Mehmet Akif Ersoy University Journal of Health Sciences Institute

MAKU J. Health Sci. Inst.
2024, 12(2): 46-52.

Research Article

Targeting Vaccine Hesitancy: A Data-Driven Approach Using AI and Public Health Data

Bekir Çetintav[1], Ahmet Yalçın[2]

[1] https://orcid.org/0000-0001-7251-1211 – Burdur Mehmet Akif Ersoy University, Veterinary Faculty, Biostatistics Department, 15030 Burdur, TÜRKİYE
[2] https://orcid.org/0009-0001-6093-8282 – Burdur Mehmet Akif Ersoy University, Institute of Science, 15030, Burdur, TÜRKİYE

ABSTRACT

This study examines H1N1 and seasonal flu vaccination behaviors using machine learning models and explainable artificial intelligence (XAI) techniques. Utilizing data from the National 2009 H1N1 Influenza Survey, we developed a predictive framework employing models such as CatBoost, XGBoost, and LightGBM. CatBoost outperformed others with an accuracy of 0.696 and an F1 score of 0.688. SHAP (Shapley Additive Explanations) was used for interpretability, providing both global insights, such as the critical role of doctor recommendations, and local insights, highlighting individual decision factors. Our findings underscore the importance of addressing vaccine skepticism and improving healthcare communication to enhance vaccination uptake. These results contribute to public health strategies aimed at increasing immunization coverage and preparing for future pandemics.

Key words: Vaccine hesitancy, Healthcare communication, Machine learning, Explainable AI, One health, Public health

Corresponding author: Bekir Çetintav

E-mail: bekircetintav@mehmetakif.edu.tr

Submission Date: November 30, 2024
Acceptance Date: December 13, 2024
Available online: December 26, 2024

INTRODUCTION

Vaccination remains one of the most effective public health measures to prevent infectious diseases and reduce morbidity and mortality globally. Despite its proven efficacy, vaccine hesitancy continues to pose significant challenges in achieving widespread immunization coverage (Larson et al., 2014). Understanding the factors driving vaccine hesitancy is critical for designing targeted interventions and improving vaccination uptake (Lincoln et al., 2022). This study focuses on predicting vaccination behavior for H1N1 and seasonal flu vaccines, leveraging machine learning models and explainable artificial intelligence (XAI) techniques to provide both predictive accuracy and transparency.

The emergence of the H1N1 influenza virus in 2009 highlighted the urgent need for effective vaccination campaigns to control pandemics. It is estimated that between 151,000 to 575,000 deaths occurred globally within the first year of the outbreak (Harding & Heaton, 2018). Despite the rapid development and deployment of the H1N1 vaccine, vaccine uptake was hindered by factors such as public distrust, misinformation, and logistical barriers (Ayachit et al., 2020). Similar challenges persist with seasonal influenza, as rapid antigenic drift necessitates annual updates to vaccine formulations, complicating public health efforts to achieve adequate coverage (Harding & Heaton, 2018).

The One Health perspective, emphasizing the interconnectedness of human, animal, and environmental health, is particularly relevant in understanding the dynamics of influenza pandemics like H1N1. Influenza A viruses, which can circulate between humans, swine, and avian hosts, are a significant zoonotic threat, as demonstrated by the 2009 H1N1 pandemic (Pappaioanou & Gramer, 2010). The ability of these viruses to undergo genetic reassortment across species underscores the critical need for integrated surveillance systems and collaborative efforts in pandemic preparedness (Kim,

DOI: 10.24998/maeusabed.1593896

2018). Additionally, mislabeling such events as "swine flu" during the pandemic led to significant economic consequences for the pork industry, further highlighting the necessity for accurate communication and intersectoral cooperation (Pappaioanou & Gramer, 2010). Vaccination, both in humans and animals, plays a crucial role in mitigating these risks. For instance, efforts to control avian influenza through poultry vaccination and enhanced biosecurity measures have proven essential in preventing spillovers into human populations (Kim, 2018). The One Health framework supports these integrative strategies, advocating for collaborative research and policymaking to address zoonotic disease risks comprehensively (Monath, 2013).

Machine learning (ML) techniques have emerged as powerful tools to predict vaccination behavior by analyzing large datasets with complex, multidimensional variables. Studies have demonstrated the effectiveness of ensemble models like CatBoost and XGBoost in capturing critical predictors of vaccine hesitancy, including socioeconomic, demographic, and behavioral factors (Altarawneh, 2023; Ahmed et al., 2022). Additionally, SHAP (Shapley Additive Explanations) has been widely used to enhance the interpretability of these models, providing actionable insights for public health stakeholders (Lundberg & Lee, 2017).

Recent advancements in ML and XAI have enabled a deeper understanding of vaccine hesitancy, highlighting the role of trust, socioeconomic factors, and cultural perceptions in shaping individual decisions (Ebulue et al., 2024; Alharbi et al., 2024). For instance, studies have shown that lack of trust in government and healthcare institutions significantly correlates with lower vaccination rates (Lincoln et al., 2022). Moreover, the influence of socioeconomic factors, such as income and education, underscores the need for equitable access to vaccination resources, particularly in underserved populations (Ebulue et al., 2024).

This study contributes to the growing body of literature by employing a new and robust methodological framework that integrates predictive modeling with interpretability. By analyzing data from the National 2009 H1N1 Influenza Survey, we aim to identify key factors influencing vaccination decisions and provide a transparent understanding of model predictions. The findings are expected to inform strategies for improving vaccine uptake and addressing barriers to immunization, ensuring better preparedness for future public health crises.

MATERIAL AND METHODS

Study Design and Ethical Statement

This study is a secondary data analysis utilizing the publicly available National 2009 H1N1 Influenza Survey dataset (CDC, 2012), which was conducted to assess public attitudes, behaviors, and vaccination uptake during the H1N1 pandemic. The study employs a mixed-methods approach, integrating machine learning techniques with explainable artificial intelligence (XAI) methods to predict and interpret vaccination behaviors. The target variable is formulated as a 3-class classification system, representing "No vaccination received," "Single vaccination" and "Double vaccination." (“Single vaccination” means patient got just one of the vaccines.) The dataset used in this study is anonymized and publicly accessible, ensuring compliance with ethical standards for secondary data analysis. Since the data does not include identifiable personal information, ethical approval was not required for its use. The principles of the Declaration of Helsinki were adhered to throughout the research, ensuring respect for the autonomy and confidentiality of respondents.

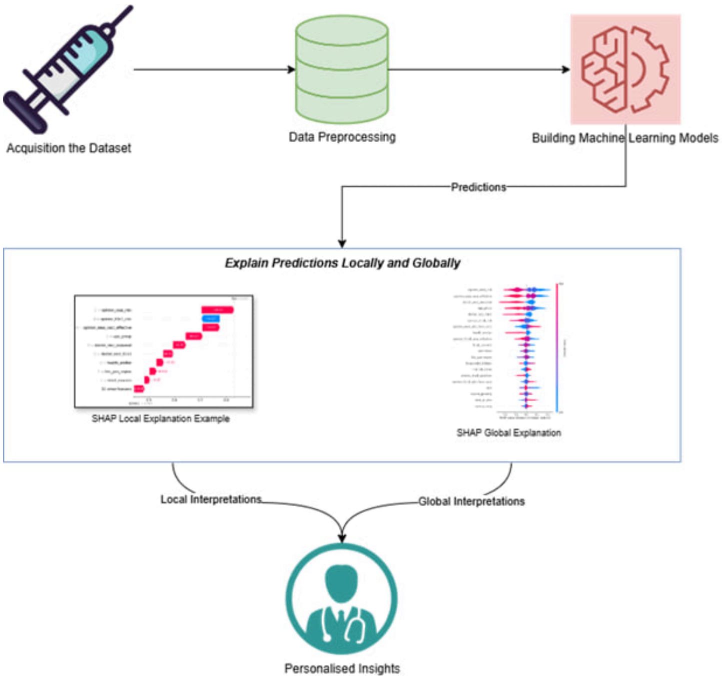

Figure 1: Framework for Generating Predictions and Interpreting Model Explanations

The methodological workflow for predicting and interpreting vaccination behavior is shown in the flowchart (Figure 1). The process begins with dataset acquisition, including features like demographics, opinions on vaccine effectiveness, risks, and doctor recommendations. After preprocessing to handle missing values, encode categorical variables, and remove irrelevant features, predictive models classify vaccination status into three categories: No vaccination, Single vaccination, and Double vaccination. Interpretability is enhanced with SHAP, providing local insights through waterfall plots for individual predictions and global explanations via summary plots to rank feature importance. This transparent framework aids public health officials in designing informed vaccination strategies.

Details of the Dataset

The dataset used in this study originates from the National 2009 H1N1 Flu Survey, conducted during late 2009 and early 2010. This large-scale survey was designed to assess public vaccination behaviors and attitudes in response to the H1N1 pandemic and seasonal influenza. The dataset provides critical insights into vaccination decisions, including demographic, socio-economic, behavioral, and health-related factors influencing individual vaccination outcomes.

The dataset comprises 36 columns and a multi-label target variable, reformulated as a 3-class classification structure for this study. The target variable combines information on two separate vaccination statuses—H1N1 vaccine and seasonal flu vaccine—into three categories:

- 0: No vaccination received
- 1: Single vaccination (either H1N1 or seasonal flu vaccine).
- 2: Double vaccination (both H1N1 and seasonal flu vaccines).

Each row corresponds to a unique respondent, identified by a respondent_id column, and includes 35 additional features capturing demographic, behavioral, and opinion-based data (Table 1).

Feature Engineering

Feature engineering was a critical step in preparing the dataset for machine learning. Missing values were addressed by replacing continuous variables with column means and filling categorical features with "No Category." Key categorical variables, such as age_group and education, were label-encoded, while non-informative columns like respondent_id and employment_occupation were removed to reduce noise. Feature selection combined domain expertise with statistical methods, including Random Forest for feature importance analysis and PCA for exploratory purposes. The refined dataset included demographic, behavioral, health-related, and opinion-based variables, offering a comprehensive foundation for modeling vaccination behavior.

Machine Learning Models

To predict vaccination status, the dataset was modeled as a 3-class classification problem with target labels representing no vaccination, single vaccination, and double vaccination. The following machine learning models were evaluated for their predictive performance:

- Random Forest Classifier: By aggregating the output of individual decision trees, the Random Forest Classifier combines several decision trees to get high accuracy. This paradigm works especially well with complicated data structures. Studies by Putri et al., (2021) and Qorib et al.(2023) demonstrate its success across different fields.
- K-Nearest Neighbors (KNN): A data point's class is predicted by the KNN model using the majority class of its closest neighbors. Studies by Suprayogi et al. (2022), Goswami & Sebastian (2022), demonstrate how well KNN works with a variety of datasets.
- XGBoost: XGBoost, which is well-known for its speed and effectiveness, builds a potent classifier by applying gradient boosting to decision trees. Studies by Cheong et al. (2021), and Nikhil et al.(2024) demonstrate the efficacy of XGBoost in a range of applications.
- LightGBM: Using a tree-based gradient boosting algorithm, LightGBM provides quick training and memory efficiency, making it ideal for huge datasets. Its scalability and efficiency on huge data are highlighted in studies by Ing et al. (2021), Gupta & Verma (2023).
- Support Vector Machine (SVM): To divide data points for classification, SVM creates a hyperplane that maximizes the margin between classes. It is frequently employed for assignments that call for a clear division of classes. The efficacy of SVM on a variety of datasets is demonstrated by studies by, To et al. (2021), and Du et al. (2017).
- CatBoost: Using a gradient boosting technique based on decision trees, CatBoost is made to effectively

handle category features. CatBoost's versatility for categorical data is demonstrated in studies by Ayachit et al. (2020), and Kim (2021).

We used accuracy, recall, precision, and F1 score as evaluation criteria to evaluate each model's performance. Every statistic offers a different viewpoint on how well the model predicts the likelihood of pet adoption.

Table 1: Vaccine Dataset Feature Descriptions

Demographic Features
- age_group: Respondent's age group.
- education: Education level.
- race: Race of the respondent.
- sex: Gender.
- income_poverty: Household income relative to the 2008 Census poverty thresholds.
- marital_status: Marital status.
- rent_or_own: Housing situation.
- employment_status: Employment status.
- hhs_geo_region: Geographic region defined by the U.S. Dept. of Health and Human Services.
- census_msa: Metropolitan statistical area classification.
- employment_industry: Employment industry (categorical).
- employment_occupation: Employment occupation (categorical).

Health-Related Factor Features:
- chronic_med_condition: Presence of chronic medical conditions.
- child_under_6_months: Regular close contact with children under six months.
- health_worker: Healthcare worker status.
- health_insurance: Health insurance status.

Behavioral Factors:
- behavioral_wash_hands: Frequently washed hands or used hand sanitizer.
- behavioral_large_gatherings: Reduced time at large gatherings.
- behavioral_antiviral_meds: Usage of antiviral medications.
- behavioral_avoidance: Avoided close contact with individuals showing flu-like symptoms.
- behavioral_face_mask: Purchased face masks.
- behavioral_outside_home: Reduced contact with people outside their household.
- behavioral_touch_face: Avoided touching eyes, nose, or mouth.

Opinions and Perceptions:
- h1n1_concern: Level of concern about H1N1 flu (0-3 scale).
- h1n1_knowledge: Level of knowledge about H1N1 flu (0-2 scale).
- opinion_h1n1_vacc_effective: Perceived effectiveness of the H1N1 vaccine (1-5 scale).
- opinion_h1n1_risk: Perceived risk of contracting H1N1 without vaccination (1-5 scale).
- opinion_h1n1_sick_from_vacc: Worry about sickness from H1N1 vaccine (1-5 scale).
- opinion_seas_vacc_effective: Perceived effectiveness of the seasonal flu vaccine (1-5 scale).
- opinion_seas_risk: Perceived risk of contracting seasonal flu without vaccination (1-5 scale).
- opinion_seas_sick_from_vacc: Worry about sickness from seasonal flu vaccine (1-5 scale).

Healthcare Access:
- doctor_recc_h1n1: Doctor's recommendation to get H1N1 vaccine.
- doctor_recc_seasonal: Doctor's recommendation to get seasonal flu vaccine.

Household Characteristics:
- household_adults: Number of adults in the household (capped at 3).
- household_children: Number of children in the household (capped at 3).

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Instances} \quad (1)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (2)$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (3)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Post-Explainability Techniques

The field of explainable artificial intelligence encompasses many methods that allow humans to interpret the output of artificial intelligence. In this way, people's opinions on the reasons behind and techniques used by artificial intelligence to generate prediction scores are influenced. To interpret the high-performing model's predictions both locally and globally, we applied SHAP (Shapley Additive Explanations), a well-liked explainability technique in explainable AI (Lundberg & Lee, 2017). We can understand how each feature affects predictions generally over the entire dataset (global explanation) and how particular qualities contribute to each prediction (local explanation) with the use of SHAP values.

Global Explanation

Global understanding is crucial for public health professionals and policymakers, as it highlights the most influential factors, such as medical recommendations, perceived vaccine efficacy, and underlying health conditions, in driving vaccination decisions. Such insights can guide more targeted health campaigns and interventions to improve vaccination uptake and enhance public health outcomes.

$$\hat{\phi}_j = \frac{1}{M} \sum_{m=1}^{M} \left(\hat{f}(x_{+j}^m) - \hat{f}(x_{-j}^m) \right) \quad (5)$$

In Equation 5, x is the sample of interest, j is the attribute index, f is the machine learning model, and M is the number of iterations. The prediction for x is "f ̂(x_(+j)^m)". However, a random number of attribute values were substituted with attribute values from a random z data points, except for the corresponding value of attribute j.To obtain all Shapley values, the process needs to be repeated for every feature.

Local Explanation

SHAP values, which illustrate the precise contribution of each feature to the vaccination likelihood for an individual respondent, provide insights into predictions at a local level. This approach is particularly valuable for case-by-case analysis, as SHAP can reveal, for example, how factors like a doctor's recommendation, perceived vaccine efficacy, or the presence of a chronic medical condition influence an individual's vaccination decision. These localized insights allow public health professionals to understand why the model assigned a specific vaccination probability to a particular respondent. Such transparency facilitates more personalized health communication strategies, enabling targeted interventions tailored to the unique concerns and circumstances of each individual.

The local SHAP value for a feature i for an instance x is given by :

where:

|S| is the number of features in subset S, f_x (S) is the model prediction using only the features in subset S,

$$\phi_j(i) = \sum_{S \subseteq N \backslash\backslash \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} \left(f_x(S \cup \{i\}) - f_x(S)\right) \quad (6)$$

f_x (S∪{i})-f_x (S) is the marginal contribution of feature i when it is added to subset S, and N is the set of all features.

This formula generates a SHAP value that represents the feature's influence on the particular prediction by calculating the weighted average of feature I's contribution over all potential feature subsets.

Computational Tools

This study leveraged Python and its ecosystem of libraries for data analysis, model development, evaluation, and explainability. pandas and numpy were employed for efficient data manipulation and numerical computations, while matplotlib and seaborn were used for data visualization. For preprocessing, scikit-learn facilitated tasks such as label encoding, missing value imputation, and train-test splitting, and was also utilized for implementing various machine learning models. Ensemble and boosting models, including XGBoost, LightGBM, and CatBoost, were developed using their respective specialized libraries. Lastly, the SHAP library was integrated to provide both local and global explanations of model predictions, ensuring transparency and interpretability.

RESULTS

The study evaluated multiple machine learning models for predicting vaccination status. CatBoost emerged as the best-performing model, achieving an

Table 2. Evaluation results of ML models

	Models	AccuracyScore	RecallScore	PrecisionScore	F1Score
1	CatBoost	0.696	0.696	0.689	0.688
2	XGBoost	0.686	0.686	0.677	0.675
3	LGBM	0.684	0.684	0.676	0.672
4	SVM	0.683	0.683	0.674	0.669
5	Random Forest	0.680	0.680	0.672	0.672
6	KNN	0.624	0.624	0.608	0.605

accuracy of 0.696 and an F1 score of 0.688, effectively handling the dataset's complexities. Gradient boosting models, XGBoost and LightGBM, followed closely with accuracy scores of 0.686 and 0.684, respectively, showcasing strong recall and precision. SVM demonstrated competitive performance with an accuracy of 0.683, while Random Forest showed balanced results with an accuracy of 0.680 and an F1 score of 0.672. KNN, though suitable for simpler tasks, achieved lower metrics, emphasizing its limitations with high-dimensional data

SHAP Global Interpretation

SHAP (Shapley Additive Explanations) values were calculated to provide a global interpretation of the CatBoost model's predictions for the "No vaccination received" group, as visualized in the summary plot (Figure 2). This analysis highlights the influence of each feature on the model's predictions, revealing both the direction and magnitude of their impact. Features like opinion_seas_risk, opinion_seas_vacc_effective, and doctor_recc_seasonal emerged as the most influential, reflecting their strong association with vaccination behavior.

High values of opinion_seas_risk (indicating a perceived low risk of seasonal flu) and opinion_seas_vacc_effective (skepticism about vaccine effectiveness) were positively associated with the prediction of not vaccinating. Conversely, the absence of doctor recommendations (doctor_recc_seasonal and doctor_recc_h1n1) strongly increased the likelihood of no vaccination, emphasizing the critical role of healthcare providers. Demographic factors like age_group and health_worker also contributed, with younger individuals and non-healthcare workers being less likely to vaccinate. Additionally, socioeconomic factors such as income and education showed smaller yet consistent effects, with higher levels generally reducing the likelihood of no vaccination.

SHAP Local Interpretation

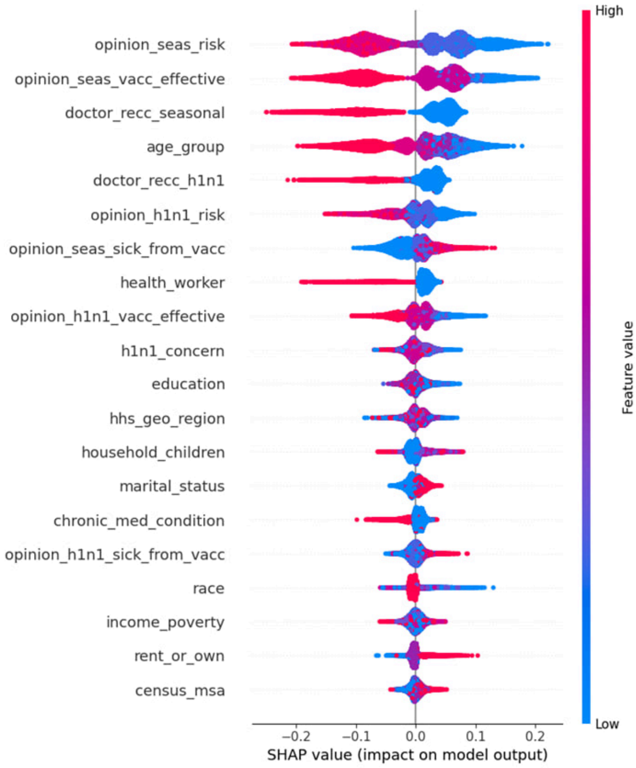

Figure 2. SHAP summary plot for global explanation of the prediction model

Two samples were selected for local interpretation: one with the target variable "Single vaccination" and the other with "No vaccination received," as these groups are critical for understanding decision patterns. The SHAP values for each sample were visualized using waterfall plots, which display the cumulative effect of each feature on the baseline prediction (mean model prediction). Features with positive contributions (red bars) increase the likelihood of the predicted outcome, while negative contributions (blue bars) decrease it, with bar lengths reflecting the magnitude of their impact (Molnar, 2020).

For the "No vaccination received" sample (Figure 3), the baseline prediction (E[f(x)] = 0.443) was adjusted by individual features to yield a final probability of f(x) = 0.83. Key factors included opinion_seas_risk, which had the largest positive contribution (+0.13), reflecting the individual's perception of low seasonal flu risk, and opinion_seas_vacc_effective (+0.07), indicating skepticism about vaccine effectiveness. The absence of doctor recommendations for both seasonal and H1N1 vaccines also strongly reinforced this outcome. Other features, such as age group and non-healthcare worker status, further contributed to the prediction.

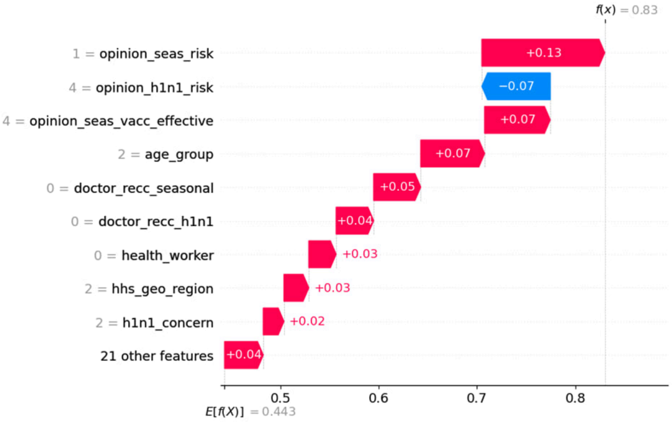

Figure 2. SHAP Waterfall plot of Sample 1 (No vaccination)

SHAP waterfall graph in Figure 4 explains the contribution of features to predicting "Single vaccination received" for a specific respondent. Starting from the baseline value (E[f(x)] = 0.349), feature contributions led to a final probability of f(x) = 0.95. The strongest predictor was the doctor's recommendation for seasonal vaccination (+0.15), followed by a belief in the high effectiveness of the seasonal vaccine (+0.11) and the respondent's age group (+0.10), which aligns with groups more likely to vaccinate. While the absence of an H1N1 vaccine recommendation contributed positively (+0.08), the perceived low risk of seasonal flu slightly decreased the likelihood (-0.04). Other factors, such as a positive perception of vaccine effectiveness and minimal concern about vaccine side effects, added minor but consistent positive impacts.

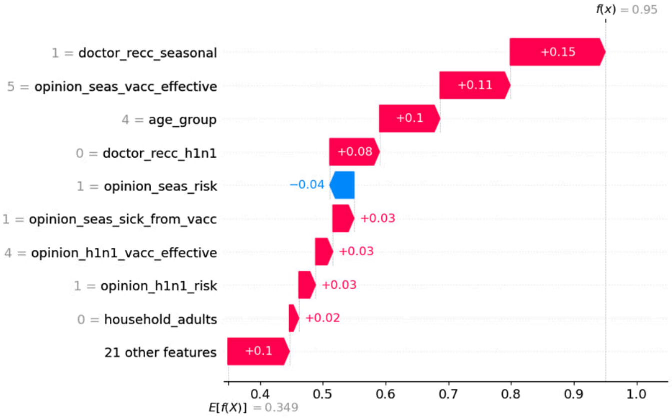

Figure 4. SHAP Waterfall plot of Sample 2 (Single vaccination)

DISCUSSION

In our study, we utilized machine learning models and explainable artificial intelligence (XAI) methods to predict and interpret behaviors related to H1N1 and seasonal flu vaccinations. These approaches provided critical insights into the factors influencing both individual and population-level vaccination decisions. The results, derived from the CatBoost model, highlighted the importance of perceptions of vaccine effectiveness, doctor recommendations, and demographic characteristics in shaping vaccination behaviors. SHAP analyses offered both global and local interpretations, contributing not only to understanding model performance but also to uncovering the underlying reasons behind vaccination decisions.

The global SHAP analysis revealed that doctor recommendations (for both seasonal and H1N1 vaccines), opinions on vaccine effectiveness, and demographic factors such as age group were the most influential features. These findings align with prior studies emphasizing the pivotal role of healthcare providers in vaccination uptake and the impact of public perceptions of vaccine efficacy (Pappaioanou & Gramer, 2010; Kim, 2021). Such insights suggest that targeted public health interventions should focus on enhancing doctor-patient communication and addressing vaccine skepticism through educational campaigns. For instance, training healthcare providers to proactively recommend vaccinations and effectively communicate their benefits could directly improve vaccination rates.

The local interpretation of a sample classified as "No vaccination received" highlighted specific barriers to vaccination, including a low perceived risk of seasonal flu, skepticism about vaccine effectiveness, and the absence of doctor recommendations. These results are consistent with findings from Ebulue et al. (2024), which noted that perceived risks alone are insufficient to motivate preventive actions. Addressing these barriers may involve tailored educational initiatives that counteract misinformation about vaccine efficacy and targeted outreach efforts in communities with low vaccination uptake.

In contrast, the local interpretation for a "Single vaccination received" case underscored the critical role of positive interactions with healthcare providers and trust in vaccine effectiveness. The individual's decision was strongly influenced by the recommendation of a seasonal vaccine and a belief in its efficacy. These insights suggest strategies such as implementing reminder systems for individuals who receive one vaccine to encourage follow-up vaccinations, and emphasizing the complementary benefits of both seasonal and H1N1 vaccinations in public health messaging.

CONCLUSION

In conclusion, the integration of global and local insights through SHAP provides actionable strategies to improve vaccination rates. Globally, the findings support the prioritization of doctor recommendations and efforts to combat vaccine skepticism. Locally, personalized interventions, such as reminders and tailored messages, can address individual concerns. Together, these approaches offer a robust framework for enhancing public health strategies and preparing for future pandemics.

ETHICAL APPROVAL

The dataset used in this study is anonymized and publicly accessible, ensuring compliance with ethical standards for secondary data analysis. Since the data does not include identifiable personal information, ethical approval was not required for its use.

ACKNOWLEDGEMENTS

We extend our gratitude to the esteemed researchers who made this dataset openly available.

AUTHOR CONTRIBUTIONS

Conceptualization, B.C., methodology, B.C and A.Y.; software, A.Y; validation, B.C and A.Y.; formal analysis, B.C..; investigation, B.C and A.Y.; resources, B.C.; data curation, B.C.; writing—original draft preparation, B.C and A.Y.; writing—review and editing, TBD.; visualization, A.Y.; supervision, B.C. All authors have read and agreed to the published version of the manuscript.

CONFLICT OF INTEREST

The authors declared that there is no conflict of interest.

RESEARCH FUNDING

There is no financial support for this study.

DATA AVAILABILITY

The data utilized in this study was acquired from an open-source report: Centers for Disease Control and Prevention, 2012. “2009 H1N1 Pandemic (H1N1pdm09 virus),”, accessed 14 October 2024. Available at https://www.cdc.gov/flu/pandemic-resources/2009-h1n1-pandemic.html. An alternative version of this dataset is also accessible via Kaggle: https://www.kaggle.com/datasets/arashnic/flu-data .

REFERENCES

Ahmed, J., Green, R., Alauddin, M. H., Saha, G., 2022. Explainable machine learning approaches to assess COVID-19 vaccination uptake. Preprints. (DOI: 10.20944/preprints202206.0115.v1)

Alharbi, R., Chan-Olmsted, S., Chen, H., Thai, M. T., 2024. Deep learning framework with multi-perspective social behaviors for vaccine hesitation. Social Network Analysis and Mining, 14, 140. (DOI: 10.1007/s13278-024-01301-1)

Altarawneh, L., 2023. Interpretable deep learning and transfer learning-based spatial-temporal modeling for vaccines demand prediction. Doctoral Dissertation, Binghamton University.

Ayachit, S. S., Kumar, T., Deshpande, S., Sharma, N., Chaurasia, K., Dixit, M., 2020. Predicting H1N1 and seasonal flu: Vaccine cases using ensemble learning approach. In: Proc. 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), IEEE, pp. 172-176.

CDC, 2012. 2009 H1N1 Pandemic (H1N1pdm09 virus). Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/flu/pandemic-resources/2009-h1n1-pandemic.html (Accessed: 14 October 2024).

Cheong, Q., Au-Yeung, M., Quon, S., Concepcion, K., Kong, J. D., 2021. Predictive modeling of vaccination uptake in US counties: A machine learning–based approach. Journal of Medical Internet Research, 23(11), e33231.

Du, J., Xu, J., Song, H., Liu, X., Tao, C., 2017. Optimization on machine learning-based approaches for sentiment analysis on HPV vaccines-related tweets. Journal of Biomedical Semantics, 8, 1–7.

Ebulue, C. C., Ekkeh, O. V., Ebulue, O. R., Ekesiobi, C. S., 2024. Leveraging machine learning for vaccine distribution in resource-limited settings. International Medical Science Research Journal, 4(5), 544–557. (DOI: 10.51594/imsrj.v4i5.1120)

Goswami, M., Sebastian, N. J., 2022. Performance analysis of logistic regression, KNN, SVM, Naïve Bayes classifier for healthcare application during COVID-19. In: Proc. Innovative Data Communication Technologies and Application: ICIDCA 2021, Springer, Singapore, pp. 645–658.

Gupta, H., Verma, O. P., 2023. Vaccine hesitancy in the post-vaccination COVID-19 era: A machine learning and statistical analysis-driven study. Evolutionary Intelligence, 16(3), 739–757.

Harding, A. T., Heaton, N. S., 2018. Efforts to improve the seasonal influenza vaccine. Vaccines 6(19). (DOI: 10.3390/vaccines6020019)

Ing, S. H., Abdullah, A. A., Harun, N. H., Kanaya, S., 2021. COVID-19 mRNA vaccine degradation prediction using LR and LGBM algorithms. Journal of Physics: Conference Series, 1997(1), 012005.

Kim, J. S., 2021. Covid-19 prediction and detection using machine learning algorithms: CatBoost and linear regression. American Journal of Theoretical and Applied Statistics, 10(5), 208.

Kim, S. H., 2018. Challenge for One Health: Co-Circulation of Zoonotic H5N1 and H9N2 Avian Influenza Viruses in Egypt. Viruses, 10(3), 121. (DOI: 10.3390/v10030121)

Larson, H. J., Jarrett, C., Eckersberger, E., Smith, D. M. D., Paterson, P., 2014. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: A systematic review of published literature. Vaccine, 32(19), 2150–2159. (DOI: 10.1016/j.vaccine.2014.01.081)

Lincoln, T. M., Schlier, B., Strakeljahn, F., et al., 2022. Taking a machine learning approach to optimize prediction of vaccine hesitancy in high-income countries. Scientific Reports, 12, 2055. (DOI: 10.1038/s41598-022-05915-3)

Lundberg, S. M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, pp. 4765–4774.

Monath, T. P., 2013. Vaccines against diseases transmitted from animals to humans: A one health paradigm. Vaccine, 31(44), 4859–4864. (DOI: 10.1016/j.vaccine.2013.07.037)

Nuwarda, R. F., Rozek, L. S., Nihlén Fahlquist, J., et al., 2021. Socioeconomic and cultural influences on vaccine uptake. Vaccine, 39(15), 2000–2012. (DOI: 10.1016/j.vaccine.2021.02.040)

Pappaioanou, M., Gramer, M., 2010. Lessons from Pandemic H1N1 2009 to Improve Prevention, Detection, and Response to Influenza Pandemics from a One Health Perspective. ILAR Journal, 51(3), 268–280. (DOI: 10.1093/ilar.51.3.268)

Putri, V. M., Masjkur, M., Suhaeni, C., 2021. Performance of SMOTE in a random forest and naive Bayes classifier for imbalanced Hepatitis-B vaccination status. Journal of Physics: Conference Series, 1863(1), 012073.

Qorib, M., Oladunni, T., Denis, M., Ososanya, E., Cotae, P., 2023. Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. Expert Systems with Applications, 212, 118715.

Suprayogi, S., Sari, C. A., Rachmawanto, E. H., 2022. Sentiment analysis on Twitter using the K-Nearest Neighbors (KNN) algorithm against COVID-19 vaccination. Journal of Applied Intelligent Systems, 7(2), 135–145.

To, Q. G., To, K. G., Huynh, V. A. N., Nguyen, N. T., Ngo, D. T., Alley, S. J., Vandelanotte, C., 2021. Applying machine learning to identify anti-vaccination tweets during the COVID-19 pandemic. International Journal of Environmental Research and Public Health, 18(8), 4069.