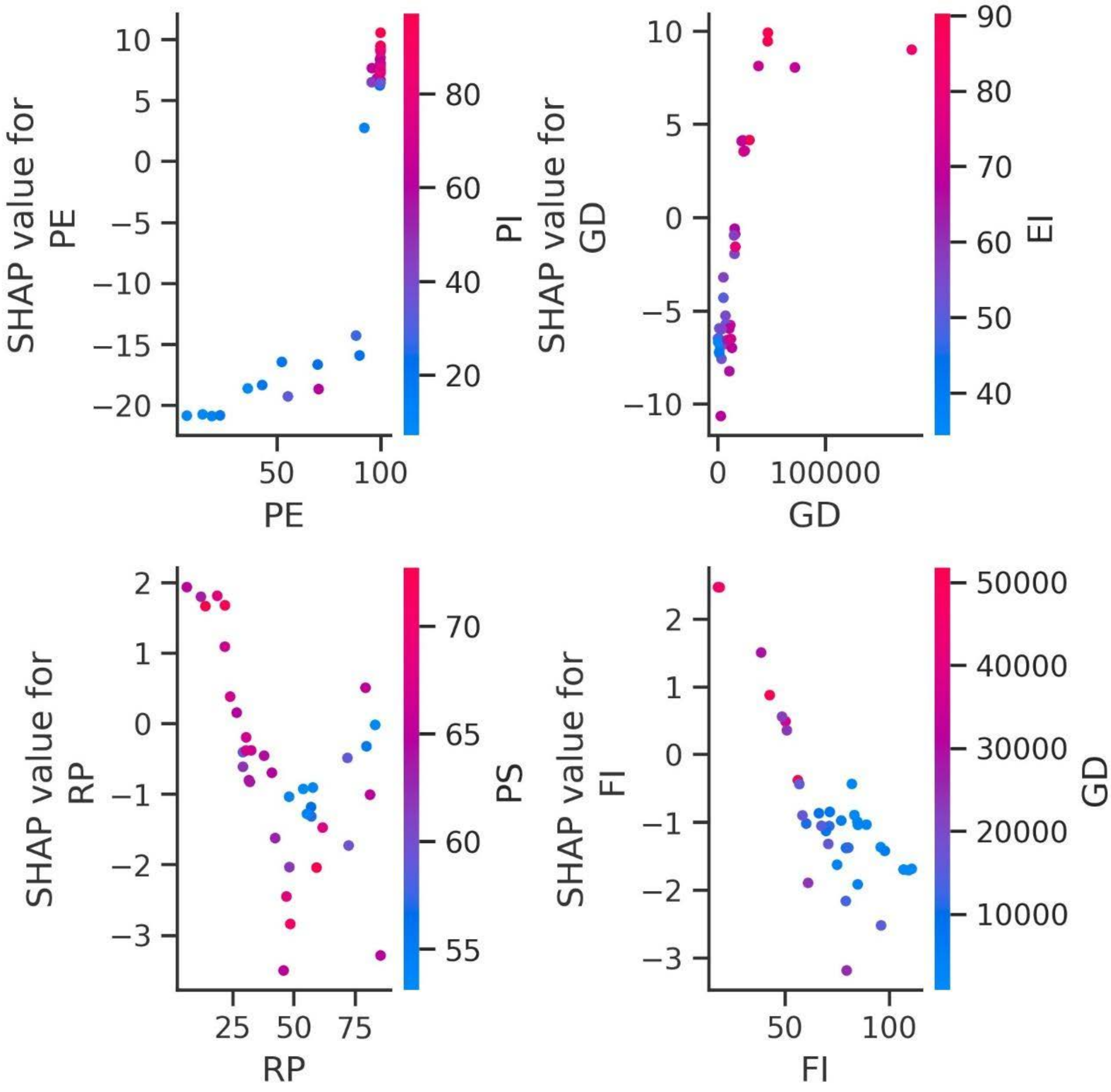


Figure 12. SHAP feature dependence plots for the testing data.

5.4.3. Feature Interaction

Pairwise interaction effects are shown in Fig. 13 where diagonals are the main effects and off diagonals are the interaction effects. From this absolute mean plot, we observe that the average main effect is higher for GDP and electrification coverage, i.e, these features have a greater influence on the RF model's predictions. There also exists a moderate interaction between them.

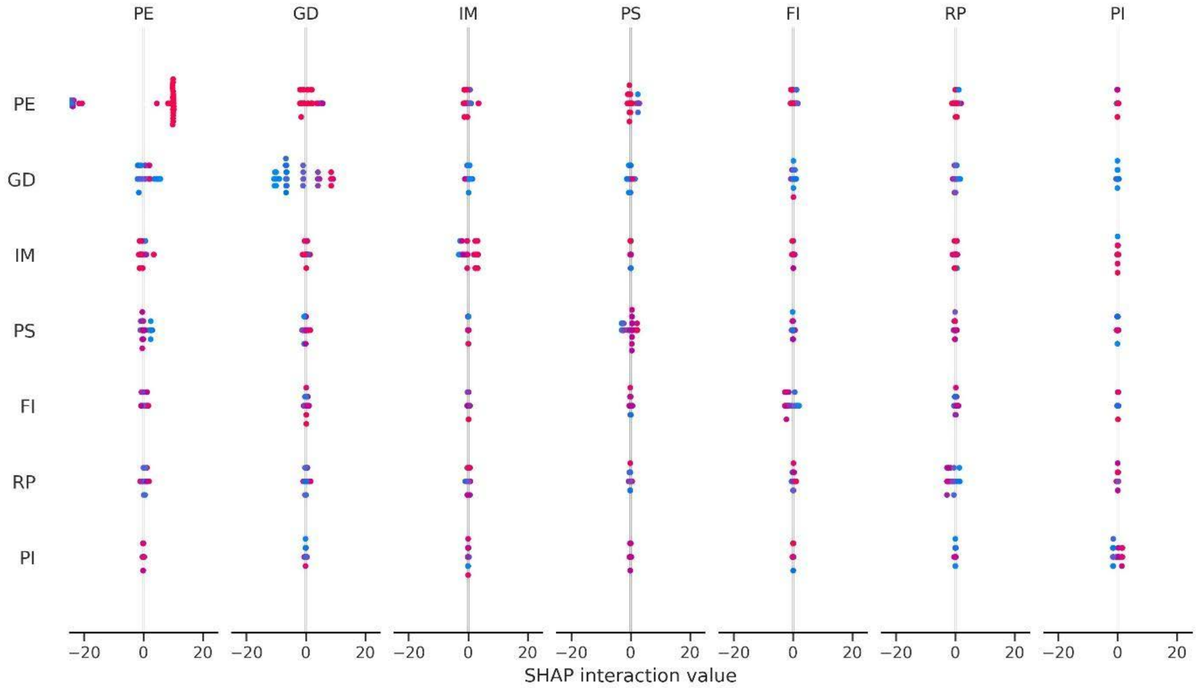


Figure 13. SHAP interaction value summary plot with the main and interaction effects of each predictor.

5.4.4. Local feature explanations

We can visualize the decision mechanism of a black-box model RF using the SHAP decision plot (Fig. 14). In Fig. 14, the vertical line represents the base value, and the colored line starting from the base value shows how models make predictions by adding SHAP values of the corresponding feature. Also, predictor values are included next to the prediction line in parenthesis to make the interpretations easy. In this figure, the features are arranged in descending order in terms of contribution to show the impact of each on the final predictions. The prediction line started with the base value of 48.8 for the 8th (Central African Republic) and 30th (Turkey) test samples. From the decision plot for the Central African Republic, we see that low GDP and poor energy infrastructure have significant contributions to the lower vaccination uptake which is aligned with our findings in Fig. 16b and Fig. 16a. However, we observe a different scenario for the 30th observation (Turkey), i.e., higher GDP and good energy infrastructure have significant impacts on the higher vaccination uptake although the Fragile state index had a negative influence on the predictions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2022 doi:10.20944/preprints202206.0115.v1

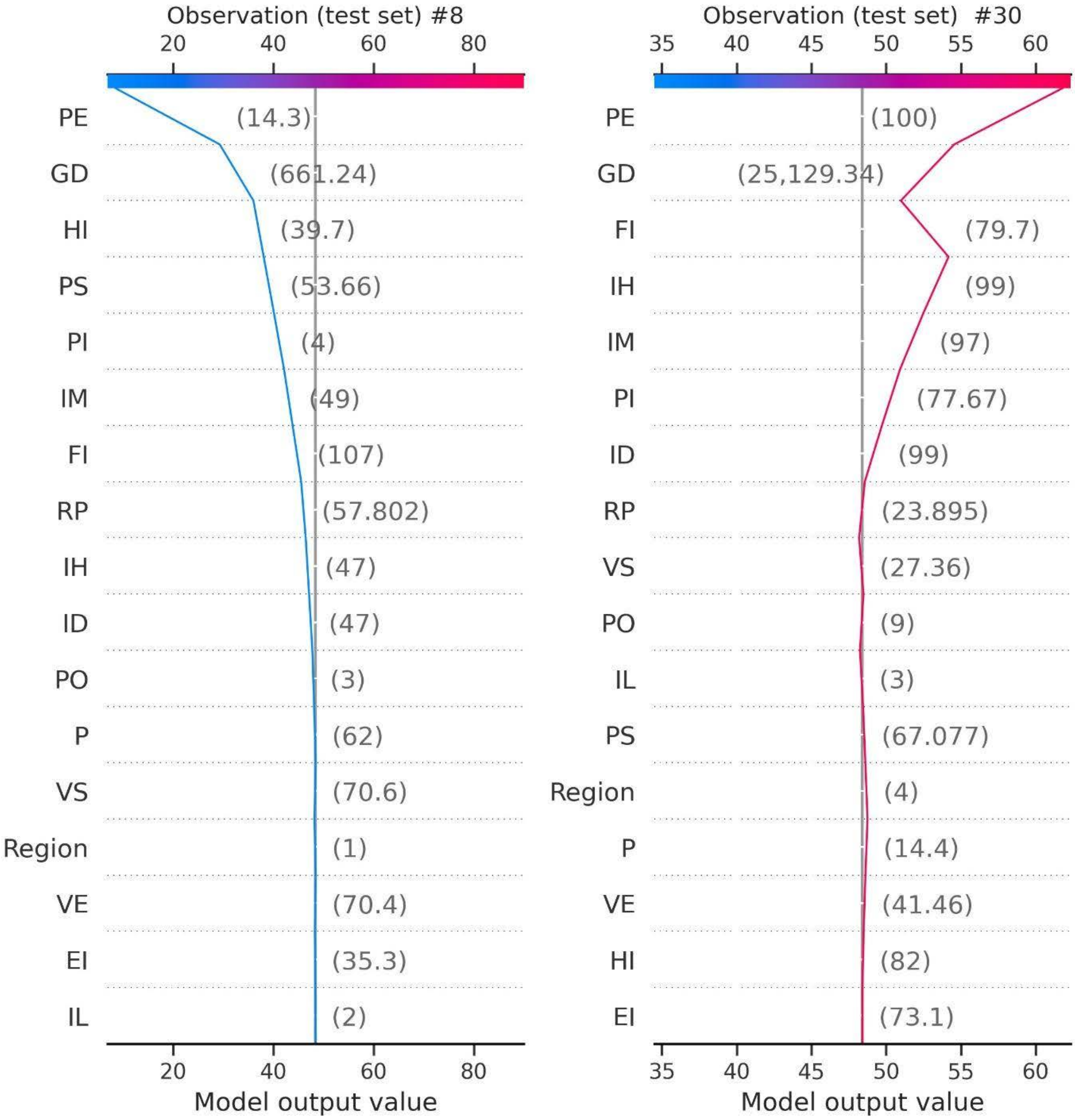
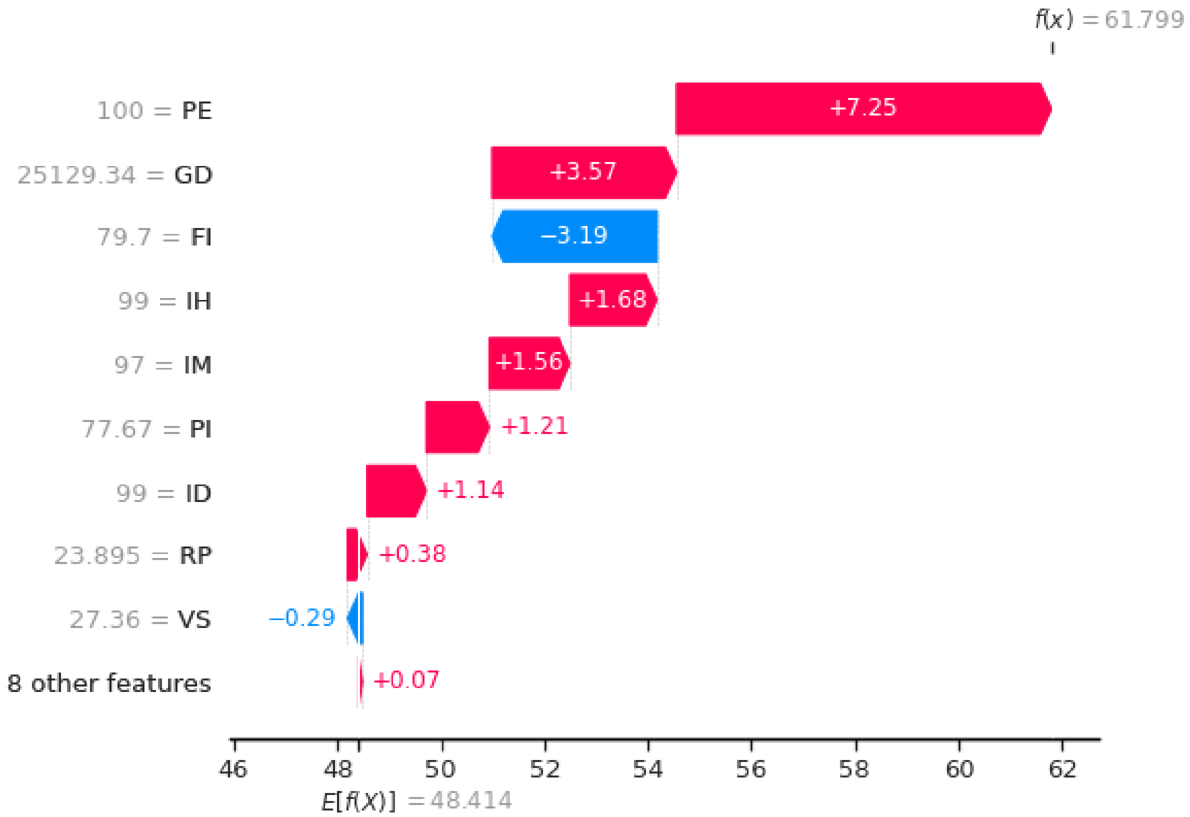


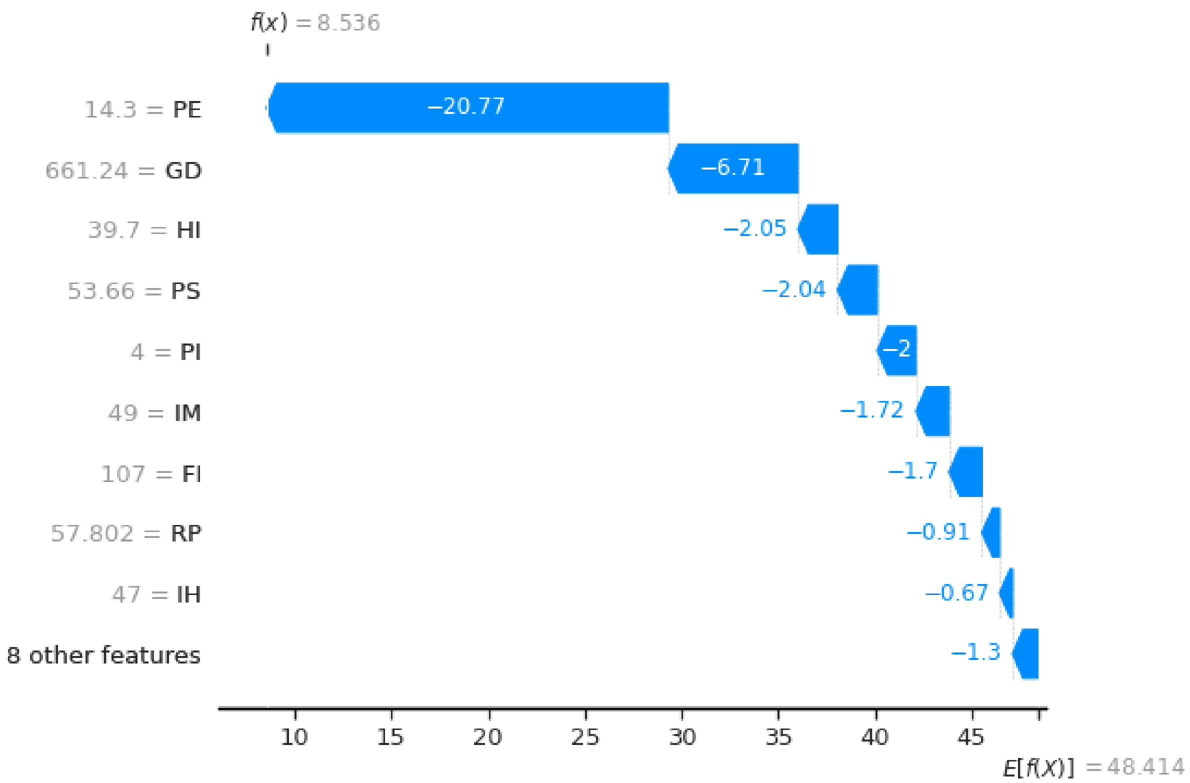
Figure 14. SHAP decision plot to explain predictions for each particular country.

The waterfall plot in Fig. 15a gives the visual representation of individual predictions by explaining the feature contributions. Final predictions are made by accumulating the positive (red) or negative (blue) contributions of each feature from the base expected value. From Fig. 15b, we see that the predicted vaccination uptake for observation Central African Republic is 8.54 which accumulation of each contribution with the base value of 48.41. Similarly, the RF model predicted that 61.80% of people will be fully vaccinated in Turkey. Moreover, feature contributions are aligned with our decision plots Fig. 14. We can achieve similar explanations by using the force plot in Fig. 16a.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2022 doi:10.20944/preprints202206.0115.v1

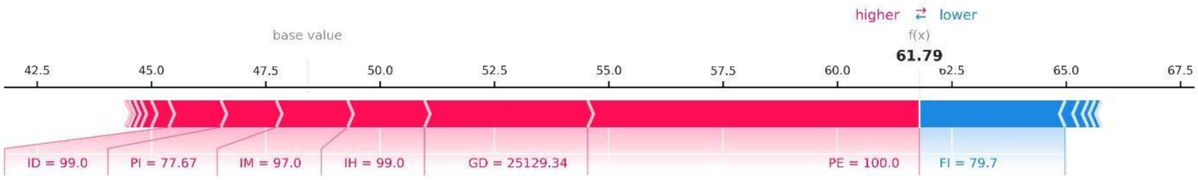


(a) Predicted vaccination uptake in Turkey

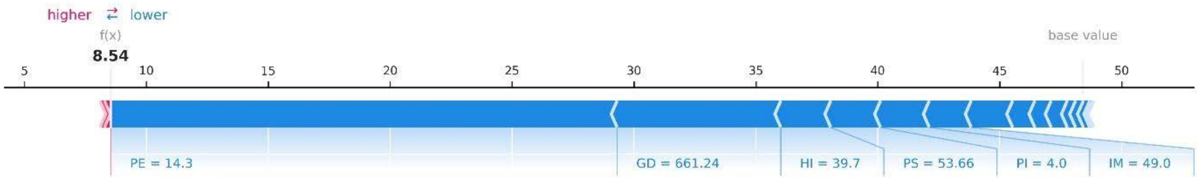


(b) Predicted vaccination uptake in Central Africa

Figure 15. Waterfall plots to explain the predicted vaccination uptake of two countries from the test sample.



(a) Predicted vaccination uptake on March 9 2022 in Turkey

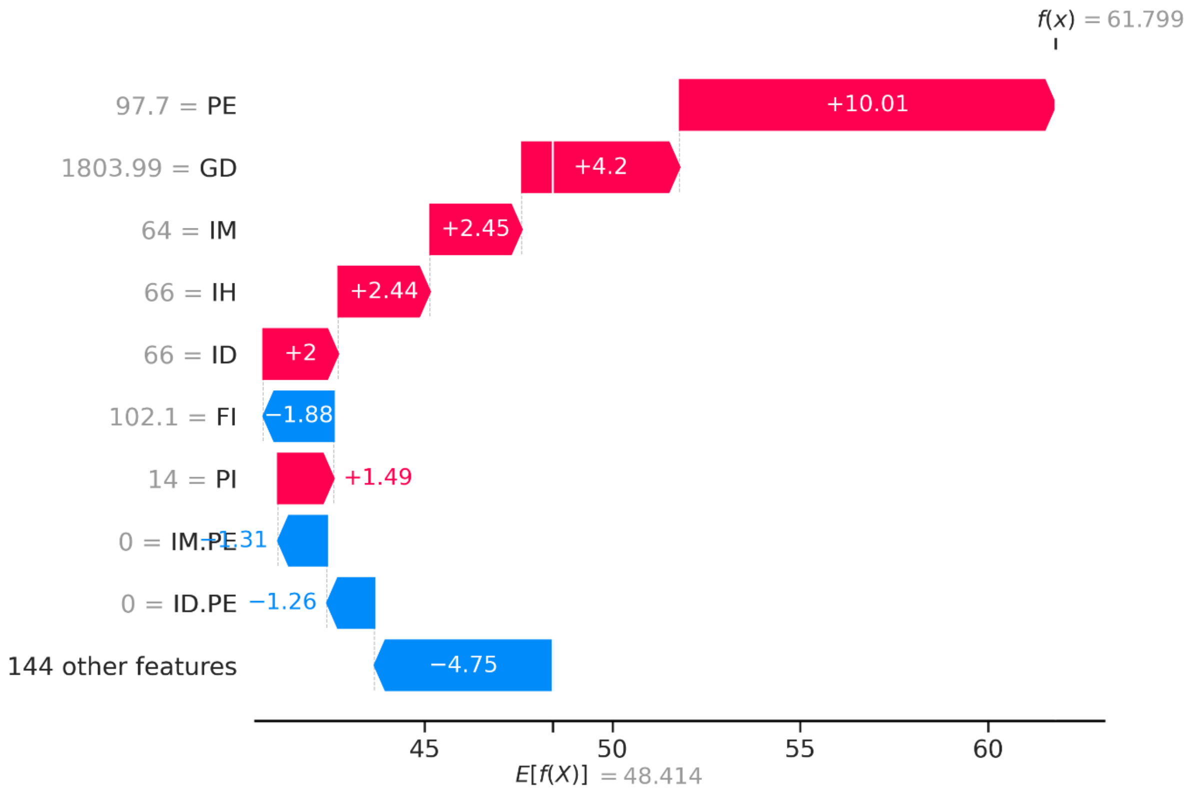


(b) Predicted vaccination uptake on March 9 2022 in Central African Republic

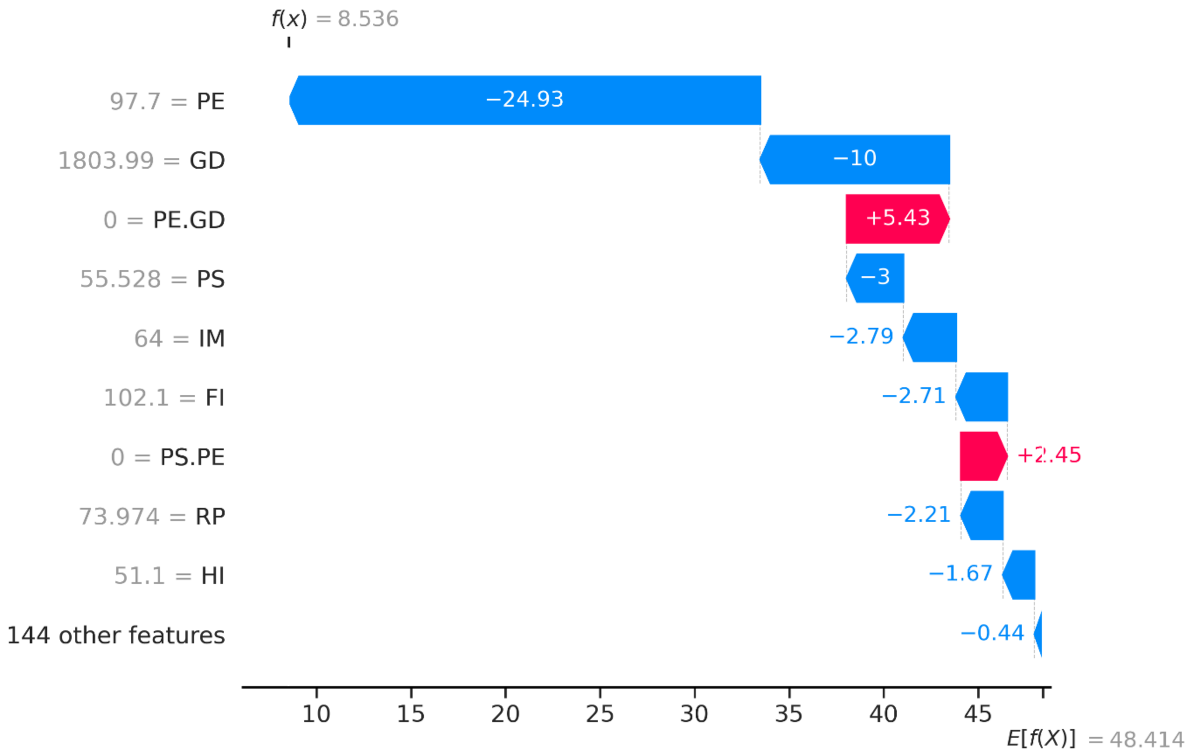
Figure 16. Force plots to explain the predicted vaccination uptake of two countries from the test sample.

Fig. 16 reveals that some features are responsible for increasing (red-colored) the vaccination uptake and some are responsible for decreasing (blue-colored) the percentage of fully vaccinated people of a country. In Fig. 16, predicted vaccination uptake is an accumulation of individual feature contributions where the red-colored arrow of features drives the predictions above the base value (to the right side of the base value) and the blue-colored arrow of features pulls the predictions below the base value (to the left side of the base value). The explanatory variables in Fig. 16a that propels the predicted value to be 61.79, which is almost close to the ground-truth value of 63.35. Feature values of PE, GD among others which are in red color push the predicted vaccination uptake higher, and FI feature in blue color pushes the predicted vaccination uptake lower. Feature values of PE, GD and FI contributed more to achieving such predicted value. The predicted vaccination uptake in Fig. 16b is heavily influenced by the PE and GD feature values and drives the prediction to a lower value. It is obvious from this force plot that the electrification coverage has a significant influence to change the direction of the predictions for both observations. Since the electrification coverage value for the Central African Republic is 14.3 which is smaller than the average value of 82.4, it facilitates the predictions to left. In contrast, it moves the predictions to the right as 100% (higher than the average 82.4) of people in Turkey have access to electricity. Fig. 17b is a waterfall plot for SHAP interaction values of observations 8 and 30. We see that the access to electricity main effect has decreased the predicted vaccination uptake by 24.93 whereas the PE.GD interaction effect has increased the predicted vaccination uptake by 5.43 for test sample 8. However, access to electricity main effect has increased the predicted vaccination uptake by 10.01 for test sample 30. Therefore, it is evident that the local importance is consistent with the global feature importance as shown in subsection 5.4.1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2022 doi:10.20944/preprints202206.0115.v1



(a) Predicted vaccination uptake on March 9 2022 in Turkey



(b) Predicted vaccination uptake on March 9 2022 in Central African Republic

Figure 17. Waterfall plots to explain the predicted vaccination uptake of two countries from the test sample.

In our analysis, we find electrification coverage (PE) is the most relevant predictor of the uptake of vaccination. Provision to electricity access has positive effects on household socioeconomic outcomes, such as education, health, and environment. Moreover, access

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 June 2022 doi:10.20944/preprints202206.0115.v1

to electricity increases the connectivity of the household and contributes to the improvement in socioeconomic welfare [114]. The benefit of electricity access also extends to increased availability of health services, access to communication technologies, and appropriate storage of vaccines and medicines [99]. Electricity coverage, by making modern technology (TV, Radio, Computer, Mobile Phone, etc.) available to the public, contributes to raising public awareness. Most importantly, electricity is a requirement for better vaccine storage, which can reduce vaccine waste and increase vaccine coverage. Therefore, our analysis reasonably reveals why coverage of the electricity is the most important predictor in vaccine uptake.

Another top feature that significantly predicts COVID-19 vaccination uptake is Gross Domestic Product (GDP). This finding is in line with [115-117]. The association between GDP and higher vaccination rate is, however, intuitive. Countries with higher GDP are more capable of allocating resources on healthcare. This consequently leads to a high rate of vaccination. As Roghani and Panahi [116] have a notably higher GDP, this can contribute to the higher vaccination rate due to the fact that GDP, national strategy, and proper utilization of the resources of a country are likely to be highly associated. A study conducted by Kazemi et al. [117] also found evidence of the association between median per capita income and overall vaccination adoption. The authors used the Random Forest algorithm for their analysis on a data set of 142 countries. Findings suggest that high income countries are likely to have higher COVID-19 vaccination adoption rates compared with low- and middle-income countries. On the other hand, Duan et al. [115] investigate the relationship between income level of a country and its COVID-19 vaccination coverage rates by conducting a cross-sectional ecological study on 138 countries. They find that the countries with higher income are more successful to have higher COVID-19 vaccination coverage rate. This finding is found robust when country specific demographic and health parameters are considered. GDP, which is an indicative measure of the relative prosperity of a country, also explains why inequality in accessing the COVID-19 vaccine is an issue for developing and underdeveloped countries of the world. Kazemi et al. [117] point out that the disparity in vaccine adoption is due to the fact that countries are different in terms of their income (low, medium, and high-income). Whereas developed countries succeeded in securing vaccines instantly when it was invented, many developing and underdeveloped countries had to deal with budget shortages in purchasing vaccines and providing required healthcare at the time of the pandemic. The disparity is still extant today. It is therefore not surprising that, as of April 3rd, 2022, a study found the highest-income countries had a per-capita vaccine rate of 195.39 vaccinations per 100 people, whereas lower middle income and lower-income countries had a vaccination rate of 114.56 per 100 and 22.10 per 100, respectively [118]. Securing vaccines from the producers and its distribution requires huge monetary allocation which many countries find as constraints in ensuring mass vaccination for their citizens. We find the percentage of measles immunization as another important feature predicting vaccination uptake. The study found evidence of a greater rate of vaccine hesitancy in regions where the childhood vaccine immunization rate is lower and vice versa [119]. This could partly be due to the fact that the benefit from childhood immunization is conducive in building public confidence on a newly offered vaccine. In addition to this, countries that have already-established public health infrastructures could readily offer vaccination to their citizens. This may in turn contribute to a higher rate of vaccination uptake. The wide acceptance of childhood immunization like measles can explain people's readiness in accepting COVID-19 vaccines. Fragility in a country could hinder vaccination campaigns and may limit vaccination access for certain segments of the population, including women and ethnic minorities. Also, fragility could have an adverse impact on vaccine supply chains for some developing countries [120]. This intuition is supported by our finding too as we find fragile state index lowers the predicted percentage of the fully vaccinated population.

6. Conclusion

In this research, we consider a couple of social, political, and economic factors that influence the uptake of COVID- 19 vaccines. Variables that we extracted for 182 countries are income level of a country, concern regarding vaccine safety and effectiveness, percentage of immunization against DPT, HepB3, and Measles, percentage of old and working age population, percentage of people living in rural areas, GDP per capita, poverty rate, percentage of people having internet access and availability of electricity, education index, human development index, Fragile state index, and availability of approved COVID-19 vaccines. We introduced an explainable machine learning technique to analyze the data of these 182 countries. Our results showed the following findings:

- Availability of internet access and electricity, the percentage of children of ages 12-23 months who took the Measles vaccine, and countries' GDP are the crucial factors for the strong acceptance of COVID-19 vaccine.
- Fragile state index and the percentage of people living in rural areas have a negative influence on the acceptance of COVID-19 vaccine.
- Percentage of poverty, income level, and older age people have no or little influence on the COVID-19 vaccine uptake.

These findings can be helpful in effective policy making to combat any future public health crisis like the COVID-19 pandemic. From our recent experience, we see wealthy nations have disproportionately controlled the production and distribution of the COVID-19 vaccines. The observed inequality in distributing COVID-19 vaccination among nations may linger the pandemic situation as a combined effort is required to end the pandemic. On the other hand, many developing and underdeveloped countries have been overlooking the necessity of improving public health infrastructure for ages. Additionally, the ongoing instability in different parts of the world restricts the distribution of COVID-19 vaccines in the conflict region. These challenges need to be addressed on a priority basis to solve the problem, Vaccine Hesitancy if there is any.

References

[1] Worldometer, https://www.worldometers.info/gdp/gdp-per-capita/[accessed 03 September 2021].

[2] Wu YC, Chen CS, Chan YJ. The outbreak of COVID-19: An overview. J. Chin. Med. Assoc. 2020; 83(3): 217-220. DOI: https://doi.org/10.1097/JCMA.0000000000000270

[3] Ahmed J, Jaman MH, Saha G, Ghosh P, Effect of environmental and socio-economic factors on the spreading of COVID-19 at 70 cities/provinces. Heliyon. 2021; 7: e06979. DOI: https://doi.org/10.1016/j.heliyon.2021.e06979

[4] Hawlader MDH, Rahman ML, Nazir A, Ara T, Haque MMA, Saha S, et al. COVID-19 vaccine acceptance in South Asia: a multi-country study. International Journal of Infectious Diseases. 2022; 114: 1–10. DOI: https://doi.org/10.1016/j.ijid.2021.09.056

[5] Mondal P, Sinharoy A, Su L, Sociodemographic predictors of COVID-19 vaccine acceptance: a nationwide us-based survey study. Public Health. 2021; 198: 252–259. DOI: https://doi.org/10.1016/j.puhe.2021.07.028

[6] Gagneux-Brunon A, Botelho-Nevers E, Bonneton M, Peretti-Watel P, Verger P, Launay O, et al. Public opinion on a mandatory COVID-19 vaccination policy in France: a cross-sectional survey. Clinical Microbiology and Infection. 2022; 28: 433–439. DOI: https://doi.org/10.1016/j.cmi.2021.10.016

[7] Tucker JS, D'Amico EJ, Pedersen ER, Garvey R, Rodriguez A, Klein DJ. COVID-19 vaccination rates and attitudes among young adults with recent experiences of homelessness. Journal of Adolescent Health. 2022; 70: 504–506. DOI: https://doi.org/10.1016/j.jadohealth.2021.11.017

[8] Byrne T, Patel P, Shrotri M, Beale S, Michie S, Butt J, et al. Trends, patterns and psychological influences on COVID-19 vaccination intention: Findings from a large prospective community cohort study in England and Wales (virus watch). Vaccine. 2021; 39: 7108–7116. DOI: https://doi.org/10.1016/j.vaccine.2021.09.066

[9] Mouter, N, de Ruijter A, Ardine de Wit G, Lambooij MS, van Wijhe M, van Exel J, et al. "please, you go first!" preferences for a COVID-19 vaccine among adults in the Netherlands. Social Science Medicine. 2022; 292:

114626. DOI: https://doi.org/10.1016/j.socscimed.2021.114626

[10] Malik AA, McFadden SM, Elharake J, Omer SB. Determinants of COVID-19 vaccine acceptance in the US. EClinicalMedicine. 2020; 26: 100495. DOI: https://doi.org/10.1016/j.eclinm.2020.100495

[11] Urrunaga-Pastor D, Bendezu-Quispe G, Herrera-Añazco P, Uyen-Cateriano A, Toro-Huamanchumo CJ, Rodriguez-Morales AJ, et al. Cross-sectional analysis of COVID-19 vaccine intention, perceptions and hesitancy across Latin America and the Caribbean. Travel Medicine and Infectious Disease. 2021; 41: 102059. DOI: https://doi.org/10.1016/j.tmaid.2021.102059

[12] Cascini F, Pantovic A, Al-Ajlouni Y, Failla G, Ricciardi W. Attitudes, acceptance and hesitancy among the general population worldwide to receive the COVID-19 vaccines and their contributing factors: A systematic review. EClinicalMedicine. 2021; 40: 101113. DOI: https://doi.org/10.1016/j.eclinm.2021.101113

[13] Kessels R, Luyten J, Tubeuf S. Willingness to get vaccinated against COVID-19 and attitudes toward vaccination in general. Vaccine. 2021; 39: 4716–4722. DOI: https://doi.org/10.1016/j.vaccine.2021.05.069

[14] Gbeasor-Komlanvi F, Afanvi K, Konu Y, Agbobli Y, Sadio A, Tchankoni M, et al. Prevalence and factors associated with COVID-19 vaccine hesitancy in health professionals in Togo, 2021. Public Health in Practice. 2021; 2: 100220. DOI: https://doi.org/10.1016/j.puhip.2021.100220

[15] Patelarou E, Galanis P, Mechili EA, Argyriadi A, Argyriadis A, Asimakopoulou E, et al. Factors influencing nursing students' intention to accept COVID-19 vaccination: A pooled analysis of seven European countries. Nurse Education Today. 2021; 104: 105010. DOI: https://doi.org/10.1016/j.nedt.2021.105010

[16] Potdar M, Potdar S, Potdar M. A study of gender disparities towards COVID-19 vaccination drive in Maharashtra state, India. Diabetes Metabolic Syndrome: Clinical Research Reviews. 2021; 15: 102297. DOI: https://doi.org/10.1016/j.dsx.2021.102297

[17] Prickett KC, Habibi H, Carr PA. COVID-19 vaccine hesitancy and acceptance in a cohort of diverse New Zealanders. Lancet Reg Health West Pac. 2021; 14: 100241. DOI: https://doi.org/10.1016/j.lanwpc.2021.100241

[18] Guillon M, Kergall P. Factors associated with COVID-19 vaccination intentions and attitudes in France. Public Health. 2021; 198: 200–207. DOI: https://doi.org/10.1016/j.puhe.2021.07.035

[19] Mir HH, Parveen S, Mullick NH, Nabi S. Using structural equation modeling to predict Indian people's attitudes and intentions towards COVID-19 vaccination. Diabetes Metabolic Syndrome: Clinical Research Reviews. 2021; 15: 1017–1022. DOI: https://doi.org/10.1016/j.dsx.2021.05.006

[20] Mottelson A, Vandeweerdt C , Atchapero M, Luong T, Holz C, Böhm R, et al. A self-administered virtual reality intervention increases COVID-19 vaccination intention. Vaccine. 2021; 39: 6746–6753. DOI: https://doi.org/10.1016/j.vaccine.2021.10.004

[21] Bell S, Clarke R, Mounier-Jack S, Walker JL, Paterson P. Parents' and guardians' views on the acceptability of a future COVID-19 vaccine: A multi-methods study in England. Vaccine. 2020; 38: 7789–7798. DOI: https://doi.org/10.1016/j.vaccine.2020.10.027

[22] Toth-Manikowski SM, Swirsky ES, Gandhi R, Piscitello G. COVID-19 vaccination hesitancy among health care workers, communication, and policy-making. American Journal of Infection Control. 2022; 50: 20–25. DOI: https://doi.org/10.1016/j.ajic.2021.10.004

[23] Humble RM, Sell H, Dubé E, MacDonald NE, Robinson J, Driedger SM, et al. Canadian parents' perceptions of COVID-19 vaccination and intention to vaccinate their children: Results from a cross-sectional national survey. Vaccine. 2021; 39: 7669–7676. DOI: https://doi.org/10.1016/j.vaccine.2021.10.002

[24] Perry M, Akbari A, Cottrell S, Gravenor MB, Roberts R, Lyons RA, et al. Inequalities in coverage of COVID-19 vaccination: A population register-based cross-sectional study in Wales, UK. Vaccine. 2021; 39: 6256–6261. DOI: https://doi.org/10.1016/j.vaccine.2021.09.019

[25] Troiano G, Nardi A. Vaccine hesitancy in the era of COVID-19. Public Health. 2021; 194: 245–251. DOI: https://doi.org/10.1016/j.puhe.2021.02.025

[26] Khan Z, Allana R, Afzal I, Ali A, Mariam O, Aslam R, et al. Assessment of attitude and hesitancy toward a

vaccine against COVID-19 in a Pakistani population: A mix methods survey. Vacunas. 2021. DOI: https://doi.org/10.1016/j.vacun.2021.08.002

[27] Zheng H, Jiang S, Wu Q. Factors influencing COVID-19 vaccination intention: The roles of vaccine knowledge, vaccine risk perception, and doctor-patient communication. Patient Education and Counseling. 2022; 105: 277–283. DOI: https://doi.org/10.1016/j.pec.2021.09.023

[28] Al-Hanawi MK, Ahmad K, Haque R, Keramat SA. Willingness to receive COVID-19 vaccination among adults with chronic diseases in the kingdom of Saudi Arabia. Journal of Infection and Public Health. 2021; 14: 1489–1496. DOI: https://doi.org/10.1016/j.jiph.2021.08.002

[29] Borriello A, Master D, Pellegrini A, Rose JM. Preferences for a COVID-19 vaccine in Australia. Vaccine. 2021; 39: 473–479. DOI: https://doi.org/10.1016/j.vaccine.2020.12.032

[30] Chakraborty C, Sharma AR, Bhattacharya M, Agoramoorthy G, Lee SS. The current second wave and COVID-19 vaccination status in India. Brain, Behavior, and Immunity. 2021; 96: 1–4. DOI: https://doi.org/10.1016/j.bbi.2021.05.018

[31] Marzo R, Ahmad A, Abid K, Khatiwada A, Ahmed A, Kyaw T, et al. Factors influencing the acceptability of COVID-19 vaccination: A cross-sectional study from Malaysia. Vacunas. 2021. DOI: https://doi.org/10.1016/j.vacun.2021.07.007

[32] Fadda M, Suggs LS, Albanese E. Willingness to vaccinate against COVID-19: A qualitative study involving older adults from southern Switzerland. Vaccine: X. 2021; 8: 100108. DOI: https://doi.org/10.1016/j.jvacx.2021.100108

[33] Luma AH, Haveen AH, Faiq BB, Stefania M, Leonardo EG. Hesitancy towards COVID-19 vaccination among the healthcare workers in Iraqi Kurdistan. Public Health in Practice. 2022; 3: 100222. DOI: https://doi.org/10.1016/j.puhip.2021.100222

[34] Sherman S, Sim J, Dasch MHC, Amlôt R, Rubin G, Sevdalis N et al. COVID-19 vaccination acceptability in the UK at the start of the vaccination programme: a nationally representative cross-sectional survey (covaccs – wave 2). Public Health. 2022; 202: 1–9. DOI: https://doi.org/10.1016/j.puhe.2021.10.008

[35] Wong MC, Wong EL, Huang J, Cheung AW, Law K, Chong MK, et al. Acceptance of the COVID-19 vaccine based on the health belief model: A population-based survey in Hong Kong. Vaccine. 2021; 39: 1148–1156. DOI: https://doi.org/10.1016/j.vaccine.2020.12.083

[36] Wang Z, Xiao J, Jiang F, Li J, Yi Y, Min W, et al. The willingness of Chinese adults to receive the COVID-19 vaccine and its associated factors at the early stage of the vaccination programme: a network analysis. Journal of Affective Disorders. 2022; 297: 301–308. DOI: https://doi.org/10.1016/j.jad.2021.10.088

[37] Valckx S, Crèvecoeur J, Verelst F, Vranckx M, Hendrickx G, Hens N, et al. Individual factors influencing COVID-19 vaccine acceptance in between and during pandemic waves (July–December 2020). Vaccine. 2022; 40(1): 151–161. DOI: https://doi.org/10.1016/j.vaccine.2021.10.073

[38] Myers A, Ipsen C, Lissau A. COVID-19 vaccination hesitancy among Americans with disabilities aged 18-65: An exploratory analysis. Disability and Health Journal. 2022; 15: 101223. DOI: https://doi.org/10.1016/j.dhjo.2021.101223

[39] Caspi G, Dayan A, Eshal Y, Liverant-Taub S, Twig G, Shalit U, et al. Socioeconomic disparities and COVID-19 vaccination acceptance: a nationwide ecologic study. Clinical Microbiology and Infection. 2021; 27: 1502–1506. DOI: https://doi.org/10.1016/j.cmi.2021.05.030

[40] Wang J, Lyu Y, Zhang H, Jing R, Lai X, Feng H, et al. Willingness to pay and financing preferences for COVID-19 vaccination in China. Vaccine. 2021; 39: 1968–1976. DOI: https://doi.org/10.1016/j.vaccine.2021.02.060

[41] Desson Z, Kauer L, Otten T, Peters JW, Paolucci F. Finding the way forward: COVID-19 vaccination progress in Germany, Austria and Switzerland. Health Policy and Technology. 2021; 100584. DOI: https://doi.org/10.1016/j.hlpt.2021.100584

[42] Brown CC, Young SG, Pro GC. COVID-19 vaccination rates vary by community vulnerability: A county-level

analysis. Vaccine. 2021; 39: 4245–4249. DOI: https://doi.org/10.1016/j.vaccine.2021.06.038

[43] Salibi NE, Abdulrahim S, Haddad ME, Bassil S, Khoury ZE, Ghattas H, et al. P101 COVID-19 vaccine acceptance in older Syrian refugees in Lebanon: preliminary findings from a longitudinal survey. Journal of Epidemiology & Community Health. 2021; 75: A87–A87. DOI: http://dx.doi.org/10.1136/jech-2021-SSMabstracts.187

[44] Adigwe OP. COVID-19 vaccine hesitancy and willingness to pay: Emergent factors from a cross-sectional study in Nigeria. Vaccine: X. 2021; 9: 100112. DOI: https://doi.org/10.1016/j.jvacx.2021.100112

[45] Haque MMA, Rahman ML, Hossian M, Matin KF, Nabi MH, Saha S, et al. Acceptance of COVID-19 vaccine and its determinants: evidence from a large sample study in Bangladesh. Heliyon. 2021; 7: e07376. DOI: https://doi.org/10.1016/j.heliyon.2021.e07376

[46] Fisher KA, Nguyen N, Crawford S, Fouayzi H, Singh S, Mazor KM. Preferences for COVID-19 vaccination information and location: Associations with vaccine hesitancy, race, and ethnicity. Vaccine. 2021; 39: 6591–6594. DOI: https://doi.org/10.1016/j.vaccine.2021.09.058

[47] Luo C, Yang Y, Liu Y, Zheng D, Shao L, Jin J, et al. Intention to COVID-19 vaccination and associated factors among health care workers: A systematic review and meta-analysis of cross-sectional studies. American Journal of Infection Control. 2021; 49: 1295–1304. DOI: https://doi.org/10.1016/j.ajic.2021.06.020

[48] Shaham A, Chodick G, Shalev V, Yamin D. Personal and social patterns predict influenza vaccination decision. BMC Public Health. 2020; 20: 222. DOI: https://doi.org/10.1186/s12889-020-8327-3

[49] Ayachit SS, Kumar T, Deshpande S, Sharma N, Chaurasia K, Dixit M. Predicting H1N1 and seasonal flu: Vaccine cases using ensemble learning approach, In: 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN). 2020: 172–176. DOI: https://doi.org/10.1109/icacccn51052.2020.9362909

[50] Mannion N. Predictions of changes in child immunization rates using an automated approach: USA. 2020. Link: http://norma.ncirl.ie/4368/

[51] Grandhi GR, Mszar R, Vahidy F, Valero-Elizondo J, Blankstein R, Blaha MJ, et al. Sociodemographic Disparities in Influenza Vaccination Among Adults With Atherosclerotic Cardiovascular Disease in the United States. JAMA Cardiology. 2021; 6: 87–91. DOI: https://doi.org/10.1001/jamacardio.2020.3978

[52] Hayawi K, Shahriar S, Serhani M, Taleb I, Mathew S. Anti-vax: A novel Twitter dataset for COVID-19 vaccine misinformation detection. Public Health. 2022; 203: 23–30. DOI: https://doi.org/10.1016/j.puhe.2021.11.022

[53] Mewhirter J, Sagir M, Sanders R. Towards a predictive model of COVID-19 vaccine hesitancy among American adults. Vaccine. 2022; 40: 1783–1789. DOI: https://doi.org/10.1016/j.vaccine.2022.02.011

[54] Hafizh M, Badri Y, Mahmud S, Hafez A, Choe P. COVID-19 vaccine willingness and hesitancy among residents in Qatar: a quantitative analysis based on machine learning. Journal of Human Behavior in the Social Environment. 2021; 1–24. DOI: https://doi.org/10.1080/10911359.2021.1973642

[55] Mondal P, Sinharoy A, Su L. Sociodemographic predictors of COVID-19 vaccine acceptance: A nationwide US-based survey study. Public Health. 2021; 198: 252–259. DOI: https://doi.org/10.1016/j.puhe.2021.07.028

[56] Jayasurya GG, Kumar S, Singh BK, Kumar V. Analysis of public sentiment on COVID-19 vaccination using twitter. IEEE Transactions on Computational Social Systems. 2021; 1–11. DOI: https://doi.org/10.1109/TCSS.2021.3122439

[57] Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S. An analysis of COVID-19 vaccine sentiments and opinions on twitter. International Journal of Infectious Diseases. 2021; 108: 256–262. DOI: https://doi.org/10.1016/j.ijid.2021.05.059

[58] Melton CA, Olusanya OA, Ammar N, Shaban-Nejad A. Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the reddit social media platform: A call to action for strengthening vaccine confidence. Journal of Infection and Public Health. 2021; 14(10): 1505–1512. DOI: https://doi.org/10.1016/j.jiph.2021.08.010

[59] Cheong Q, Au-yeung M, Quon S, Concepcion K, Kong J. Predictive modeling of vaccination uptake in US counties: A machine learning-based approach (preprint). Journal of Medical Internet Research. 2021; 23(11): e33231. DOI: https://doi.org/10.2196/33231

[60] Carrieri V, Lagravinese R, Resce G. Predicting vaccine hesitancy from area-level indicators: A machine learning approach. Health Economics. 2021; 30: 3248–3256. DOI: https://doi.org/10.1002/hec.4430

[61] The world bank, https://data.worldbank.org/country, [accessed 03 September 2021].

[62] Human development data center, https://hdr.undp.org/en/data, [accessed 03 September 2021].

[63] Fragile states index, https://fragilestatesindex.org/, [accessed 03 September 2021].

[64] Wellcome global monitor 2020: COVID-19. https://wellcome.org/reports/wellcome-global-monitor-COVID-19/2020 [accessed 03 September 2021].

[65] Johns Hopkins University Medicine. https://coronavirus.jhu.edu/vaccines/international [accessed 03 September 2021].

[66] The New York Times. https://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html. [accessed 03 September 2021].

[67] Arce JSS, Warren SS, Meriggi NF, Scacco A, McMurry N, Voors M, et al. COVID-19 vaccine acceptance and hesitancy in low-and middle-income countries. Nature Medicine. 2021; 27: 1385–1394. DOI: https://doi.org/10.1038/s41591-021-01454-y

[68] Africa CDC. Majority of Africans would take a safe and effective COVID-19 vaccine, 2021. Link: https://africacdc.org/news-item/majority-of-africans-would-take-a-safe-and-effective-covid-19-vaccine/ [accessed 12 January 2022].

[69] Wouters OJ, Shadlen KC, Salcher-Konrad M, Pollard AJ, Larson HJ, Teerawattananon Y, et al. Challenges in ensuring global access to COVID-19 vaccines: production, affordability, allocation, and deployment. The Lancet. 2021; 397: 1023–1034. DOI: https://doi.org/10.1016/S0140-6736(21)00306-8

[70] Detoc M, Bruel S, Frappe P, Tardy B, Botelho-Nevers E, Gagneux-Brunon A. Intention to participate in a COVID-19 vaccine clinical trial and to get vaccinated against COVID-19 in France during the pandemic. Vaccine. 2020; 38: 7002–7006. DOI: https://doi.org/10.1016/j.vaccine.2020.09.041

[71] Kerr JR, Schneider CR, Recchia G, Dryhurst S, Sahlin U, Dufouil C, et al. Predictors of COVID-19 vaccine acceptance across time and countries. MedRxiv. 2020. DOI: https://doi.org/10.1101/2020.12.09.20246439

[72] Szilagyi PG, Thomas K, Shah MD, Vizueta N, Cui Y, Vangala S, et al. National trends in US public's likelihood of getting a COVID-19 vaccine—April 1 to December 8, 2020. Jama. 2021; 325: 396–398. DOI: https://doi.org/10.1001/jama.2020.26419

[73] De Figueiredo A, Simas C, Karafillakis E, Paterson P, Larson HJ. Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modeling study. The Lancet. 2020; 396: 898–908. DOI: https://doi.org/10.1016/S0140-6736(20)31558-0

[74] Larson H, De Figueiredo A, Xiahong Z, Schulz W, Verger P, Johnston I, et al. The state of vaccine confidence 2016: global insights through a 67-country survey. Ebiomedicine. 2016; 12: 295–301. DOI: https://doi.org/10.1016/j.ebiom.2016.08.042

[75] Murthy BP, Sterrett N, Weller D, Zell E, Reynolds L, Toblin RL, et al. Disparities in COVID-19 vaccination coverage between urban and rural counties—united states, December 14, 2020–April 10, 2021. Morbidity and Mortality Weekly Report. 2021; 70: 759. Link: https://www.cdc.gov/mmwr/volumes/70/wr/mm7020e3.htm

[76] Lee J, Huang Y. COVID-19 vaccine hesitancy: The role of socioeconomic factors and spatial effects. Vaccines. 2022; 10: 352. DOI: https://doi.org/10.3390/vaccines10030352

[77] Simas C, Larson HJ. Overcoming vaccine hesitancy in low-income and middle-income regions. Nat Rev Dis Primers. 2021; 7: 41. DOI: https://doi.org/10.1038/s41572-021-00279-w

[78] Machingaidze S, Wiysonge CS. Understanding COVID-19 vaccine hesitancy. Nature Medicine. 2021; 27: 1338–1339. DOI: https://doi.org/10.1038/s41591-021-01459-7

[79] Roghani A. The relationship between macro-socioeconomics determinants and COVID-19 vaccine distribution. AIMS Public Health. 2021; 8(4): 655-664. DOI: https://doi.org/10.3934/publichealth.2021052
[80] Bwire G, Ario AR, Eyu P, Ocom F, Wamala JF, Kusi KA, et al. The COVID-19 pandemic in the African continent. BMC Medicine. 2022; 20: 167. DOI: https://doi.org/10.1186/s12916-022-02367-4
[81] de Oliveira BRB, da Penha Sobral AIG, Marinho MLM, et al. Determinants of access to the SARS-CoV-2 vaccine: a preliminary approach. International Journal for Equity in Health. 2021; 20: 183. DOI: https://doi.org/10.1186/s12939-021-01520-4
[82] de Figueiredo A, Larson HJ. Exploratory study of the global intent to accept COVID-19 vaccinations. Commun Med. 2021; 1: 30. DOI: https://doi.org/10.1038/s43856-021-00027-x
[83] Kazemi M, Bragazzi NL, Kong JD. Assessing Inequities in COVID-19 Vaccine Roll-Out Strategy Programs: A Cross-Country Study Using a Machine Learning Approach. Vaccines. 2022; 10(2): 194. DOI: https://doi.org/10.3390/vaccines10020194
[84] COVID-19 and human development: Assessing the Crisis, Envisioning the Recovery, 2020 Human Development Perspectives, UNDP. Link: https://hdr.undp.org/en/hdp-covid
[85] Asundi A, O'Leary C, Bhadelia N. Global COVID-19 vaccine inequity: The scope, the impact, and the challenges. Cell Host & Microbe. 2021; 29(7): 1036–1039. DOI: https://doi.org/10.1016/j.chom.2021.06.007
[86] Van De Pas R, Widdowson MA, Ravinetto R, Srinivas PN, Ochoa TJ, Fofana TO, et al. COVID-19 vaccine equity: a health systems and policy perspective. Expert Rev Vaccines. 2022; 21(1): 25-36. DOI: https://doi.org/10.1080/14760584.2022.2004125
[87] Lawal L, Aminu MB, Murwira T, Avoka C, Yusuf SM, Harrison LO, et al. Low coverage of COVID-19 vaccines in Africa: current evidence and the way forward. Hum Vaccin Immunother. 2022; 18(1): 2034457. DOI: https://doi.org/10.1080/21645515.2022.2034457
[88] Mutombo PN, Fallah MP, Munodawafa D, Kabel A, Houeto D, Goronga T, et al. COVID-19 vaccine hesitancy in Africa: A call to action. Lancet Glob Health. 2022; 10(3): e320-e321. DOI: https://doi.org/10.1016/S2214-109X(21)00563-5
[89] Loembé MM, Nkengasong JN. COVID-19 vaccine access in Africa: Global distribution, vaccine platforms, and challenges ahead. Immunity. 2021; 54(7): 1353-1362. DOI: https://doi.org/10.1016/j.immuni.2021.06.017
[90] Jecker NS. Global sharing ofCOVID-19 vaccines: A duty of justice, not charity. Developing World Bioethics. 2022; 1–10. DOI: https://doi.org/10.1111/dewb.12342
[91] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer series in statistics, Springer, 2009. Link: https://books.google.com/books?id=eBSgoAEACAAJ.
[92] Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations, sources, and a solution. BMC Bioinformatics. 2007; 8(1): 25. DOI: https://doi.org/10.1186/1471-2105-8-25
[93] Breiman L. Random forests. Machine Learning. 2001; 45:5–32. DOI: https://doi.org/10.1023/A:1010933404324
[94] Genuer R, Poggi JM, Tuleau-Malot C. Variable selection using random forests. Pattern Recognition Letters. 2010; 31: 2225–2236. DOI: https://doi.org/10.1016/j.patrec.2010.03.014
[95] Nembrini S, König IR, Wright MN. The revival of the Gini importance?. Bioinformatics. 2018; 34: 3711–3718. DOI: https://doi.org/10.1093/bioinformatics/bty373
[96] Shapley L, Roth A., The Shapley value: essays in honor of Lloyd S. Shapley, Cambridge University Press, 1988.
[97] Lundberg SM, Lee SI. A unified approach to interpreting model predictions, In Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017; 4768–4777. DOI: https://doi.org/10.48550/arXiv.1705.07874
[98] Matsui Y, Matsui T. Np-completeness for calculating power indices of weighted majority games. Theoretical

Computer Science. 2001; 263: 305–310. DOI: https://doi.org/10.1016/S0304-3975(00)00251-6
[99] Lundberg S, Nair B, Vavilala M, Horibe M, Eisses M, Adams T, et al. Explainable machine-learning predictions for the prevention of Hypoxaemia during surgery. Nature Biomedical Engineering. 2018; 2: 749-760. DOI: https://doi.org/10.1038/s41551-018-0304-0
[100] Lundberg SM, Erion GG, Lee SI. Consistent individualized feature attribution for tree ensembles. arXiv pre-print. 2018. DOI: https://doi.org/10.48550/arXiv.1802.03888
[101] Lundberg S, Erion G, Chen H, DeGrave A, Prutkin J, Nair B, et al. From local explanations to a global under-standing with explainable ai for trees. Nature Machine Intelligence. 2020; 2: 56-67. DOI: https://doi.org/10.1038/s42256-019-0138-9
[102] Fujimoto K, Kojadinovic I, Marichal JL. Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices. Games and Economic Behavior. 2006; 55: 72–99. DOI: https://doi.org/10.1016/j.geb.2005.03.002
[103] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research. 2005; 30 :79–82. Link: https://www.jstor.org/stable/24869236
[104] Chai T, Draxler RR, Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. Geoscientific Model Development. 2014; 7: 1247–1250. DOI: https://doi.org/10.5194/gmd-7-1247-2014
[105] Habeck CG, Brickman AM. A common statistical misunderstanding in psychology and neuroscience: Do we need normally distributed independent or dependent variables for linear regression to work?. bioRxiv. 2018. DOI: https://doi.org/10.1101/305946
[106] Ernst A, Albers C. Regression assumptions in clinical psychology research practice—a systematic review of common misconceptions. PeerJ. 2017; 5: e3323. DOI: https://doi.org/10.7717/peerj.3323
[107] Li X, Wong W, Lamoureux EL, Wong TY. Are Linear Regression Techniques Appropriate for Analysis When the Dependent (Outcome) Variable Is Not Normally Distributed?. Investigative Ophthalmology Visual Science. 2012; 53(6): 3082–3083. DOI: https://doi.org/10.1167/iovs.12-9967
[108] Lott JP. A note on normality. Journal of the American Academy of Dermatology. 2015; 72: e169–e170. DOI: https://doi.org/10.1016/j.jaad.2015.01.051
[109] Schmidt AF, Finan C. Linear regression and the normality assumption. Journal of Clinical Epidemiology. 2018; 98: 146–151. DOI: https://doi.org/10.1016/j.jclinepi.2017.12.006
[110] Schober P, Boer C, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. Anesthesia & Analgesia. 2018; 126: 1763–1768.
DOI: https://doi.org/10.1213/ANE.0000000000002864
[111] Mukaka M. Statistics corner: A guide to the appropriate use of correlation coefficient in medical research. Ma-lawi medical journal: The Journal of Medical Association of Malawi. 2012; 24(3): 69–71. Link: https://pub-med.ncbi.nlm.nih.gov/23638278/
[112] Strobl C, Boulesteix A-L, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. BMC bioinformatics. 2008; 9: 307. DOI: https://doi.org/10.1186/1471-2105-9-307
[113] Javadi D, Ssempebwa J, Isunju JB, Yevoo L, Amu A, Nabiwemba E, et al. Implementation research on sustain-able electrification of rural primary care facilities in Ghana and Uganda. Health Policy and Planning. 2020; 35: ii124–ii136. DOI: https://doi.org/10.1093/heapol/czaa077
[114] Moore N, Glandon D, Tripney J, Kozakiewicz T, Shisler S, Eyers J, et al. Effects of access to electricity inter-ventions on socio-economic outcomes in low-and middle-income countries. 3ie Systematic Review 45, 2020. Link: https://www.3ieimpact.org/evidence-hub/publications/systematic-reviews/effects-access-electricity-in-terventions-socio
[115] Duan Y, Shi J, Wang Z, Zhou S, Jin Y, Zheng Z-J. Disparities in COVID-19 vaccination among low-, middle-

 doi:10.20944/preprints202206.0115.v1

, and high-income countries: the mediating role of vaccination policy. Vaccines. 2021; 9(8): 905. DOI: https://doi.org/10.3390/vaccines9080905

[116] Roghani A, Panahi S. The global distribution of COVID-19 vaccine: The role of macro-socioeconomics measures. MedRxiv. 2021.
DOI: https://doi.org/10.1101/2021.02.09.21251436

[117] Kazemi M, Bragazzi NL, Kong JD. Assessing inequities in COVID-19 vaccine roll-out strategy programs: A cross-country study using a machine learning approach. Vaccines. 2022; 10(2): 194. DOI: https://doi.org/10.3390/vaccines10020194

[118] Ritchie H, Mathieu E, Roser M. Coronavirus pandemic (COVID-19), Our World in Data, 2020. Link: https://ourworldindata.org/coronavirus.

[119] Stoeckel F, Carter C, Lyons BA, Reifler J. Association of vaccine hesitancy and immunization coverage rates in the European Union. Vaccine. 2021; 39: 3935–3939. DOI: https://doi.org/10.1016/j.vaccine.2021.05.062

[120] Elgar K, Lopert L, Carey E, Wenzl M. Coronavirus (COVID-19) vaccines for developing countries: An equal shot at recovery, Tackling Coronavirus (COVID-19)- Browse OECD Contrıbutions. 2021; 1: 22. Link: https://searchworks.stanford.edu/view/13829305