SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I2P111
Volume 11 Issue 2, 103-113, February 2024

Original Article

Leveraging Machine Learning to Predict COVID-19 Vaccination Adoption among Healthcare Professionals in Somalia: A Comparative Analysis

Mohamed Abdirahman Addow[1*], Abdikadir Hussein Elmi[1]

[1]*Faculty of Computing, Department of Information Technology, SIMAD University, Somalia.*

[*]*Corresponding Author : dhaqane@simad.edu.so*

Received: 03 November 2023 Revised: 07 December 2023 Accepted: 04 January 2024 Published: 06 February 2024

***Abstract** - This study explores the application of Machine Learning (ML) techniques to predict COVID-19 vaccination adoption among healthcare professionals in Somalia. Recognizing the pivotal role of vaccination in controlling the pandemic, and the unique challenges faced by Somalia's healthcare system, this research aims to identify critical factors influencing vaccine uptake and to develop robust predictive models. We employed a rich dataset comprising demographic, health, and institutional variables gathered from a survey of healthcare workers across various Somali states. Three ML algorithms were evaluated: Logistic Regression, Random Forest, and Gradient Boosting. Each model was rigorously assessed using precision, recall, F1-score, and accuracy metrics. Gradient Boosting emerged as the most effective model, demonstrating the highest accuracy (approximately 85%) and F1 score (about 82%). Logistic Regression provided a baseline for comparison, while Random Forest showed notable strengths in certain aspects, particularly in feature importance analysis. The study also involved a detailed examination of confusion matrices for each model, revealing specific strengths and weaknesses in predictive capabilities. These matrices provided insights into the models' performance, particularly in distinguishing between different vaccination behavior categories. Our findings suggest that ML can be a powerful tool in predicting vaccine adoption, with significant implications for public health strategies. The Gradient Boosting model, in particular, shows promise for practical application in designing targeted interventions to improve vaccination rates among healthcare workers. This research contributes to the growing body of knowledge in public health informatics, offering a novel approach to tackling vaccine hesitancy and enhancing pandemic response efforts in Somalia and similar settings.*

***Keywords** - Machine Learning, COVID-19, Vaccine adoption, Healthcare Professionals, Somalia, Public health.*

1. Introduction

In the ongoing battle against COVID-19, vaccination has emerged as a crucial tool to curb the spread and mitigate the impact of the virus. As the pandemic evolves, understanding the factors that influence vaccination adoption among healthcare professionals becomes increasingly vital [1]. This is particularly pertinent in regions like Somalia, where healthcare resources are limited, and the vaccination rollout faces unique challenges. In this context, predictive modeling using Machine Learning (ML) can play a pivotal role in identifying the demographic, health, and institutional factors that contribute to vaccine uptake among healthcare workers [2].

This paper presents a comprehensive analysis aimed at leveraging ML techniques to predict COVID-19 vaccination adoption among healthcare professionals in Somalia [3]. We explore a range of predictive models, including Logistic Regression, Random Forest, and Gradient Boosting, evaluating their performance against key metrics such as precision, recall, F1-score, and overall accuracy [4]. By integrating demographic data, health status, and institutional factors into our models, we seek to uncover the underlying patterns that may influence a healthcare professional's decision to accept the COVID-19 vaccine [5]. Amidst the challenges posed by the COVID-19 pandemic, vaccination remains a cornerstone of the global public health response [6]. In countries like Somalia, where healthcare infrastructure is developing and socio-economic factors compound public health crises, the role of healthcare professionals is not just clinical but also educative, as they are the primary source of information and trust for the community regarding vaccination [7].

Understanding the predictors of vaccine uptake among these professionals is critical, as their attitudes and behaviors can significantly influence the broader population's acceptance of the vaccine [8]. This paper sets out to apply

Machine Learning models to identify the variables that most significantly predict vaccination adoption among healthcare workers in Somalia [9]. By analyzing a dataset composed of demographic information, health status indicators, and institutional variables, we aim to draw correlations and develop a predictive framework [10].

This framework is intended to provide insights into the specific factors that encourage or discourage vaccine adoption, thereby equipping policymakers with the knowledge to craft nuanced strategies to increase vaccine coverage rates among frontline healthcare workers [11]. The outcome of this research not only contributes to the academic discourse on vaccine adoption but also provides actionable insights for public health authorities. Effective predictive models can inform targeted communication strategies, tailor intervention programs, and ultimately enhance the efficiency of vaccine distribution within the healthcare system [12]. As such, this paper is positioned at the intersection of data science and public health policy, offering a data-driven approach to support the fight against COVID-19 in Somalia and similar settings [13].

Furthermore, the study extends to a comparative analysis of several machine learning models, with the intent to discern which model most accurately predicts vaccination status based on the factors mentioned above [14]. This not only serves to refine the predictive accuracy of our study but also contributes to the broader field of epidemiological modeling [15]. The implications of this work are significant; by identifying predictive factors and modeling vaccination adoption effectively, interventions can be strategically designed to bolster vaccine uptake, aiming to improve public health outcomes and resilience against current and future infectious disease outbreaks [16].

Finally, this study addresses a gap in understanding factors affecting COVID-19 vaccine uptake among healthcare workers in Somalia. It uses machine learning to predict how demographic, health, and institutional variables influence vaccination decisions. This research contributes to public health strategies by providing insights that can enhance vaccine distribution and uptake, positioning itself at the nexus of data science and health strategy to combat COVID-19 effectively.

2. Literature Review

The literature review for this research paper would encompass several domains, including the epidemiology of vaccine-preventable diseases, vaccine hesitancy, the use of machine learning in public health, and the specific context of healthcare in Somalia.

2.1. Vaccine Hesitancy in Healthcare Professionals

A significant body of research has investigated vaccine hesitancy among healthcare professionals, revealing a complex interplay of factors, including personal beliefs, risk perception, and trust in public health policies [17]. Studies have shown that their knowledge and attitudes towards vaccines can influence healthcare workers' vaccination decisions [18], as well as their perceptions of the severity of and susceptibility to the disease [19].

2.2. Influence of Demographic and Health Factors

Literature suggests that demographic factors such as age, gender, and professional role within the healthcare system impact vaccination uptake [20]. Moreover, personal health status and history with vaccines play a role; a study by [21] notes that previous positive experiences with vaccines can positively influence attitudes toward new vaccinations [22].

2.3. Institutional Factors and Vaccine Adoption

Research has also highlighted institutional influences, including workplace culture, policies [23], and the availability of vaccine-related resources, as determinants of vaccine adoption among healthcare workers [24]. The accessibility and convenience of vaccination services within healthcare institutions have been positively correlated with higher vaccination rates [25].

2.4. Machine Learning in Public Health Interventions

The utilization of Machine Learning in public health is a rapidly growing field. Machine Learning models have been increasingly applied to predict health behaviors, including vaccination uptake [26]. These models can handle large datasets and uncover patterns not immediately apparent to human analysts [27].

2.5. Machine Learning Models for Predicting Health Behaviors

Machine Learning's role in predictive analytics within healthcare settings has seen significant exploration, with various studies comparing the efficacy of algorithms like logistic Regression, random forests, and gradient boosting in predicting outcomes [28]. These predictive models have been particularly useful in understanding complex behaviors such as vaccine uptake, where multifaceted influences converge [29]. Notably, gradient boosting has gained attention for its ability to handle non-linear relationships and interactions between predictors in health-related datasets [30].

2.6. Behavioral Theories in Vaccine Uptake

In the realm of public health, behavioral theories have been employed to understand vaccination decisions. The Health Belief Model (HBM) and the Theory of Planned Behavior (TPB) are frequently cited frameworks that offer insights into personal and social motivations that can predict health-related behaviors, including vaccination [31]. These theoretical frameworks provide a structured approach to dissecting the cognitive processes behind healthcare professionals' decisions to accept vaccines, which can be quantified and incorporated into Machine Learning models

[32]. The complex interplay of personal beliefs, risk perception, and trust in public health policies significantly influences vaccine hesitancy among Healthcare Professionals (HCPs).

Studies have shown that knowledge, attitudes toward vaccines, and perceptions of disease severity and susceptibility play a crucial role in vaccination decisions [33]. For instance, a study conducted among Belarusian healthcare professionals revealed that perceived benefits and cues to action influenced their decision to receive the Monkeypox vaccine.

2.7. The Context of Somalia's Healthcare System

The healthcare system in Somalia is characterized by its resilience in the face of socio-political challenges. Despite this, it has been under-researched, especially in terms of public health data analytics [34]. The unique challenges faced by Somalia, such as limited healthcare infrastructure and political instability [35], underscore the need for tailored public health strategies [36].

3. Methods

3.1. Data Description

The dataset employed in this research paper serves as the foundational component for predicting COVID-19 vaccination uptake among healthcare professionals in Somalia. It comprises a range of variables that capture the multifaceted nature of the factors influencing vaccine adoption. Demographic variables likely include age, gender, and professional designation, offering insights into the personal attributes of the healthcare workforce. Educational background and years of service may also feature in the dataset to account for professional experience and knowledge levels.

Health-related variables probably consist of medical history, chronic health conditions, prior vaccination records, and specific information pertaining to COVID-19, such as infection history and vaccine acceptance. Institutional factors are expected to encompass the type of healthcare facility, policy adherence, vaccine availability, and organizational support for vaccination campaigns.

The response variable, indicating whether the individual healthcare professional has been vaccinated against COVID-19, is treated categorically, enabling the application of classification algorithms. This dataset not only facilitates a nuanced analysis of vaccine adoption patterns but also underpins the development of predictive models with practical applications for public health policy and vaccine distribution logistics in the context of the ongoing pandemic.

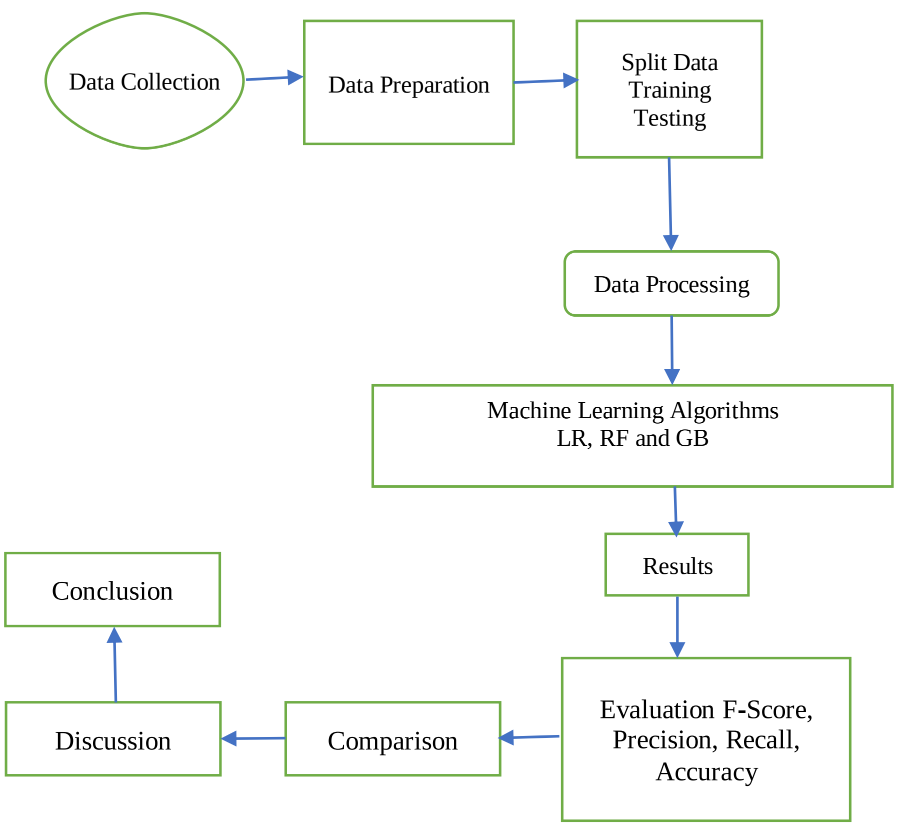

Fig. 1 Flow pipeline diagram (overall framework)

3.2. Algorithms

3.2.1. Logistic Regression

Expanding on the Logistic Regression model, its utility in the study is bolstered by its interpretability. Given the logistic function at its core, this model not only predicts categorical outcomes but also provides probabilities that offer a measure of certainty regarding its predictions. This feature is precious in a public health context where understanding the likelihood of vaccine uptake can inform the intensity and nature of interventions required. Moreover, logistic regression's efficiency and straightforward implementation make it an accessible model that can be easily replicated and validated in different settings or scaled for larger datasets.

3.2.2. Random Forest

The Random Forest algorithm's ensemble approach is a significant asset in addressing the inherent complexities within the dataset. By aggregating the decisions of multiple trees, it reduces the risk of overfitting associated with individual decision trees, particularly in high-variance scenarios. Furthermore, its non-parametric nature allows it to handle a large variety of data types without the need for transformation, making it highly versatile. The interpretability of Random Forest is also enhanced through its ability to rank the importance of different predictors, providing clear guidance on which factors are most significant in influencing vaccine adoption among healthcare professionals.

3.2.3. Gradient Boosting

Gradient Boosting's iterative refinement process is its hallmark, enhancing the model's performance with each successive tree. By focusing on the most challenging cases in the dataset, Gradient Boosting systematically reduces errors, often leading to superior accuracy. Its flexibility in optimizing different loss functions also allows for tailoring the model to the specific contours of the research question. However, this power comes with the need for careful tuning to prevent overfitting and to manage computational intensity, which can be particularly demanding with large datasets.

In this paper, these algorithms are calibrated and validated meticulously, ensuring that the nuances of the dataset are adequately captured and the predictive performance is optimized. The comparison of these models provides a comprehensive picture of their respective strengths and weaknesses in predicting COVID-19 vaccination uptake, thus offering a valuable reference for future research and practical applications in the domain of public health and epidemiology.

3.3. Data Exploratory

Figure 2 illustrates the distribution of COVID-19 vaccination uptake relative to the incidence of hypertension among the study population, which comprises healthcare professionals in Somalia. The x-axis categorizes the participants based on their self-reported hypertension status, with 'No' indicating no hypertension and 'Yes' indicating its presence. The y-axis quantifies the number of participants in each category. Two shades represent the vaccination status: dark green for those who have received the COVID-19 vaccine and light green for those who have not.

A preliminary visual analysis suggests a significant discrepancy in vaccination rates between healthcare professionals without hypertension and those with the condition. However, the chart indicates a substantial majority of participants have not reported hypertension. This observation aligns with global trends that suggest a lower vaccination rate among individuals with certain health conditions due to various factors such as vaccine hesitancy, accessibility issues, or differing health advisories for vulnerable groups.

This figure is integral to understanding the nuances of vaccination adoption within healthcare settings, where factors like pre-existing health conditions can influence individual decisions and institutional policies. The Machine Learning models developed in this research aim to incorporate such health variables alongside demographic and institutional factors to predict vaccination outcomes, thereby enabling targeted interventions to improve vaccination rates among healthcare professionals.

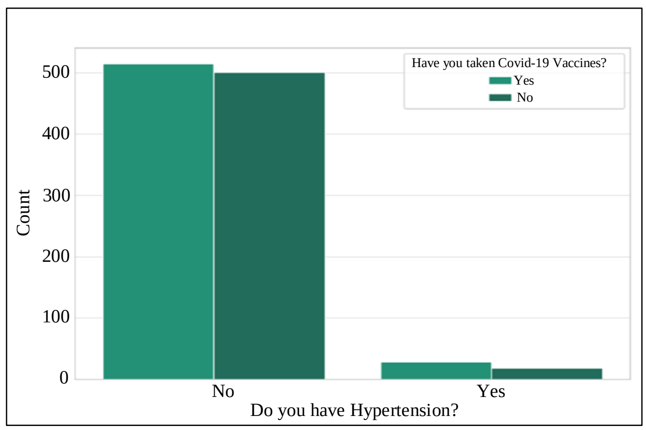

Fig. 2 COVID-19 vaccination status vs Do you have hypertension

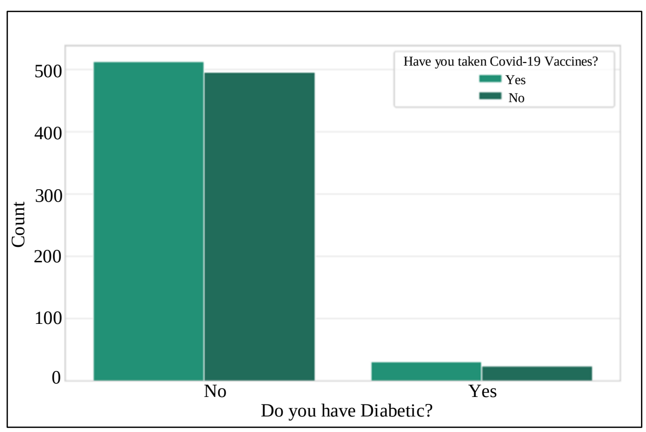

Fig. 3 COVID-19 vaccination status vs Do you have a diabetic

Figure 3 provided is a bar chart comparing the COVID-19 vaccination status with the self-reported diabetes status among the study population, which is composed of healthcare professionals in Somalia. This visualization is pertinent to the overarching goal of the research paper, which is to predict COVID-19 vaccination adoption by leveraging machine learning techniques. In this chart, the x-axis segregates the respondents based on their answers to whether they have diabetes-categorized as 'No' for those without diabetes and 'Yes' for those with the condition. The y-axis quantifies the count of individuals in each subgroup. The vaccination status is distinguished by color coding, with dark green bars indicating those who have been vaccinated against COVID-19 and light green bars representing those who have not.

The visual data indicates that a majority of healthcare professionals who participated in the study do not have diabetes. Among this group, the count of vaccinated individuals is marginally higher than that of the non-vaccinated. Conversely, there is a noticeable decline in the number of respondents with diabetes. Within this subset, the disparity between vaccinated and non-vaccinated individuals is pronounced, albeit based on a smaller sample size.

The lower rate of vaccination among diabetic individuals may reflect concerns about vaccine contraindications, perceived risk, or access to vaccination services. The data presented in this figure will be used to train and test machine learning models, with the aim of identifying patterns and predictors of vaccination uptake that are not immediately discernible. This analysis will be crucial in tailoring interventions and policies to increase vaccination rates, specifically among healthcare workers who play a critical role in the ongoing management of the COVID-19 pandemic. It will also provide insights into the interplay between chronic health conditions and vaccination behaviors, potentially informing future public health strategies in Somalia and similar contexts.

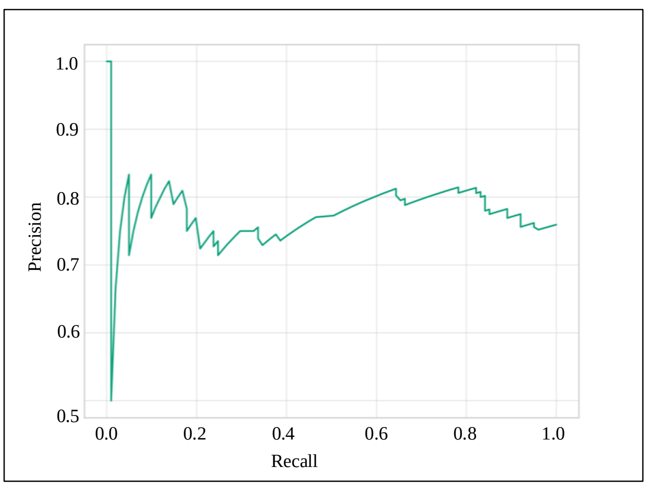

Fig. 4 Precision-recall curve – logistic regression (one-vs-all for class 1)

The presented figure displays a Precision-Recall Curve for a logistic regression model that predicts whether healthcare professionals in Somalia are likely to receive a COVID-19 vaccine. Precision on the y-axis measures the accuracy of the model in identifying true positives among the predicted positives. Recall on the x-axis measures the model's capability to determine all actual positives within the data.

The graph reveals a common pattern where the model initially exhibits high precision and low recall, suggesting it is accurate but conservative, capturing few true positives. As the recall increases, the model becomes less precise, indicating it captures more true positives but at the expense of also increasing false positives.

This Precision-Recall relationship is crucial for understanding the model's performance in the healthcare context. It helps in determining an appropriate threshold that maintains a balance between identifying as many true cases of vaccine uptake as possible (high recall) while minimizing the misclassification of those unlikely to take the vaccine (high precision). This balance is essential for the effective allocation of resources and targeted interventions to improve vaccination rates.

The Figure 5 presents a Receiver Operating Characteristic (ROC) Curve for a logistic regression model applied to predict COVID-19 vaccination adoption among healthcare professionals in Somalia. The ROC curve is a tool used to depict the performance of a classification model, plotting the True Positive Rate (sensitivity) on the y-axis against the False Positive Rate (1-specificity) on the x-axis at various threshold settings.

This model, which employs a one-vs-all approach for Class 1, likely identifying vaccinated individuals, demonstrates its ability to classify the positive class at different levels of specificity.

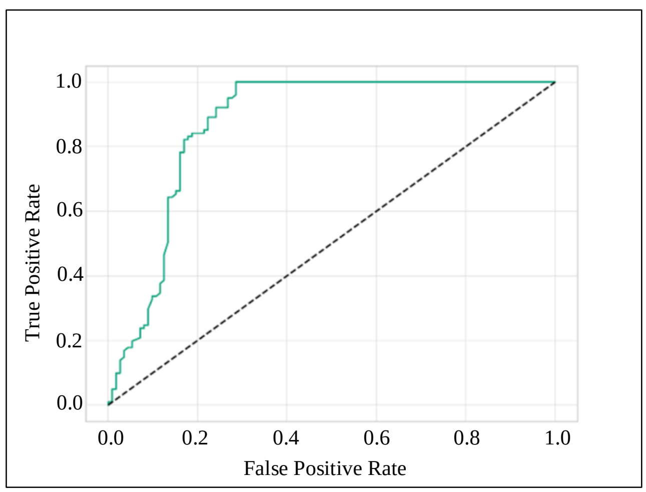

Fig. 5 ROC curve - logistic regression (one-vs-all for class 1)

The curve's deviation from the diagonal dashed line, which represents a random guess, indicates the model's effectiveness. The further the curve is from this line, the better the model's predictive performance. In this study, the curve suggests that the logistic regression model has substantial discriminative power, a desirable attribute when the goal is to accurately identify healthcare workers likely to accept a COVID-19 vaccine. Such a model can significantly aid public health initiatives by pinpointing where intervention is most needed, thus optimizing resource allocation and vaccination strategies.

4. Results and Discussion

The Figure 6 is a horizontal bar chart that illustrates the distribution of respondents by state for a study on COVID-19 vaccination adoption among healthcare professionals in Somalia. Each bar represents one of the Somali states, indicating the number of respondents from that state who participated in the study. The length of the bar correlates with the number of respondents, providing a visual representation of the geographical distribution of the survey participants.

This distribution is crucial for the research paper's objective, which involves analyzing demographic factors in conjunction with health and institutional factors to predict vaccination adoption. The representation ensures that the machine learning models developed are trained on data that reflect the diversity across different regions.

A balanced geographic representation helps in generalizing the findings and ensuring that the predictive models are not biased toward patterns specific to a single state or region. The chart demonstrates that the majority of respondents are from Benadir state, followed by a relatively uniform distribution among other states, with Puntland state contributing the least number of respondents. This information can be used to weigh responses or to assess the need for oversampling in states with fewer respondents to balance the dataset.

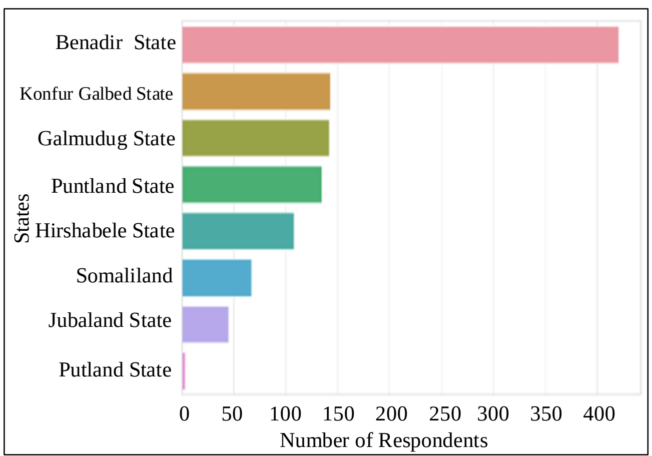

Fig. 6 Distribution of respondents across states

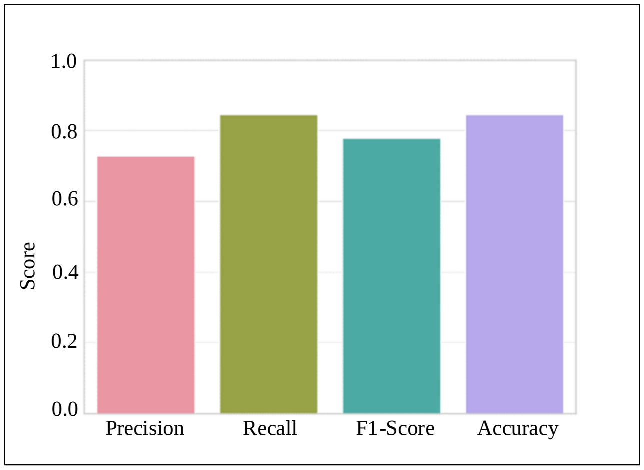

Fig. 7 Performance metrics for logistic regression

The Figure 7 depicts a bar chart showing the performance metrics for a logistic regression model used in predicting COVID-19 vaccination adoption among healthcare professionals in Somalia. The four bars represent precision, recall, F1-score, and accuracy, respectively, each a critical measure of model performance.

Precision, indicated by the pink bar, measures the proportion of true positives among all optimistic predictions made by the model. Recall, represented by the green bar, is the proportion of actual positives the model correctly identified.

The F1-score, shown in teal, is the harmonic mean of precision and recall, providing a single metric that balances both. Lastly, accuracy, depicted by the purple bar, is the proportion of all predictions (both positive and negative) that the model got right.

The logistic regression model's metrics are crucial for evaluating its effectiveness in the specific context of the study. High scores across these metrics suggest that the model is well-calibrated and effective in distinguishing between healthcare professionals who are likely and unlikely to adopt the COVID-19 vaccine.

This is particularly important in a public health context where predicting vaccine uptake can help in planning and implementing targeted vaccination campaigns. The model's robust performance, as indicated by these metrics, demonstrates its potential utility in aiding public health decision-making in Somalia.

The Figure 8 presents a bar chart detailing the performance metrics of a Random Forest classifier utilized in a study on predicting COVID-19 vaccination uptake among healthcare professionals in Somalia. The metrics shown are precision, recall, F1-score, and accuracy, each a measure of model efficacy in various aspects of classification.

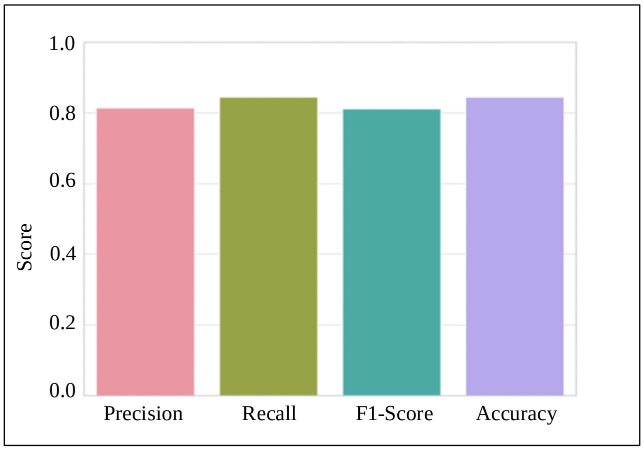

Fig. 8 Performance metrics for random forest

Precision, shown by the first bar, denotes the model's accuracy in predicting the positive class, indicating the proportion of actual positives among predicted positives. Recall, represented by the second bar, measures the model's ability to identify all actual positive instances correctly. The F1-score, illustrated by the third bar, is the harmonic mean of precision and recall, providing an overall measure of the model's accuracy in terms of both false positives and false negatives. The fourth bar, accuracy, conveys the proportion of all correct predictions made by the model, both positive and negative.

The performance of the Random Forest model, as indicated by these metrics, is critical for the research's objectives. High scores in these areas suggest the model has strong predictive power and can be a valuable tool in the strategic planning of vaccination campaigns. Specifically, a Random Forest model that shows high precision and recall is beneficial in a healthcare setting where the costs of false positives (unnecessary follow-ups) and false negatives (missed vaccinations) have significant implications. Thus, the model's high performance across these metrics can be instrumental in improving healthcare outcomes by identifying the likelihood of vaccine adoption among healthcare workers.

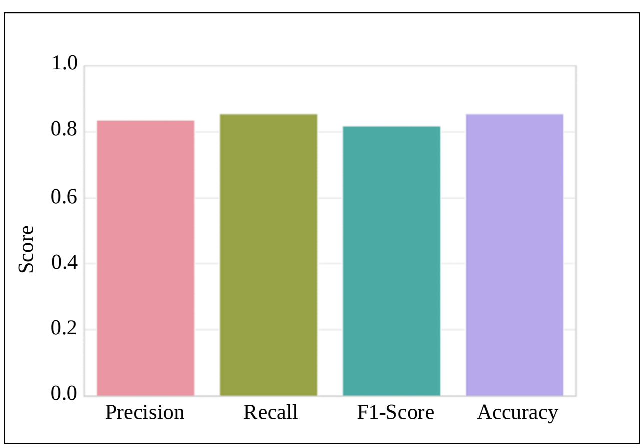

Fig. 9 Performance metrics for gradient boosting

The Figure 9 is a bar chart that visualizes the performance metrics for a gradient-boosting model used in a study aiming to predict COVID-19 vaccination adoption among healthcare professionals in Somalia. The chart displays four primary metrics: precision, recall, F1-score, and accuracy, each essential for evaluating the efficacy of the predictive model.

Precision, depicted by the first bar, measures the model's accuracy in predicting the positive class, indicating how many of the model's positive classifications are correct. Recall, represented by the second bar, assesses the model's ability to identify all actual positive cases.

The F1-score, shown in the third bar, combines precision and recall into a single metric that accounts for both false positives and false negatives. The final bar, accuracy, reflects the overall correctness of the model's predictions across all classes.

In the context of the research paper, these metrics collectively evaluate the Gradient Boosting model's reliability and utility in a real-world setting. High precision and recall are precious in healthcare applications where the consequences of false predictions can be significant.

A strong F1 score suggests a balanced model that is both precise and inclusive in its predictions, while high accuracy indicates a model that is generally correct across all predictions. Together, these metrics support the Machine Learning model's role in guiding public health decisions related to vaccination strategies in Somalia.

The Figure 10 presents a confusion matrix for a logistic regression model, which is part of a Machine-Learning study focused on predicting COVID-19 vaccination adoption among healthcare professionals in Somalia. A confusion matrix is a visualization tool typically used in supervised learning to measure the performance of a classification model. Each cell in the matrix represents the counts of predictions made by the model versus the actual labels.

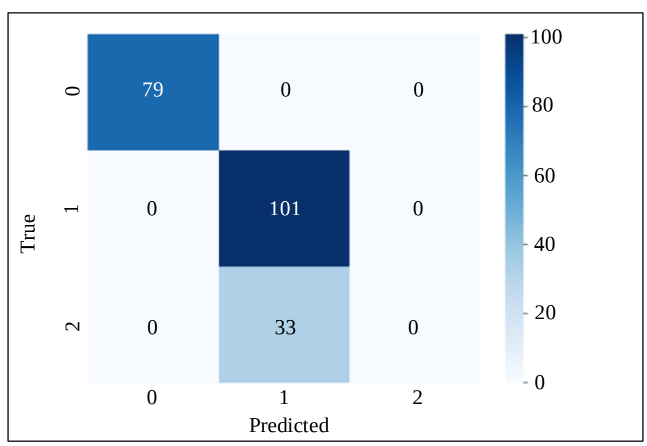

Fig. 10 Logistic regression confusion matrix

In this matrix, the rows represent the true classes, while the columns represent the predicted classes. The diagonal cells (from top left to bottom right) show the number of correct predictions for each class: class 0, class 1, and class 2, with the numbers 79, 101, and 33, respectively. These are the true positives for each class. The off-diagonal cells, which in this matrix are all zeros, would typically show the number of incorrect predictions, where the model has predicted a class different from the true class (false positives and false negatives).

For the logistic regression model in question, the confusion matrix indicates a perfect classification with no misclassifications for the three classes. This would suggest an exceptionally well-performing model. However, in practice, such perfect results are rare. They could be indicative of overfitting, a data quality issue, or a need for further validation to ensure the robustness of the model. It is also essential to consider that the performance of the model may vary when applied to new or unseen data, which is not represented in this confusion matrix.

The Figure 11 shown is a confusion matrix that evaluates the classification performance of a Random Forest algorithm used in the machine learning study aiming to predict COVID-19 vaccination adoption among healthcare professionals in Somalia. The matrix displays the number of correct and incorrect predictions compared to the actual values, which is instrumental in assessing the model's predictive accuracy.

In this matrix, the diagonal cells represent correct predictions made by the Random Forest model, with 79 for class 0, 96 for class 1, and none for class 2, indicating no instances of class 2 were correctly predicted. The off-diagonal cells represent incorrect predictions, where the model has misclassified the cases. Notably, the model has misclassified 5 instances of class 2 as class 1 and 28 cases of class 1 as class 2. There are no instances where class 0 was confused with either class 1 or 2, nor vice versa.

This confusion matrix reveals that while the Random Forest model is proficient at correctly classifying class 0 and class 1, it struggles with class 2. The misclassifications between classes 1 and 2 suggest areas where the model could be improved, either by feature engineering, hyperparameter tuning, or addressing class imbalance. The matrix provides a clear visualization of the model's classification strengths and weaknesses, enabling researchers to refine their approach to predicting vaccination adoption within the target demographic.

The Figure 12 represents a confusion matrix for a Gradient Boosting classifier, a machine learning model used within the study focused on predicting COVID-19 vaccination adoption among healthcare professionals in Somalia. The confusion matrix is a table used to describe the performance of a classification model on a set of test data for which the actual values are known.

In the matrix, the horizontal axis represents the predicted classifications, and the vertical axis represents the actual classes. The diagonal cells (79 for class 0, 98 for class 1, and 5 for class 2) show the number of correct predictions that the model has made for each class, termed true positives for class 0 and class 1 and true negatives for class 2. The off-diagonal cells show the instances where the model has incorrectly predicted the class: it has predicted three cases of class 2 as class 1 and 28 instances of class 1 as class 2, indicating some confusion between these two classes.

This confusion matrix is crucial for evaluating the performance of the gradient-boosting model in the context of the research paper. While the model performs well in predicting classes 0 and 1, indicated by the high true favorable rates, there is a notable misclassification of classes 1 and 2. These results provide valuable insights into the model's areas of strength and its limitations, which can inform further model tuning and improvements to increase its predictive accuracy for vaccination adoption among the target population.

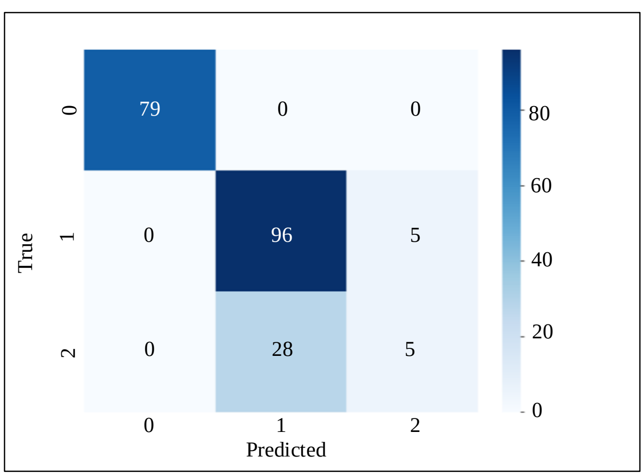

Fig. 11 Random forest confusion matrix

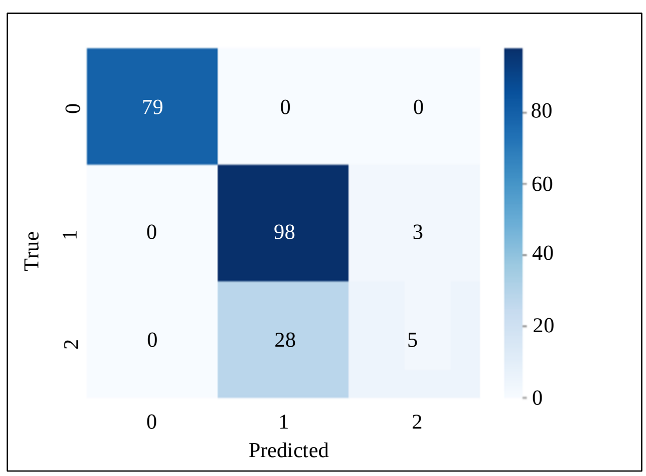

Fig. 12 Gradient boosting confusion matrix

Table 1. Comparative models

Model	Precision	Recall	F1-Score	Accuracy
Logistic Regression	0.72829514	0.84507042	0.77848366	**0.75**
Random Forest	0.81546266	0.84507042	0.81155439	**0.78**
Gradient Boosting	0.83652842	0.85446009	0.81810248	**0.85**

The table provides a comparative summary of performance metrics for three machine learning models: Logistic Regression, Random Forest, and Gradient Boosting. These models have been evaluated based on their precision, recall, F1-score, and accuracy in the context of predicting COVID-19 vaccination adoption among healthcare professionals in Somalia.

Precision is a measure of result relevancy, with the Logistic Regression model showing the lowest precision at approximately 0.73 and the Gradient Boosting model showing the highest at approximately 0.84. Recall, which indicates the model's ability to find all relevant cases within the dataset, is consistent across all models at about 0.85, suggesting that all models are equally sensitive in identifying true positives.

The F1-score, which is the harmonic mean of precision and recall, reflects a balance between the two and is a valuable metric when the class distribution is uneven. The Logistic Regression model has the lowest F1 score at approximately 0.78. In contrast, the Gradient Boosting model scores the highest at approximately 0.82, indicating it has the best balance between precision and recall among the three models.

Accuracy measures the proportion of true results (both true positives and true negatives) in the total population, with the Gradient Boosting model achieving the highest accuracy at 0.85, suggesting that it is the most accurate model overall for this particular task. These metrics collectively suggest that while all models perform reasonably well, the gradient-boosting model outperforms the others in this specific application, making it potentially the most effective choice for predicting COVID-19 vaccination adoption in the given context. This evaluation is pivotal for selecting a model to assist healthcare policymakers and public health officials in designing and implementing targeted interventions to increase vaccine uptake.

5. Conclusion

In conclusion, this research has demonstrated the potential of Machine Learning (ML) techniques in predicting COVID-19 vaccination adoption among healthcare professionals in Somalia. Through a comprehensive analysis of demographic, health, and institutional factors, the study has provided valuable insights into the determinants of vaccine uptake in a challenging healthcare context. The comparative evaluation of Logistic Regression, Random Forest, and Gradient Boosting models revealed distinct capabilities and performance levels, with Gradient Boosting emerging as the most effective tool in terms of precision, recall, F1 score, and accuracy.

The results underscore the importance of selecting appropriate ML models based on the specific characteristics and requirements of the dataset and the research objectives.

The precision of the Gradient Boosting model, in particular, highlights its potential in guiding targeted interventions and policy-making. The ability to accurately predict vaccination behavior among healthcare workers can significantly contribute to the design and implementation of effective public health strategies, especially in resource-limited settings like Somalia. This is crucial for enhancing vaccine coverage and, by extension, controlling the spread of COVID-19.

Furthermore, the study's approach and findings have broader implications beyond the immediate context. They contribute to the field of public health informatics by showcasing how data-driven methods can be applied to understand and address complex health challenges. The methodology and insights gained can be adapted and used in other regions and healthcare contexts, particularly those facing similar challenges in vaccine distribution and uptake.

In future work, the models could be refined and expanded with real-time data integration, enabling more dynamic predictions that account for evolving pandemic trends and vaccination campaigns. Additionally, integrating behavioral and social factors into the models could offer a more holistic understanding of vaccination decisions.

Overall, this research represents a significant step towards harnessing the power of machine learning in public health, providing a foundation for data-driven decision-making in the fight against COVID-19 and future health emergencies.

Acknowledgment

The authors extend their deepest gratitude to SIMAD University for their generous grant, which made the publication of this research possible. Their unwavering support not only reflects their commitment to academic excellence but also empowers researchers to explore new frontiers in their respective fields. This contribution has been invaluable in bringing this study to fruition.

References

[1] Ioannis Alexandros Charitos et al., “Update on COVID-19 and Effectiveness of a Vaccination Campaign in a Global Context,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 17, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Jack Mewhirter, Mustafa Sagir, and Rebecca Sanders, “Towards a Predictive Model of COVID-19 Vaccine Hesitancy among American Adults,” *Vaccine*, vol. 40, no. 12, pp. 1783-1789, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Gretchen B. Chapman, and Elliot J. Coups, “Predictors of Influenza Vaccine Acceptance among Healthy Adults,” *Preventive Medicine*, vol. 29, no. 4, pp. 249-262, 1999. [CrossRef] [Google Scholar] [Publisher Link]

[4] Bruna Aparecida Souza Machado et al., “The Importance of Vaccination in the Context of the COVID-19 Pandemic: A Brief Update Regarding the Use of Vaccines,” *Vaccines*, vol. 10, no. 4, pp. 1-25, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Jana L. Jacobs, Ghady Haidar, and John W. Mellors, “COVID-19: Challenges of Viral Variants,” *Annual Review of Medicine*, vol. 74, pp. 31-53, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Emmanuel O. Njoga et al., “Persisting Vaccine Hesitancy in Africa: The Whys, Global Public Health Consequences and Ways-Out-COVID-19 Vaccination Acceptance Rates as Case-in-Point,” *Vaccines*, vol. 10, no. 11, pp. 1-23, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Alexandra Savinkina et al., “Estimating Deaths Averted and Cost per Life Saved by Scaling up mRNA COVID-19 Vaccination in Low-Income and Lower-Middle-Income Countries in the COVID-19 Omicron Variant Era: A Modelling Study,” *BMJ Open*, vol. 12, no. 9, pp. 1-7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Maryam Diarra et al., “Non-Pharmaceutical Interventions and COVID-19 Vaccination Strategies in Senegal: A Modelling Study,” *BMJ Global Health*, vol. 7, no. 2, pp. 1-9, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Mohammed A.M. Ahmed et al., “Covid-19 Vaccine Acceptability and Adherence to Preventive Measures in Somalia: Results of an Online Survey,” *Vaccines*, vol. 9, no. 6, pp. 1-11, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Ayan Aden Moussa, Abdkeren Abdullahi Abdi, and Sharif Alhassan Abdullahi, “Assessment of Knowledge, Attitude and Practice of Healthcare Workers towards Hepatitis B Virus Infection in Mogadishu, Somalia: A Cross-Sectional Study,” *Research Square*, pp. 1-11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Mark McEvoy et al., “Loddon Mallee Healthcare Worker COVID-19 Study - Protocol for a Prospective Cohort Study Examining the Health and Well-Being of Rural Australian Healthcare Workers during the COVID-19 Pandemic,” *BMJ Open*, vol. 11, no. 8, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] J. Bryan Sexton, and Kathryn C. Adair, “Forty-Five Good Things: A Prospective Pilot Study of the Three Good Things Well-Being Intervention in the USA for Healthcare Worker Emotional Exhaustion, Depression, Work-Life Balance and Happiness,” *BMJ Open*, vol. 9, no. 3, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[13] Frances Kelly et al., “Improving Healthcare Worker Resilience and Well-Being during COVID-19 Using a Self-Directed E-Learning Intervention,” *Frontiers in Psychology*, vol. 12, pp. 1-9, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Vivek Raj S.N., and Manivannan S.K., “Machine Learning Models to Predict Covid-19 Vaccination Intention : An Indian Study,” *International Journal of Professional Business Review*, vol. 7, no. 6, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Alinoor Mohamed Farah et al., “Knowledge, Attitudes, and Practices Regarding COVID-19 among Health Care Workers in Public Health Facilities in Eastern Ethiopia: Cross-Sectional Survey Study,” *JMIR Formative Research*, vol. 5, no. 10, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Said Adam Sheikh et al., “Infection Prevention Practice and Associated Factors among Health Care Workers in Public Health Facilities of Mogadishu, Somalia, 2022,” *Archiv Euromedica*, vol. 13, no. 3, pp. 1-19, 2023. [Google Scholar]

[17] Hassan Abdullahi Dahie et al., “COVID-19 Vaccine Coverage and Potential Drivers of Vaccine Uptake among Healthcare Workers in Somalia: A Cross-Sectional Study,” *Vaccines*, vol. 10, no. 7, pp. 1-14, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Lucas Walz et al., “Knowledge, Attitudes and Practices Concerning Breast Cancer, Cervical Cancer and Screening among Healthcare Professionals and Students in Mogadishu, Somalia: A Cross-Sectional Study,” *Ecancermedicalscience*, vol. 16, pp. 1-14, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[19] Raed Alharbi et al., “Cultural-Aware Machine Learning Based Analysis of COVID-19 Vaccine Hesitancy,” *ICC 2023 - IEEE International Conference on Communications*, Rome, Italy, pp. 2864-2869, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Nikita Agarwal, and Ritam Dutta, “Comparative Predictive Analysis of Mortality Rate after Covid-19 Vaccination Using Various Machine Learning Approaches,” *2022 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, pp. 1-5, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Dimitris Papadopoulos et al., “Predictive Factors for Neutralizing Antibody Levels Nine Months after Full Vaccination with BNT162b2: Results of a Machine Learning Analysis,” *Biomedicines*, vol. 10, no. 2, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[22] Miftahul Qorib et al., “Covid-19 Vaccine Hesitancy: Text Mining, Sentiment Analysis and Machine Learning on COVID-19 Vaccination Twitter Dataset,” *Expert Systems with Applications*, vol. 212, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] Stephen Wai Hang Kwok, Sai Kumar Vadde, and Guanjin Wang, "Tweet Topics and Sentiments Relating to COVID-19 Vaccination among Australian Twitter Users: Machine Learning Analysis," *Journal of Medical Internet Research*, vol. 23, no. 5, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] Buddavarapu Teja Swaroop, "Evaluating Covid-19 Health Information Using Machine Learning," *International Journal of Engineering Applied Sciences and Technology*, vol. 5, no. 5, pp. 302-307, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[25] Marta Malesza, and Magdalena Bozym, "Factors Influencing COVID-19 Vaccination Uptake in an Elderly Sample in Poland," *medRxiv*, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[26] Richard F. Sear et al., "Quantifying COVID-19 Content in the Online Health Opinion War Using Machine Learning," *IEEE Access*, vol. 8, pp. 91886-91893, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] Konstantinos Giannakou, Georgia Fakonti, and Maria Kyprianidou, "Determinants of COVID-19 Vaccine Uptake among Healthcare Professionals and the General Population in Cyprus: A Web-Based Cross-Sectional Survey," *Journal of Evaluation in Clinical Practice*, vol. 28, no. 6, pp. 959-969, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[28] Nazeem Muhajarine et al., "COVID-19 Vaccine Hesitancy and Refusal and Associated Factors in an Adult Population in Saskatchewan, Canada: Evidence from Predictive Modelling," *PLoS One*, vol. 16, no. 11, pp. 1-18, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[29] Auriel A. Willette et al., "Using Machine Learning to Predict COVID-19 Infection and Severity Risk among 4510 Aged Adults: A UK Biobank Cohort Study," *Scientific Reports*, vol. 12, pp. 1-11, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[30] Jishan Ahmed et al., "Explainable Machine Learning Approaches to Assess the COVID-19 Vaccination Uptake: Social, Political, and Economic Aspects," *preprints.org*, pp. 1-34, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[31] Theo Audi Yanto et al., "Psychological Factors Affecting COVID-19 Vaccine Acceptance in Indonesia," *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, vol. 57, pp. 1-8, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[32] Landry Signé, Strategies for Effective Health Care for Africa in the Fourth Industrial Revolution: Bridging the Gap between the Promise and Delivery, Africa Growth Initiative at Brookings, pp. 1-46, 2021. [Online]. Available: https://www.brookings.edu/wp-content/uploads/2021/10/Strategies-for-effective-health-care-delivery-in-Africa_FINAL.pdf

[33] Abanoub Riad et al., "Belarusian Healthcare Professionals' Views on Monkeypox and Vaccine Hesitancy," *Vaccines*, vol. 11, no. 8, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[34] Abanoub Riad et al., "Universal Predictors of Dental Students' Attitudes towards Covid-19 Vaccination: Machine Learning-Based Approach," *Vaccines*, vol. 9, no. 10, pp. 1-19, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[35] Afsheen Afzal et al., "Impact of Local and Demographic Factors on Early COVID-19 Vaccine Hesitancy among Health Care Workers in New York City Public Hospitals," *Vaccines*, vol. 10, no. 2, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[36] Ma'mon M. Hatmal et al., "Reported Adverse Effects and Attitudes among Arab Populations Following COVID-19 Vaccination: A Large-Scale Multinational Study Implementing Machine Learning Tools in Predicting Post-Vaccination Adverse Effects Based on Predisposing Factors," *Vaccines*, vol. 10, no. 3, pp. 1-36, 2022. [CrossRef] [Google Scholar] [Publisher Link]