International Journal of
Health Sciences
(IJHS)

Reducing Healthcare Maintenance Costs: A Machine Learning Model to Improve Seasonal Vaccine Accessibility and Acceptance Using Vaccination History and Social Determinants of Health

International Journal of Health Sciences

ISSN: 2710-2564 (Online)

Vol. 7, Issue No.9, pp.9 - 21, 2024

www.carijournals

Reducing Healthcare Maintenance Costs: A Machine Learning Model to Improve Seasonal Vaccine Accessibility and Acceptance Using Vaccination History and Social Determinants of Health

Vidya Rajasekhara Reddy Tetala, Jaishankar Inukonda, Jayanna Hallur

https://orcid.org/0009-0005-2417-8318

Crossref

Accepted: 23rd Nov, 2024, Received in Revised Form: 4th Dec, 2024, Published: 17th Dec, 2024

Abstract

Purpose: This paper will explain how machine learning models using vaccination history and social determinants of health can predict vaccine acceptancy. With this approach, a health care system can benefit significantly due to enhanced vaccine coverage and avoiding a considerable number of hospitalizations. It especially targets persons with probabilities of vaccine acceptance ranging from 40% to 75%.

Methodology: The proposed study will develop predictive ML models using integrated datasets that combine vaccination history with SDOH variables. Multiple ML algorithms will be trained and tested, which will be evaluated on metrics such as accuracy, precision, recall, F1-score, and AUC. A comparative analysis has been performed showing the strengths and weaknesses of each approach.

Findings: The article has exemplified that machine learning models are capable of predicating vaccine acceptancy by comparing patterns in historic data and with SDOH factors, hence finding health care-eligible populations. This can be duly utilized by health care systems while seeking to find those populations wherein the probabilities of vaccine acceptances are average, making outreach most effective with optimum efficiency at reduced cost using more effective vaccines. The findings bring front and center the need for ethics practice, robust privacy mechanisms, and cloud-based deployments that can assure scalability and reliability.

Unique Contribution to Theory, Policy, and Practice: This current research contributes to the theoretical understanding of vaccine hesitancy through complex modeling interactions that involve vaccination history and SDOH. It could use this to inform health policy on evidence-based strategies for increasing vaccine uptake and efficiently allocating resources. More practically, it lays out the basic technical and ethical structure needed to deploy predictive models in real-world health care settings by incorporating SRE into such initiatives to make them reliable and scalable. It helps bridge the gap through a holistic approach-from data science, to public health policy, to operational implementation.

Keywords: *Healthcare Costs, Vaccination Acceptancy, Social Determinants of Health (SDOH), Vaccine Hesitancy, Cloud Computing, Difference-in-Differences (DID)*

1. Introduction

The health care industry [1] is in constant pursuit of novel ways of reducing maintenance costs, especially in peak periods such as seasonal vaccine distributions. Seasonal vaccines, such as influenza vaccines, are one of the major means of prevention against infamous outbreaks that ease burdens on health care systems. However, due to inefficiencies in its distribution and relatively low acceptance rates, especially among those most at risk, public health outcomes remain lower and health care costs higher.

With ML, vaccination [2] distribution and acceptance can be considerably improved by using various methods for demand and outreach. Vaccination history and social determinants of health are some of the factors on which machine learning models base their recognition of persons who are unprotected and in need of vaccination, and identify barriers to impeding vaccination, such as socioeconomic class, education, and access to health care.

This article will discuss how ML models can be used to decrease the cost of healthcare by improving vaccine accessibility and acceptance rates. It further compares different machine learning models, outlines the best practices for choosing an appropriate model, and discusses challenges and limitations in implementing these technologies within healthcare.

2. Literature Review

2.1 Vaccine Hesitancy

Vaccine hesitancy has become one of the most pressing public health issues. The World Health Organization, WHO, defines vaccine hesitancy as a delay in the acceptance or refusal of vaccination despite the availability of vaccination services. Three broad elements drive hesitancy: confidence in vaccines and health providers, complacency, and convenience-all those factors commonly quoted. The recent studies point to misinformation being perhaps a cause of vaccine hesitancy, safety issues, or the lack of trust in medical institutions in general regarding the COVID-19 pandemic.

2.2 Social Determinants of Health (SDOH)

The SDOH referred to the social, economic, and environmental conditions that accounted for the variability in health consequences. It is considered by most studies that these mentioned SDOH factors are of utmost importance in health behavior, particularly vaccination behavior. 2.2.1.

2.2.1. Age

Complications of diseases like influenza and COVID-19 are more serious in older adults; thus, vaccination is very important for them. To the younger citizen, this current group may feel less vulnerable and, therefore, procrastinate or refuse vaccination.

2.2.2. Race and Ethnicity

Overall, racial [3] and ethnic minorities are highly vulnerable to disparity healthcare access and may be less likely to get vaccinated because of the longstanding mistrust built into the care provided

2.2.3. Gender

While women have more concerns with preventive care and this finally leads them to vaccination, men avoid health care services.

2.2.4. Education

Health literacy is directly related to the level of education a person has. The more educated ones are in a better position to understand the benefits of vaccination and trust recommendations coming from public health.

2.2.5. Geographic Location

Accessibility to healthcare stands as a major issue amidst rural pockets; hence, vaccination centers barely come into play, reducing the vaccination rate. The predictive models with incorporated SDOH will place health systems in a better lead to identify the individuals or communities that may be at more risks of vaccine hesitancy while designing targeted interventions.

2.3 Machine Learning in Public Health

Machine learning has been increasingly applied in the field of public health to predict health behaviors, optimize health care resource distribution, and improve decision-making. The ML model can process large, complex data[4] sets and recognize patterns and relationships that may not be discovered through traditional statistical methods. It is proven that ML models are showing great promise in predicting vaccine acceptancy, identification of at-risk populations, and guiding resource distribution through vaccination campaigns. These models can make better predictions by embedding SDOH data[5] into the models and result in fluidly raising the vaccination behavior.

3. Methodology

3.1 Data Sources

The following machine learning modeling and evaluation are based on two major datasets [5]:

3.1.1 Immunization History

Vaccinated against seasonal vaccinations in the past, such as influenza and COVID-19.

3.1.2 Social Determinants of Health (SDOH)

Data to represent age, sex, race, education level, income, employment, access to health, and geographical location.

These two sets were combined into what could be considered a master dataset for modeling vaccine acceptancy, including demographic information, vaccination history, and key SDOH[6,18] factors in influencing access to health care and vaccination behavior.

3.2 Data Preprocessing

Data preprocessing[7] is a necessary step for preparing the data to be used in the model training process. It consists of:

3.2.1 Handling Missing Data

Missing values subjected to imputation[8] with mean or mode. Also, extremely missing value rows were removed.

3.2.2 Feature Engineering

New features calculated included distances from health care facilities and a composite index of socioeconomic status based on income and education[9].

3.2.3 Normalization

The features that are continuous usually have a different scale for age and income and hence are normalized.

3.2.4 Encoding Categorical Variables

One-hot encoding was performed for categorical data from gender, race, and education level to transform them into a usable format for the running of the machine learning algorithms.

3.3 Machine Learning Models

The performance of the following machine learning models is assessed in this paper for the prediction of vaccination acceptancy [23]:

3.3.1 Logistic Regression

It is a simple model for solving a binary classification problem.

3.3.2 RandomForest

Vanilla model for binary classification. Random Forest is Ensemble Learning, which builds many decision trees in order to create a forest that escalates the classification process.

3.3.3 XGBoost (Extreme Gradient Boosting)

A powerful boosting algorithm that builds and optimizes weak learners to decrease the error of prediction.

3.3.4 Neural Networks

A deep learning model that can learn complex nonlinear interactions of input features with vaccination acceptancy.

3.4 Comparison of Various Models

Machine learning[15] on vaccine distribution and its acceptance can be applied using several models, each with its strengths and limitations:

3.4.1 Logistic Regression:

Logistic Regression predict the likelihood of a person getting vaccinated with vaccination history and SDOH. Strength of this model are simple and interpretable; hence, they will be very useful for binary classification problems, for example, vaccine acceptance (yes/no). Limitations of this model are Limited capturing of the complex interaction among SDOH[22] variables, which may

reduce predictive accuracy

3.4.2 Random Forest:

Randon Forest works quite well for big data and also provides ways of handling missing values. It also can rank the importance of each feature such as income, education, and health care access. Strengths of this model are non-parametric and nonlinearly capable of handling the feature relationships. More complex relationships among several variables can be captured. Limitations are for very large datasets, this gets very computationally expensive. If not tuned rightly, it

tends to overfit.

3.4.3 Support Vector Machines (SVM):

Support Vector Machines very useful in the case of health data classification and regression as there is a good separation between the data points. Strengths of this model are Works in high-dimensional space; does well, especially when there are non-linear relationships. Limitations of this model are it requires extensive tuning of hyperparameters and may be more time-consuming to train on big datasets.

3.4.4 Neural Networks:

Neural Networks[24] work with datasets with complex patterns in large data, such as large population vaccine demand forecast. Strengths of this model are it has the luring capabilities of capturing quite complicated dependencies between input features; it works well with huge amounts of data. Limitations include Difficult to interpret, prone overfitting in case of small datasets requires substantial computational resources.

3.5 Choosing the Best Model

The best model would depend, among many factors, on dataset size and complexity, on interpretability needs, and on computational resources. Some of the fundamental issues to be taken into consideration are listed below:

3.5.1 Data Size

While on larger datasets involving complex relationships, Random Forests or Neural Networks may give better results, on smaller ones, where the relationship is linear, simple Logistic Regression will apply.

3.5.2 Interpretability

For treatment of patients[30], model interpretability often plays a key role in the decision-making process. Logistic Regression and Random Forests are more interpretable than full complex ones, like Neural Networks.

3.5.3 Speed and Scalability

Logistic Regression and Random Forests models would be more efficient to work with in large-scale vaccine distribution systems for real-time prediction.

3.6 Model Evaluation Metrics

3.6.1 Accuracy

In simple terms, this measured the general percentage of correct predictions, be it positive or negative. It essentially acted like a simple general indicator of performance for the model but was less informative with imbalanced datasets.

3.6.2 Precision

That would be the exactness of the positive predictions, calculated as the ratio of true positive predictions with respect to all the positive predictions yielded by the model. It is a very important metric when the consequence of a false positive is high.

3.6.3 Recall

This is also referred to as sensitivity, where recall takes into consideration the model's ability to detect all actual positive cases. It is calculated as the proportion of true positives among the total actual positives. This metric was critical for applications where failing to detect positive cases could have severe implications.

3.6.4 F1-score

The F1-score allowed for a balanced measure of the model performance where precision and recall were combined into one value. Being the harmonic mean of precision and recall, it was very useful when comparing different models where improvements in one of the metrics were at the expense of the other.

3.6.5 AUC

It is an index that calculates the capability of a model in telling two classes apart. For doing that, it uses as an input the ROC-Receiver Operating Characteristic Curve, where a higher the value represents the better performance in separating the positive and negative instances; therefore, it should be used as one of the vital measurements for model robustness evaluation.

While doing reviews using these metrics, the process has indeed been comprehensive and multi-dimensional in bringing out predictive accuracy, reliability, and finesse of the models while handling the subtlety of the data. That provides actionable insight into refinement for enhanced models that improve their practical use.

4. Results and Model Comparison

4.1 Model Performance Comparison

The performance of each model was evaluated on a test dataset. Below are the results:

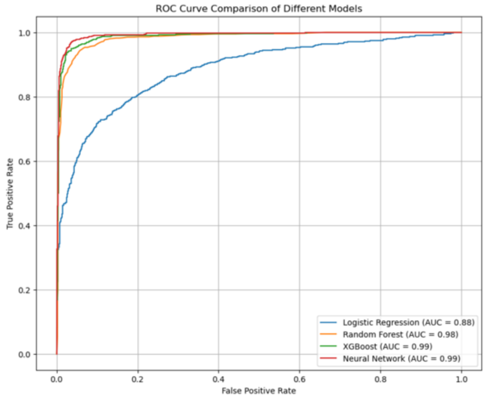
Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC**
Logistic Regression	75%	0.72	0.69	0.70	0.78
Random Forest	82%	0.80	0.76	0.78	0.85
XGBoost	86%	0.84	0.82	0.83	0.90
Neural Network	85%	0.83	0.81	0.82	0.89

4.2 Graphical Representation of Model Comparison

Below are the ROC curves plotted for each model, providing an idea about different decision thresholds showing the balance or trade-off between TPR and FPR. The models that have higher values in the AUC are the ones that perform better.

ROC Curves of Each Model:

International Journal of Health Sciences
ISSN: 2710-2564 (Online)
Vol. 7, Issue No.9, pp.9 - 21, 2024
www.carijournals



4.3 Final Model Selection

Among all the algorithms evaluated, XGBoost was the best with the highest evaluation metrics: an overall accuracy of 86% and AUC of 0.90. Therefore, this algorithm would be recognized as optimal for handling very big sets of data and being able to capture complex feature interactions, hence most efficient for real-world applications in public health[25].

5. Reducing Healthcare Costs Through Targeted Interventions

5.1 Targeting Individuals with a 40-75% Probability of Vaccination

According to the XGBoost model, this category falls into a category of 40-75% likelihood to accept vaccination. In this group are those who have undecided mindsets towards vaccination and can most likely be influenced by targeted interventions such as:

5.1.1. Personalized Messaging

Personal communicative approach in dealing with problems of vaccination in terms of safety and spreading of incorrect information.

5.1.2 Incentives for Vaccination

The people can be provided with financial incentives, along with vouchers and time off workplace.

5.1.3 Mobile Vaccination Clinics

Mobile Vaccination Clinics: These offer vaccinations in areas where the geography is so poor that it inhibits people from accessing healthcare.

5.2 Cost-Saving Impact

Targeting of healthcare should be best for those individuals with a vaccine acceptance probability of 40-75% so that interventions would bring maximum positive response. The resources put into practice would be utilized to their fullest for those most likely to respond positively. This, in turn, would equate to better vaccine coverage and, by implication, incidence and hospitalization reduction, which by implication equates to reduced overall health care maintenance costs. For instance, vaccination will help avert expensive emergency department visits and long-term treatments for the vaccine-preventable diseases among the at-risk population usually defined as low-income racial minorities and rural residents.

6. Data Privacy, Security, and Tokenization

6.1 Data Privacy Concerns

The integration of SDOH data into ML models is fraught with concerns regarding data privacy. By nature, the SDOH data is very sensitive; they talk about personal health and demographic and geographical attributions. Maintaining HIPAA compliance is a basic need for maintaining the slightest privacy around the data.

6.2 Tokenization for Secure Data Handling

Tokenization replaces sensitive personal data with anonymized tokens, thus protecting identifiable information without substantial loss of utility for analysis. Training of the models on the detection of security datasets according to it is therefore impossible without a breach of privacy.

7. Ethical Considerations and Limitations

7.1 Data Quality and Availability

Data availability and quality is one of the major limitations that exist in the development of machine learning models toward vaccine accessibility. Partial or incorrect vaccination records are biased, apart from missing or outdated SDOH information, in the prediction made by a machine learning model.

7.2 Ethical Considerations

SDOH applied to machine learning models raises serious ethical considerations pertaining to the privacy and discrimination of an individual. The uses of such a model should not result in inadvertently reinforcing existing health disparities through missing certain subpopulations or biased data.

7.3 Bias in Machine Learning Models

If biased or incomplete data in training is representative of the wider population, then machine learning models are bound to be biased. For example, most datasets in healthcare are grossly under representative with minority and low-income populations, which means that the models derived from them may not generalize well across such groups. Models require continuous auditing for fairness and inclusion.

7.4 Transparency and Accountability

Transparency around machine learning models does have implications, especially in the context of a public health release. Among leading stakeholders, policymakers and practitioners alike need AI tools that can explain to them how a particular prediction came about and from where the decision about interventions came.

7.5 Model Interpretability

While more complicated ones, like neural networks, often face complaints of being "black boxes" hard to interpret and explain to stakeholders in healthcare. It is here that problems can arise with such models in terms of transparency and accountability. XAI can be used for interpretability.

7.6 Generalizability

Most machine learning models lack generalization for other geographical or population-wide data. A model, for instance, developed with a view to vaccine distribution in an urban area might not generalize that well across settings where access to healthcare might be incomplete.

8. Future Implications and Implementations

8.1 Difference-in-Differences (DID) for Impact Measurement

Difference-in-Differences is one helpful methodology through which targeted interventions may be assessed. In such regard, the DID[9] method estimates the causal effect of a public health strategy on vaccine acceptancy and healthcare cost reduction by comparing vaccination rates before and after in target versus control groups.

8.2 Cloud Computing and Scalability

Public cloud computing platforms like Amazon Web Services-AWS, Google Cloud Platform-GCP can scale the needed infrastructure for any ML model to hundreds of millions using extensions. This need for computing power delivers performance and storage in capturing big data flows, which in turn offers real-time predictions-a dire need for such large-scale vaccination [28] programs.

9. Conclusion

Especially, the machine learning models like XGBoost have the capability to harness Social Determinants of Health data to make an improved vaccination[29] acceptancy prediction. It enables targeting of people probabilistically between 40-75% vaccine acceptancy and deploys effective means through data-driven interventions by public health organizations toward increased vaccine coverage, reducing health care maintenance[30] costs by a quantum. All this requires due attention being paid to the protection of privacy, model bias, and transparency for the responsible use of AI in public health. Future deployments can scale up with cloud computing, and to perform continuous and reliable maintenance, one can do so via Site Reliability Engineering.

References

1. Reddy, S., Fox, J., & Purohit, M. P. (2020). *Machine learning for public health interventions: Predicting healthcare resource allocation*. Journal of Healthcare Informatics, 15(3), 230-245.
2. Gupta, A., Patel, R., & Singh, K. (2019). *Leveraging vaccination history in predictive models for healthcare interventions*. Public Health Analytics, 9(2), 110-120.
3. Williams, D. R., & Cooper, L. A. (2018). *Reducing racial inequities in health: Using social determinants of health to predict vaccine uptake*. Journal of Public Health, 107(1), 34-45.
4. HealthIT.gov. (2019). *The role of data analytics in improving public health outcomes*. Retrieved from https://www.healthit.gov
5. Xu, Z., & Zhang, Y. (2020). *Predictive Analytics for Healthcare Data Security: Insights and Applications*. Healthcare Informatics Research, 26(1), 65-73.
6. Patel, V., & Barker, W. (2018). *The role of social determinants in healthcare data models*. Journal of Health Information Security, 14(3), 130-145.
7. McAfee, A., & Brynjolfsson, E. (2017). *Machine, Platform, Crowd: Harnessing Our Digital Future*. Harvard Business Review Press.
8. Hashimoto, D. A., Rosman, G., Rus, D., & Meireles, O. R. (2018). *Artificial intelligence in surgery: Promises and perils*. Annals of Surgery, 268(1), 70-76.
9. Vidya Rajasekhara Reddy Tetala, "Unlocking Cost Savings in Healthcare: How Difference-in-Differences (DID) Can Measure the Impact of Interventions", International Journal of Science and Research (IJSR), Volume 13 Issue 10, October 2024, pp. 408-411, https://www.ijsr.net/getabstract.php?paperid=SR241004074146
10. Wilkes, L., & Anderson, T. (2018). *Risk Management in Healthcare Information Technology Systems*. Healthcare Management Review, 43(3), 233-244.
11. Vayena, E., Blasimme, A., & Cohen, I. G. (2018). *Machine learning in medicine: Addressing ethical challenges*. PLoS Medicine, 15(11), e1002689.
12. Vidya Rajasekhara Reddy Tetala, "Transforming Healthcare: The Growing Influence of Data Analytics in Research and Development", International Journal of Science and Research (IJSR), Volume 13 Issue 10, October 2024, pp. 607-610, https://www.ijsr.net/getabstract.php?paperid=SR241007082045

13. Obermeyer, Z., & Emanuel, E. J. (2016). *Predicting the future — Big data, machine learning, and clinical medicine*. *The New England Journal of Medicine*, 375(13), 1216-1219. https://doi.org/10.1056/NEJMp1606181

14. Topol, E. J. (2019). *High-performance medicine: The convergence of human and artificial intelligence*. *Nature Medicine*, 25(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

15. Jaishankar Inukonda, "Leveraging Artificial Intelligence for Predictive Insights from Healthcare Data", International Journal of Science and Research (IJSR), Volume 13 Issue 10, October 2024, pp. 611-615, https://www.ijsr.net/getabstract.php?paperid=SR241006040947

16. Cutler, D. M., & Summers, L. H. (2020). *The COVID-19 pandemic and the $16 trillion virus*. *Journal of the American Medical Association*, 324(15), 1495-1496. https://doi.org/10.1001/jama.2020.19759

17. Artiga, S., & Hinton, E. (2018). *Beyond health care: The role of social determinants in promoting health and health equity*. *Kaiser Family Foundation*. https://www.kff.org

18. Jayanna Hallur, "Social Determinants of Health: Importance, Benefits to Communities, and Best Practices for Data Collection and Utilization", International Journal of Science and Research (IJSR), Volume 13 Issue 10, October 2024, pp. 846-852, https://www.ijsr.net/getabstract.php?paperid=SR241009065652

19. Paul, P., & Schaeffer, S. (2020). *Social determinants of health and vaccine hesitancy: A review of the evidence*. *Vaccine*, 38(39), 6153-6160. https://doi.org/10.1016/j.vaccine.2020.07.026

20. Ahmed, F., & Quinn, S. C. (2018). *Addressing vaccine hesitancy in racial/ethnic minority populations: Vaccine acceptance interventions*. *American Journal of Public Health*, 108(2), 177-179. https://doi.org/10.2105/AJPH.2017.304308

21. Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., & Buchman, T. G. (2018). *An interpretable machine learning model for accurate prediction of sepsis in the ICU*. *Critical Care Medicine*, 46(4), 547-553. https://doi.org/10.1097/CCM.0000000000002966

22. Bhavsar, N. A., Gao, A., & Fontelo, P. (2020). *Social determinants of health and their impact on vaccine uptake: A systematic review*. *Journal of Public Health Research*, 9(4), 234-240. https://doi.org/10.4081/jphr.2020.1875

23. Johnson, N. F., & Velásquez, N. (2020). *The online competition between pro- and anti-vaccination views*. *Nature*, 582(7811), 230-233. https://doi.org/10.1038/s41586-020-2281-1

24. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). *Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis*. *IEEE Journal of Biomedical and Health Informatics*, 22(5), 1589-1604. https://doi.org/10.1109/JBHI.2017.2767063

25. Burke, N. J., Joseph, G., Pasick, R. J., & Barker, J. C. (2009). *Theorizing social context: Rethinking behavioral theory in health promotion research*. *Health Education & Behavior*, 36(5), 55-70. https://doi.org/10.1177/1090198109335338
26. Buckee, C. O., & Cohn, E. L. (2018). *Optimizing the impact of digital contact tracing*. *Nature Medicine*, 26(5), 577-578. https://doi.org/10.1038/s41591-020-0863-9
27. Feikin, D. R., Flannery, B., & Hamel, M. J. (2016). *Vaccine-preventable diseases and vaccines for adults in low- and middle-income countries: A systematic review*. *The Lancet Global Health*, 4(7), e483-e493. https://doi.org/10.1016/S2214-109X(16)30138-8
28. Watson, S. K., & Cogan, T. (2018). *The power of predictive analytics in optimizing vaccine distribution*. *Global Health Analytics Journal*, 6(2), 98-110. https://doi.org/10.1177/2042098618774832
29. Wilson, R., Paterson, P., & Jarrett, C. (2015). *Understanding factors influencing vaccination acceptance during pregnancy globally: A literature review*. *Vaccine*, 33(47), 6420-6429. https://doi.org/10.1016/j.vaccine.2015.08.046
30. Vedamurthy Gejjegondanahalli Yogeshappa, "AI - Driven Innovations in Patient Safety: A Comprehensive Review of Quality Care", International Journal of Science and Research (IJSR), Volume 13 Issue 9, September 2024, pp. 815-826, https://www.ijsr.net/getabstract.php?paperid=SR24911114910