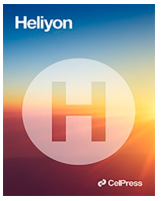
Heliyon 10 (2024) e27594

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Predicting immunogenicity of COVID-19 vaccines in hemodialysis patients with renal disease

Saeed Awad M Alqahtani [a], Waleed H. Mahallawi [b], Suliman Alomar [c,*]

[a] Physiology Department, Taibah University, Saudi Arabia
[b] Medical Laboratory Technology Department, College of Applied Medical Sciences, Taibah University, Madinah, Saudi Arabia
[c] Zoology Department, College of Science, King Saud University, P.O. Box: 2455, 11451, Riyadh, Saudi Arabia

ARTICLE INFO

Keywords:
Chronic kidney disease
Hemodialysis
COVID-19
Machine learning
Immunogenicity

ABSTRACT

Individuals who are diagnosed with chronic kidney disease, particularly those receiving maintenance hemodialysis treatment, face a greater likelihood of suffering from severe symptoms and fatality due to COVID-19. This study aimed to explore the optimal vaccination approach for these individuals. The study used data analysis tasks such as data preprocessing, cleaning, and exploration, and machine learning models including linear regression, random forest, XGBoost, gradient boosting, AdaBoost, decision trees, Lasso, and ridge regression were used to construct the predictive model. The study found that the Lasso model performed the best overall in predicting anti-S IgG antibodies levels in response to COVID-19 vaccines for people with kidney failure with MAE of 8.81, RMSE of 19.59, and R^2 value of 0.93. The adjusted R^2 value for the Lasso model was also 0.93, indicating that the model's ability to explain the variance in the data was not affected by the number of predictors in the model. The Random Forest model best predicted the duration of immunogenicity, with R^2 and adjusted R^2 values of 0.71 and 0.69, respectively. The ensemble model that includes all eight models, i.e., Ridge, Lasso, Linear Regression, Random Forest, AdaBoost, Gradient Boosting, XGBoost, and Decision Tree, has the best performance with the lowest MAE, the lowest RMSE, the highest R2, and the highest adjusted R2 values of 3.91, 5.00, 0.73, and 0.72, respectively. However, further research is required to validate these models and extend their application to different populations and vaccine types, as well as considering other factors that may affect immune response to COVID-19 vaccines. These findings can be helpful in improving vaccination strategies and promoting public health.

1. Introduction

The COVID-19 pandemic has had a significant impact on patients with chronic kidney disease (CKD), particularly those on maintenance hemodialysis. These patients are at a higher risk of severe illness and death from COVID-19 due to their underlying comorbidities and immunocompromised state [1]. The efficacy and immunogenicity of COVID-19 vaccines in these patients is an important area of research as it can help to determine the best vaccination strategies for this population.

Previous studies have shown that patients with CKD are at an increased risk of severe illness and death from COVID-19 [1,2]. Additionally, patients on maintenance hemodialysis are at a higher risk due to their underlying comorbidities and

* Corresponding author.
E-mail address: syalomar@ksu.edu.sa (S. Alomar).

https://doi.org/10.1016/j.heliyon.2024.e27594
Received 2 August 2023; Received in revised form 1 March 2024; Accepted 4 March 2024
Available online 7 March 2024

immunocompromised state [1]. Several studies have also reported that the immune response to COVID-19 vaccines is impaired in patients with CKD, including those on maintenance hemodialysis [3–5].

The vaccine for COVID-19 is considered as an important tool to reduce the serious effects of the pandemic, however, it remains unclear if the vaccines are as effective in patients with CKD [6]. Furthermore, the vaccines for COVID-19 have been shown to have variable efficacy and immunogenicity in CKD patients on maintenance hemodialysis [7], and this highlights the need for more research to determine the best vaccination strategies for this population.

Studies have also shown that the immune response to COVID-19 vaccines may be impaired in CKD patients on maintenance hemodialysis [3,4]. Kolb et al. found that dialysis patients had lower antibody titers and higher rates of seropositivity after vaccination with the Pfizer-BioNTech vaccine compared to the general population. Piotrowska et al. also reported that the immune response to the Pfizer-BioNTech vaccine was impaired in end-stage renal disease patients on hemodialysis, peritoneal dialysis, and kidney transplant recipients. Additionally, these patients had lower levels of local and systemic immunity after vaccination. Chen et al. conducted a meta-analysis of studies on the immunogenicity of COVID-19 vaccines in patients with end-stage kidney disease. The authors found that the immunogenicity rates were lower in these patients compared to the general population, and the results were consistent across the different vaccines studied. The authors concluded that the immunogenicity of COVID-19 vaccines may be impaired in patients with end-stage kidney disease.

In addition to the impaired immune response, patients on maintenance hemodialysis may also have other factors that can affect the efficacy of the vaccines such as being on immunosuppressive drugs and having anemia [6].

Further, studies have shown that the immunity provided by COVID-19 vaccines may not be long-lasting in CKD patients on maintenance hemodialysis [8]. Munro et al. conducted a blinded, multicenter, randomized, controlled trial in the UK, which investigated the safety and immunogenicity of seven COVID-19 vaccines as a third dose (booster) following two doses of BNT162b2 (Pfizer) or ChAdOx1 (AstraZeneca). The study included a subgroup of patients with chronic kidney disease, including those on hemodialysis. The results showed that the immunogenicity of the vaccines was lower in this subgroup, with a lower antibody response and a higher rate of seropositivity after the third dose.

In addition, some studies have suggested that patients on maintenance hemodialysis may require more frequent or higher doses of COVID-19 vaccines to achieve an adequate immune response [5,6]. However, more research is needed to confirm these findings and to determine the optimal dosing and vaccination schedule for CKD patients on maintenance hemodialysis.

Furthermore, studies have shown that the immunity provided by COVID-19 vaccines may not be long-lasting in CKD patients on maintenance hemodialysis [9]. Dekervel et al. conducted a study on the humoral response of BNT162b2 vaccine in patients on hemodialysis and found that the antibody levels were decreasing rapidly over time in this population. This highlights the need for more research to determine how long the immunity provided by COVID-19 vaccines lasts in CKD patients on maintenance hemodialysis, and whether or not these patients will require booster doses in the future.

Pfizer-BioNTech, Moderna, and AstraZeneca are three of the COVID-19 vaccines that have been authorized for emergency use worldwide. Studies have shown that the Pfizer-BioNTech and Moderna vaccines are the most commonly studied in CKD patients on maintenance hemodialysis. AstraZeneca vaccine, on the other hand, has been studied less in this population [3–5,7].

In terms of efficacy, studies have shown that the Pfizer-BioNTech and Moderna vaccines have a lower efficacy in CKD patients on maintenance hemodialysis compared to the general population [4,5]. AstraZeneca vaccine efficacy in CKD patients on maintenance hemodialysis has been studied less and more research is needed to determine the efficacy of this vaccine in this population.

In terms of immunogenicity, studies have reported that the antibody titers and rates of seropositivity are lower in CKD patients on maintenance hemodialysis who have received the Pfizer-BioNTech and Moderna vaccines [3,4]. The AstraZeneca vaccine immunogenicity in CKD patients on maintenance hemodialysis has been studied less, and more research is needed to determine the immunogenicity of this vaccine in this population.

It should be noted that, while some studies have shown a lower efficacy and immunogenicity of the Pfizer-BioNTech, Moderna, and AstraZeneca vaccines in CKD patients on maintenance hemodialysis, other studies have shown similar results to the general population [8]. Therefore, more research is needed to confirm the findings and to determine the best vaccination strategies for CKD patients on maintenance hemodialysis.

In summary, previous studies have shown that the immunity provided by COVID-19 vaccines may be impaired in CKD patients on maintenance hemodialysis, with lower antibody titers and higher rates of seropositivity. The Pfizer-BioNTech, Moderna, and AstraZeneca vaccines have been authorized for emergency use worldwide, and studies have shown that the Pfizer-BioNTech and Moderna vaccines are the most commonly studied in CKD patients on maintenance hemodialysis. The studies have reported that the efficacy and immunogenicity of the Pfizer-BioNTech, Moderna, and AstraZeneca vaccines may be lower in CKD patients on maintenance hemodialysis, but more research is needed to confirm the findings and to determine the best vaccination strategies for this population. Factors such as being on immunosuppressive drugs and having anemia may also affect the efficacy of the vaccines in these patients. Machine learning (ML) models can be used to predict immunogenicity in CKD patients on maintenance hemodialysis, which can help to identify patients who are at risk of waning immunity and who may need additional booster doses.

Machine learning is a type of artificial intelligence (AI) that utilizes algorithmic approaches to enable machines to solve problems without explicit computer programming, as stated by Ref. [10]. ML has become increasingly important in the field of medicine, as it can optimize the care of patients with chronic diseases, inform precision medicine approaches, and facilitate clinical trials. The number of articles utilizing ML in the medical field has been rapidly increasing, particularly in the areas of diagnostics and drug discovery. According to Accenture data, AI applications in healthcare can potentially save the United States healthcare sector $150 billion annually by 2026, as reported by Ref. [11]. This demonstrates the significant potential of ML in healthcare and explains the growing involvement of AI companies in the field of medicine, from diagnosis to treatment and drug development.

As far as we are aware, there have been no previous studies that have utilized machine learning to predict the immunogenicity of COVID-19 vaccines in patients with chronic renal disease undergoing hemodialysis. Therefore, the primary objective of this study is to develop a machine learning model that can accurately predict the immunogenicity of COVID-19 vaccines in this specific patient population. We used a variety of the most common machine learning models used in research, including linear regression, random forest, XGBoost, gradient boosting, AdaBoost, decision trees, Lasso, and ridge models.

2. Materials and methods

This study looked at a group of adult patients (n = 172) who were receiving hemodialysis treatment at a single center in Madinah, Saudi Arabia. The data collected for this study are presented in Table 1. In addition, blood samples were taken from the patients and the samples were frozen and later tested for the presence of SARS-CoV-2 anti-S IgG antibodies using a commercial ELISA kit. The study obtained ethical approval from the Research Ethics Committee of the College of Applied Medical Sciences, Taibah University. The ELISA test involved using 172-well plates and adding plasma and a sample diluent buffer to each well. After incubation and washing, horseradish peroxidase-labeled anti-human IgG antibody was added, followed by substrates A and B. The optical densities were then measured, and a cut-off value was calculated using a formula based on the mean absorbance of the negative controls.

Descriptive statistics were used to summarize the characteristics of the study sample. Mean and standard deviation were used for continuous variables, while frequency and percentage were used for categorical variables.

In this study, the Google Colaboratory platform was utilized to run codes using python libraries for data preprocessing, cleaning, and exploration of data analysis. The necessary libraries such as pandas, numpy, matplotlib, and seaborn were imported, and the Colab notebook was connected to Google Drive. The dataset was then read and stored as a DataFrame. Data cleaning tasks such as handling missing values, dealing with outliers, and removing duplicates were performed. Descriptive analysis was conducted to calculate statistics such as the mean, median, mode, standard deviation, etc. Feature engineering techniques were applied to improve the performance of the predictive model. The dataset was split into training and testing sets, and a predictive model was constructed using machine learning models: linear regression, random forest, XGBoost, gradient boosting, AdaBoost, decision trees, Lasso, and ridge regression. The model was trained on the training set and evaluated on the testing set using Coefficient of Determination (R^2), adjusted R^2, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Normalized Root Mean Squared Error (NRMSE) to measure its accuracy. Finally, the model was saved for future use once its performance was deemed satisfactory.

3. Results

3.1. Variable characteristics

The dataset consists of 172 individuals with a mean age of 50 ± 17 years, who received a mean of 2 ± 1 vaccine doses with mean number of weeks since the last vaccine dose is 18 ± 9. The mean anti-S IgG antibodies level (AU/mL) is 278 ± 70, with a mean number of months since kidney failure of 65 ± 60. Sixty-seven percent of the study sample was female (n = 118). According to this data, diabetes is the most frequent cause of renal failure (37%, n = 63). All patients included in this study were on hemodialysis three time per week. Most patient had no previous Covid19 infection, 137 observations which represent 80%. The majority (n = 160, 93%) of the patients have positive result for Anti IgG antibody. Pfizer were the most frequent as the first and second type of vaccine given to patients with 121 and 112 observations (70%, 65%) respectively. About half of patient did not get vaccine (n = 81, 47%). These results provide valuable information about the characteristics of the individuals in the dataset and can be used to identify patterns and relationships between variables and in building a predictive model by identifying important predictor variables.

Table 1
Data collected for this study.

Variable Name	Description
Age	Age of the patient
Sex	Gender of the patient
Cause_of_kidney_failure	Cause of kidney failure
Doses_of_vaccines	Number of vaccine doses received by the patient
Previous_Covid19_Infection	Whether the patient has previously had a Covid-19 infection
AU_mL	The level of antibodies in the patient's blood, measured in arbitrary units
Anti_IgG_antibody_results	Results of the Anti-IgG antibody test
_1st_type	Type of vaccine received for the first dose
_2nd_type	Type of vaccine received for the second dose
_3rd_type	Type of vaccine received for the third dose (if applicable)
Months_since_kidney_failure	Number of months since the patient experienced kidney failure
Time_since_last_vaccine_dose_weeks	Number of weeks since the patient received their last vaccine dose

3.2. Machine learning models

3.2.1. Predicting anti-S IgG antibodies level (AU/mL) level

The results provided are the evaluation metrics of eight different single machine learning models Table 2 that were trained to predict the target variable "AU_mL" using a dataset. Lasso model performed the best overall. The Lasso model has a MAE of 8.81, indicating that on average, the model's predictions deviate by approximately 8.81 units. The RMSE for the Lasso model is 19.59, which implies that the model's predictions deviate by approximately 19.59 units on average. The R^2 value for the Lasso model is 0.93, signifying that the model accounts for 93% of the variance in the data. The adjusted R^2 value for the Lasso model, which accounts for the number of predictors in the model, is 0.93, providing a more accurate representation of the model's ability to explain the variance in the data. The NRMSE for the Lasso model is 0.26, providing a means to evaluate the model's performance relative to other models with varying data scales.

An ensemble model is a combination of multiple individual models, which are combined to improve the overall performance of the model. Table 3 shows that the best ensemble model is the one that includes Lasso model only. It has the lowest MAE and RMSE values, and the highest R^2 and adjusted R^2 values. It also has the lowest NRMSE, which means that it performs better than the other models in comparison to the data scale.

3.2.2. Predicting duration of immunogenicity

Same models used to predict duration of immunogenicity in weeks (*Time_since_last_vaccine_dose_weeks)*. The results show that all the single models perform reasonably well, with R^2 values ranging from 0.54 to 0.71. The Random Forest model has the highest R^2 value of 0.71 which indicates that the model has a good fit to the data. The Decision Tree model has the lowest R^2 value of 0.54, indicating that it does not fit the data as well as the other models.

In terms of MAE, all the models have relatively low MAE values, with the Random Forest model having the lowest value of 4.18, indicating that it makes the least amount of error in its predictions. The Decision Tree model has the highest MAE value of 4.85, indicating that it makes more errors in its predictions compared to other models. The RMSE and NRMSE values are similar across all models, indicating that the models have similar levels of precision.

Overall, the model with the highest R^2 and Adj_ R^2 values is the Random Forest model. It has an R^2 value of 0.71 and an Adj_R2 value of 0.69, which indicates that it can explain 71% and 69% of the variance in the data, respectively. This means that it is the best model in predicting. Other metrics like MAE and RMSE are also relatively low for this model which is also a sign of good performance, Table 4.

As shown in Table 5, the ensemble model that has the best performance is the one that includes all the eight models; Ridge, Lasso, Linear Regression, Random Forest, AdaBoost, Gradient Boosting, XGBoost and Decision Tree. It has the lowest MAE, the lowest RMSE, the highest R^2 and the highest Adjusted R^2, 3.91, 5.00, 0.73, 0.72 0.52 respectively.

On the other hand, the individual models like Lasso, Linear Regression, and Random Forest Regressor are not performing as well as the ensemble models that include them. This suggests that the combination of multiple models can lead to better predictions, as it can leverage the strengths of each individual model.

4. Discussion

To the best of our knowledge, this is the first study to employ machine learning techniques for the prediction of anti-S IgG antibody levels and the duration of immunogenicity. The results of this study provide valuable insights into the characteristics of individuals with kidney failure who received COVID-19 vaccines and their immune response to the virus. The mean age of the study population was 50 years, and the majority of patients were female. The most common cause of kidney failure was diabetes. All patients were on hemodialysis three times per week, and the majority had no previous COVID-19 infection. Pfizer was the most frequently administered vaccine, with approximately half of patients not receiving any vaccine doses.

The study also evaluated eight different machine learning models for predicting anti-S IgG antibodies levels in response to COVID-19 vaccines. The Lasso model performed the best overall, with MAE of 8.81, RMSE of 19.59, and R^2 value of 0.93. The adjusted R^2 value for the Lasso model was also 0.93, indicating that the model's ability to explain the variance in the data was not affected by the number of predictors in the model. Additionally, the NRMSE value for the Lasso model was 0.26, suggesting that its performance was better than other models, even when data scales varied.

Table 2
Evaluation of the machine learning models.

Model	MAE	RMSE	R^2	Adjusted R^2	NRMSE
Lasso	8.81	19.59	0.93	0.93	0.26
Linear Regression	12.03	21.07	0.92	0.92	0.28
Ridge	13.12	22.14	0.91	0.91	0.30
Random Forest	12.17	29.82	0.84	0.83	0.40
XGBoost	12.57	33.00	0.81	0.79	0.44
Gradient Boosting	12.90	34.98	0.78	0.77	0.47
AdaBoost	14.18	38.60	0.74	0.72	0.51
Decision Tree	11.25	42.14	0.68	0.66	0.56

Table 3
Evaluation of different ensemble models.

Models	MAE	RMSE	R^2	Adjusted R^2	NRMSE
lasso	8.81	19.59	0.93	0.93	0.26
lasso, linear_regression	9.81	19.91	0.93	0.92	0.27
ridge, lasso	10.57	20.49	0.93	0.92	0.27
ridge, lasso, linear_regression	10.88	20.52	0.93	0.92	0.27
linear_regression	12.02	21.07	0.92	0.92	0.28
ridge, lasso, linear_regression, random_forest	11.28	21.62	0.92	0.91	0.29
lasso, linear_regression, random_forest	10.83	21.82	0.92	0.91	0.29
ridge, lasso, linear_regression, random_forest, ada_boost	10.93	21.89	0.91	0.91	0.29
ridge	13.12	22.14	0.91	0.91	0.30
lasso, linear_regression, random_forest, ada_boost	10.64	22.67	0.91	0.90	0.30
ridge, lasso, linear_regression, random_forest, ada_boost, gradient_boosting	10.90	23.31	0.90	0.90	0.31
linear_regression, random_forest	12.23	24.59	0.89	0.89	0.33
ridge, lasso, linear_regression, random_forest, ada_boost, gradient_boosting, xgboost	11.46	25.43	0.89	0.88	0.34
lasso, linear_regression, random_forest, ada_boost, gradient_boosting, xgboost	11.10	25.44	0.89	0.88	0.34
lasso, linear_regression, random_forest, ada_boost, gradient_boosting	11.52	26.06	0.88	0.87	0.35
ridge, lasso, linear_regression, random_forest, ada_boost, gradient_boosting, xgboost, dt	11.43	26.69	0.87	0.86	0.36
linear_regression, random_forest, ada_boost, gradient_boosting	12.21	27.81	0.86	0.85	0.37
linear_regression, random_forest, ada_boost	12.52	28.20	0.86	0.85	0.38
lasso, linear_regression, random_forest, ada_boost, gradient_boosting, xgboost, dt	11.97	28.90	0.85	0.84	0.39
linear_regression, random_forest, ada_boost, gradient_boosting, xgboost, dt	11.39	29.04	0.85	0.84	0.39
linear_regression, random_forest, ada_boost, gradient_boosting, xgboost	12.45	29.43	0.85	0.84	0.39
random_forest, ada_boost, gradient_boosting, xgboost, dt	11.18	29.47	0.85	0.84	0.39
random_forest, ada_boost	12.84	29.98	0.84	0.83	0.40
Decision_Tree	8.12	32.93	0.81	0.79	0.44
xgboost	12.57	33.00	0.81	0.79	0.44
random_forest, ada_boost, gradient_boosting, xgboost	12.89	33.35	0.80	0.79	0.44
random_forest	13.60	33.38	0.80	0.79	0.44
random_forest, ada_boost, gradient_boosting	13.07	33.49	0.80	0.79	0.45
gradient_boosting, xgboost	13.12	34.02	0.79	0.78	0.45
gradient_boosting	13.13	34.20	0.79	0.78	0.46
ada_boost, gradient_boosting, xgboost	12.49	34.22	0.79	0.78	0.46
ada_boost, gradient_boosting	12.39	34.88	0.78	0.77	0.46
ada_boost	14.29	37.14	0.75	0.74	0.50
gradient_boosting, xgboost, Decision_Tree	13.51	37.51	0.75	0.73	0.50
ada_boost, gradient_boosting, xgboost, Decision_Tree	13.60	37.58	0.75	0.73	0.50
xgboost, Decision_Tree	13.80	40.48	0.71	0.69	0.54

Table 4
Evaluation of Machine Learning models.

Model	MAE	RMSE	R^2	Adjusted R^2	NRMSE
Random Forest	4.18	5.25	0.71	0.69	0.54
Ridge	4.12	5.30	0.70	0.68	0.55
Linear Regression	4.12	5.39	0.69	0.67	0.56
Gradient Boosting	4.27	5.52	0.68	0.65	0.57
XGBoost	4.42	5.58	0.67	0.65	0.58
Lasso	4.42	5.72	0.65	0.63	0.59
AdaBoost	4.56	5.75	0.65	0.62	0.59
Decision Tree	4.85	6.55	0.54	0.51	0.68

Furthermore, the study assessed an ensemble model consisting of multiple individual models combined to improve the overall performance. The results indicated that the best ensemble model included only the Lasso model. It had the lowest MAE and RMSE values and the highest R^2 and adjusted R^2 values, as well as the lowest NRMSE value, indicating its superior performance relative to other models.

The results of this study show that predicting the duration of immunogenicity can be reasonably well achieved using various machine learning models. The R^2 values range from 0.54 to 0.71, with the Random Forest model having the highest value of 0.71, indicating a good fit to the data. The Decision Tree model has the lowest R^2 value of 0.54, which indicates that it does not fit the data as well as the other models.

In terms of mean absolute error (MAE), all the models have relatively low values, with the Random Forest model having the lowest value of 4.18. This indicates that it makes the least amount of error in its predictions. The Decision Tree model has the highest MAE value of 4.85, indicating that it makes more errors in its predictions compared to other models. The RMSE and NRMSE values are similar across all models, indicating that the models have similar levels of precision.

Overall, the Random Forest model is the best model in predicting the duration of immunogenicity, as it has the highest R^2 and

Table 5
Evaluation of ensemble models.

Ensemble_Models	MAE	RMSE	R^2	Adjusted R^2	NRMSE
Ridge, Lasso, LinearRegression, RandomForest, AdaBoost, GradientBoosting, XGBoost, DecisionTree	3.91	5.00	0.73	0.72	0.52
Ridge, Lasso, LinearRegression, RandomForest, AdaBoost, GradientBoosting, XGBoost	4.03	5.05	0.73	0.71	0.52
LinearRegression, RandomForest, AdaBoost, GradientBoosting, XGBoost, DecisionTree	3.98	5.09	0.72	0.71	0.53
Ridge, Lasso, LinearRegression, RandomForest, AdaBoost, GradientBoosting	4.06	5.10	0.72	0.70	0.53
Lasso, LinearRegression, RandomForest, AdaBoost, GradientBoosting, XGBoost	4.04	5.11	0.72	0.70	0.53
Lasso, LinearRegression, RandomForest, AdaBoost, GradientBoosting, XGBoost, DecisionTree	4.09	5.11	0.72	0.70	0.53
LinearRegression, RandomForest	4.01	5.13	0.72	0.70	0.53
LinearRegression, RandomForest, AdaBoost, GradientBoosting	4.05	5.13	0.72	0.70	0.53
RandomForest, AdaBoost, GradientBoosting, XGBoost, DecisionTree	3.99	5.15	0.72	0.70	0.53
Lasso, LinearRegression, RandomForest	4.09	5.15	0.72	0.70	0.53
Lasso, LinearRegression, RandomForest, AdaBoost, GradientBoosting	4.08	5.15	0.72	0.70	0.53
Ridge, Lasso, LinearRegression, RandomForest, AdaBoost	4.08	5.15	0.72	0.70	0.53
Ridge, Lasso, LinearRegression, RandomForest	4.10	5.15	0.72	0.70	0.53
LinearRegression, RandomForest, AdaBoost, GradientBoosting, XGBoost	4.08	5.18	0.71	0.69	0.53
Lasso, LinearRegression, RandomForest, AdaBoost	4.10	5.19	0.71	0.69	0.54
RandomForest	4.05	5.19	0.71	0.69	0.54
LinearRegression, RandomForest, AdaBoost	4.07	5.20	0.71	0.69	0.54
RandomForest, AdaBoost, GradientBoosting, XGBoost	4.12	5.27	0.70	0.68	0.54
RandomForest, AdaBoost, GradientBoosting	4.09	5.28	0.70	0.68	0.54
Ridge	4.12	5.30	0.70	0.68	0.55
Ridge, Lasso, LinearRegression	4.18	5.32	0.70	0.68	0.55
Ridge, Lasso	4.26	5.36	0.69	0.67	0.55
Lasso, LinearRegression	4.22	5.37	0.69	0.67	0.55
LinearRegression	4.12	5.39	0.69	0.67	0.56
RandomForest, AdaBoost	4.15	5.41	0.69	0.67	0.56
GradientBoosting	4.13	5.43	0.69	0.66	0.56
GradientBoosting, XGBoost	4.22	5.46	0.68	0.66	0.56
AdaBoost, GradientBoosting, XGBoost	4.24	5.47	0.68	0.66	0.56
AdaBoost, GradientBoosting, XGBoost, DecisionTree	4.29	5.51	0.68	0.65	0.57
AdaBoost, GradientBoosting	4.31	5.53	0.67	0.65	0.57
XGBoos	4.42	5.58	0.67	0.65	0.58
GradientBoosting, XGBoost, DecisionTree	4.36	5.62	0.66	0.64	0.58
AdaBoost	4.45	5.69	0.66	0.63	0.59
Lasso	4.42	5.72	0.65	0.63	0.59
XGBoost, DecisionTree	4.58	5.80	0.64	0.62	0.60
DecisionTree	5.02	6.66	0.53	0.50	0.69

adjusted R^2 values of 0.71 and 0.69, respectively. This means that it is able to explain 71% and 69% of the variance in the data, respectively. Other metrics such as MAE and RMSE are also relatively low for this model, which is a sign of good performance.

Interestingly, the ensemble model that includes all eight models, i.e., Ridge, Lasso, Linear Regression, Random Forest, AdaBoost, Gradient Boosting, XGBoost, and Decision Tree, has the best performance with the lowest MAE, the lowest RMSE, the highest R^2, and the highest adjusted R^2 values of 3.91, 5.00, 0.73, and 0.72, respectively. This suggests that combining multiple models can lead to better predictions, leveraging the strengths of each individual model.

This study has some limitations. Firstly, the study population only includes individuals with kidney failure who received COVID-19 vaccines and were on hemodialysis three times per week. This may limit the generalizability of the findings to other populations with kidney failure who are not on hemodialysis or who received different vaccine types. Secondly, while the study evaluated multiple machine learning models, it only considered a limited number of predictors for immune response and duration of immunogenicity. Other potential predictors such as comorbidities, vaccination history, and genetic factors may influence the immune response to COVID-19 vaccines and were not included in the analysis. Thirdly, the study did not assess the clinical implications of the predicted antibody levels or duration of immunogenicity. It is unclear whether the predicted outcomes correlate with actual protection against COVID-19 infection or disease progression. Lastly, the study relied on self-reported vaccination status, which may be subject to recall bias or inaccurate reporting. Additionally, the study did not assess the frequency or severity of adverse events following vaccination.

5. Conclusion

Overall, the study's findings suggest that the Lasso model is an effective tool for predicting anti-S IgG antibodies levels in response to COVID-19 vaccines in individuals with kidney failure. The results of this study suggest that machine learning models can predict the duration of immunogenicity, with the Random Forest model and the ensemble model being the most effective. The results also highlight the importance of evaluating multiple machines learning models and using ensemble models to improve overall performance. In addition, these findings can be helpful in improving vaccination strategies and promoting public health. However, future research is needed to validate these models and extend their application to different populations and vaccine types and consider additional predictors that may impact immune response to COVID-19 vaccines.

Funding

The authors would like to thank the Researchers Supporting Project Number (RSP2024R35), King Saud University, Riyadh, Saudi Arabia.

Data availability statements

The data that support the findings of this study are available on request from the corresponding author, Dr. Suliman Alomar.

CRediT authorship contribution statement

Saeed Awad M Alqahtani: Writing – original draft, Validation, Supervision, Software, Methodology, Investigation. **Waleed H. Mahallawi:** Writing – review & editing, Visualization, Resources, Methodology, Investigation. **Suliman Alomar:** Writing – review & editing, Validation, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D. Geetha, A. Kronbichler, M. Rutter, D. Bajpai, S. Menez, A. Weissenbacher, S. Anand, E. Lin, N. Carlson, S. Sozio, et al., Impact of the COVID-19 pandemic on the kidney community: Lessons learned and future directions, Nat. Rev. Nephrol. 18 (11) (2022) 724–737.
[2] K. El Karoui, A.S. De Vriese, COVID-19 in dialysis: clinical impact, immune response, prevention, and treatment, Kidney Int. 101 (6) (2022) 883–894.
[3] T. Kolb, S. Fischer, L. Müller, N. Lübke, J. Hillebrandt, M. Andrée, M. Schmitz, C. Schmidt, S. Küçükköylü, L. Koster, et al., Impaired immune response to SARS-CoV-2 vaccination in dialysis patients and in kidney transplant recipients, Kidney360 2 (6) (2021) 1491–1498.
[4] M. Piotrowska, M. Zieliński, L. Tylicki, B. Biedunkiewicz, A. Kubanek, Z. Ślizień, K. Polewska, P. Tylicki, M. Muchlado, J. Sakowska, et al., Local and systemic immunity are impaired in end-stage-renal-disease patients treated with hemodialysis, peritoneal dialysis and kidney transplant recipients Immunized with BNT162b2 pfizer-BioNTech SARS-CoV-2 vaccine, Front. Immunol. 13 (2022) 832924.
[5] J.J. Chen, T.H. Lee, Y.C. Tian, C.C. Lee, P.C. Fan, C.H. Chang, Immunogenicity rates after SARS-CoV-2 vaccination in people with end-stage kidney disease: a systematic review and meta-analysis, JAMA Netw. Open 4 (8) (2021) e2131749.
[6] N. Babel, C. Hugo, T.H. Westhoff, Vaccination in patients with kidney failure: lessons from COVID-19, Nat. Rev. Nephrol. 18 (2022) 708–723.
[7] R. Patel, M. Kaki, V.S. Potluri, P. Kahar, D. Khanna, A comprehensive review of SARS-CoV-2 vaccines: pfizer, Moderna & Johnson & Johnson, Hum. Vaccines Immunother. 18 (5) (2022) 1291–1298.
[8] A.P.S. Munro, L. Janani, V. Cornelius, P.K. Aley, G. Babbage, D. Baxter, M. Bula, K. Cathie, K. Chatterjee, K. Dodd, et al., Safety and immunogenicity of seven COVID-19 vaccines as a third dose (booster) following two doses of ChAdOx1 nCov-19 or BNT162b2 in the UK (COV-BOOST): a blinded, multicentre, randomised, controlled, phase 2 trial, Lancet 398 (10287) (2021) 2258–2276.
[9] M. Dekervel, N. Henry, M. Torreggiani, L.M. Pouteau, J.P. Imiela, C. Mellaza, A.S. Garnier, A. Dujardin, M. Asfar, A. Ducancelle, et al., Humoral response to a third injection of BNT162b2 vaccine in adults previously vaccinated with Pfizer-BioNTech COVID-19 mRNA vaccine, Lancet 397 (2021) 1335–1342.
[10] K.Y. Ngiam, W. Khor, Big data and machine learning algorithms for health-care delivery, Lancet Oncol. 20 (5) (2019) e262–e273.
[11] M.H. Stanfill, D.T. Marc, Health information management: implications of artificial intelligence on healthcare data and information management, Yearbook of medical informatics 28 (1) (2019) 56–64.