# Classification of an Individuals Vaccination Status Using Ensemble Hard Voting Classifier

**[1]Saranya S and [2]Agusthiyar R**
[1,2] Department of Computer Science and Application, SRM Institute of Science and Technology, Ramapuram, Chennai, India.
[1]saranyasivapragasam88@gmail.com, [2]agusthir@srmist.edu.in.

Correspondence should be addressed to Agusthiyar R : agusthir@srmist.edu.in

**Article Info**
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202404091
Received 30 March 2024; Revised from 25 May 2024; Accepted 29 July 2024.
Available online 05 October 2024.


**Abstract –** Vaccination is a proactive medical immunization procedure where an inactivated form of a disease-causing agent (such as a virus) is administered to boost the body's defense systems. Efficient management of vaccination status is crucial in healthcare management, disease eradication, community immunity ("herd immunity"), disease prevention, and global health security. Ensuring precise monitoring and validation of an individual's vaccination status is indispensable, especially in the context of emerging diseases and epidemics. This study evaluates the likelihood of individuals obtaining vaccination for the H1N1 virus and the seasonal flu vaccine. Ensemble methods combine the predictions of multiple base classifiers to enhance overall performance. One such method, the hard voting classifier, aggregates the votes from each base classifier and selects the class with the majority vote as the final prediction. This approach leverages the strengths of different classifiers, reducing the risk of individual model biases and improving generalization using metrics such as precision, recall, accuracy, and F1-score are employed to assess the system's effectiveness. The results demonstrate how data-driven methods can address population wellness and improve vaccination rates using an ensemble method. The proposed ensemble hard voting classifier achieved accuracies of 0.905 and 0.907 on the H1N1 and seasonal vaccine datasets, respectively. Using an ensemble approach like the hard voting classifier enhances prediction accuracy and robustness, ultimately leading to better decision making in public health initiatives.

**Keywords** – Classification, Voting Classifier, Preprocessing, Model Selection, Public Health, Performance Analysis, Vaccination Status, Accuracy, Linear Regression, Structured Data.

## I. INTRODUCTION

In October 2009, a vaccination against the H1N1 flu virus became widely accessible. During the latter part of 2009 and the early months of 2010, the USA conducted the National 2009 H1N1 Flu Investigation. In the above poll, members were asked if they had recently received the H1N1 and seasonal flu vaccines, as well as specific inquiries. These further inquiries centered the survey gathered information on participants' social, economic, and demographic characteristics, along with their perceptions of disease risks and vaccine efficacy, as well as behaviors geared toward preventing transmission [1]. Greater knowledge regarding how these traits are related to personal immunization habits will help guide future public health efforts.

Machine learning techniques are widely utilized for predicting in a wide range of areas, including healthcare and robotic vehicles [2]. Furthermore, machine learning techniques are crucial in natural language processing, robots, videos, pictures, and audio processing are only a few examples. ML methods use a consistent approach that is diametrically opposite to conventional programming syntax that employs expressions with conditions. [3]. In this study, we provide a cutting-edge method for categorizing an individual's vaccination status in relation to two critical vaccines: the H1N1 influenza vaccine and the seasonal influenza vaccine. Our approach uses cutting-edge machine learning algorithms to quickly and reliably categorize people according to their vaccination status. The prevention of the spread of infectious diseases and the protection of both individuals and communities have rendered vaccinations an indispensable component of contemporary public health initiatives [4].

The fundamental principle of public health is vaccination, which is essential in controlling the spread of infectious diseases and reducing the severity of illnesses. It is a powerful tool that safeguards both individuals and communities by establishing immunity against harmful pathogens [5]. The emergence of novel infectious diseases, such as H1N1 (commonly known as swine flu), and the yearly recurrence of seasonal influenza present continuous challenges for healthcare systems globally. The need to understand vaccine uptake patterns, the factors that affect elements vaccination decisions, includes the advancement of strategies to enhance vaccination rates has never been more critical.

In recent years, the intersection of healthcare data, advanced computational methods, and artificial intelligence has revolutionized the way we approach medical research and public health. Machine learning (ML) techniques have shown remarkable capabilities in analyzing complex datasets, uncovering patterns, and making predictions. Previous research has employed XGBoost is (extreme gradient boosting) was an intelligent A tree-based machine learning methodology utilizing a gradient-boosting framework to build models that are predictive of influenza vaccination acceptance [6] and pediatric immunizations [7]. Leveraging these techniques, particularly in the context of vaccine uptake, provides an unprecedented opportunity to gain deeper insights into the dynamics of vaccination coverage, identify vulnerable populations, and develop targeted interventions [8]. Because there is no efficient therapy for healing, efforts to reduce the prevalence and severity of vaccines are mostly focused on immunization [9,10]. Despite exceptional vaccine production and unparalleled levels of Because of the rapid pace of development and manufacture, the global danger posed by the influenza virus is far from across, particularly to those at greater risk [11]. It is a critical tool for transforming biological information into useful data, performing the highest-level clinical studies, and enhancing medicine. As previously stated, the demand for vaccine classification led to numerous breakthroughs in ML approaches [12].

This paper focuses on the classification of individuals' vaccination status for both the H1N1 vaccine and seasonal influenza vaccine. We aim to harness the power of ML algorithms to accurately predict whether individuals have received these essential vaccinations. By doing so, we aspire to contribute valuable knowledge that not only aids in understanding the factors that influence vaccine decisions but also aids public health authorities in formulating effective strategies to improve vaccine coverage [13]. It's encouraging to see that current research has used machine learning algorithms to forecast diseases and identify vaccines [14]. With its capacity to absorb and analyze massive volumes of complicated data to produce precise predictions and classifications, machine learning has in fact established itself as an effective tool in healthcare and medical research [15].

Common supervised machine learning techniques frequently utilized in healthcare and disease prediction include: Logistic Regression, Random Forest, Support Vector Machines, Gradient Boosting, Naive Bayes, and Ensemble Methods. Various Algorithms for directed learning have been developed also been employed various scenarios [16-18]. To address existing human health issues caused by vaccination, this study tries to anticipate people's proclivity to be immunized by vaccination so that a strategy for convincing people to become immunized can be developed.

Among following major models have been predicted for the future days.

- To determine the vaccination trend based on the survey report status for both H1N1 and seasonal influenza vaccines.
- To determine the correctness of vaccinations To analyze the increase in newly diagnosed cases of vaccination and region.
- To create a Classification Model and test them with preprocessed dataset.
- To construct an ensemble model based on individual algorithm scores during the ML algorithm phase
- To compare the selected models and finalize the robust classification model, precision, recall, F1 score, and accuracy metrics will be utilized.

In conclusion, the classification of an individual's vaccination status using machine learning algorithms holds substantial promise for enhancing healthcare management. With the increasing complexity of healthcare data and the pressing need for accurate disease prevention measures, the fusion of technology and public health becomes paramount. This study aims to bridge this gap by presenting an approach that has the potential to reshape how we ascertain and manage vaccination status in an era characterized by data-driven solutions.

The structure of this research study is as follows: Section 2 examines the research in this area, including a full assessment of several Artificial intelligence and collaborative methods. Section 3 describes the suggested methodology, which employs a hard voting classifier in conjunction with an ensemble ML algorithms Section 4 presents the outcomes and findings of the provided technique have been compared and evaluated against conventional machine learning algorithms. and cutting-edge methods after the suggested methodology's examination.

## II. REVIEW OF LITERATURE

Shmueli et al. [2021] [19] In the research paper entitled "Predicting Intention to Receive COVID-19 Vaccine Among the General Population Using the Health Belief Model and the Theory of Planned Behavior Model", the research focuses on understanding the variables impacting people's willingness using both the Theory of Planned Behaviour Model and the Health Belief Model to agree to receive the COVID-19 vaccine are employed to assess various psychological and behavioral determinants. The research explores factors including perceived vulnerability to COVID-19, perceived seriousness of the illness, perceived advantages of vaccination, and perceived obstacles. attitudes, subjective norms, and perceived behavioral control. By analyzing survey data from the general population, the study offers insights into the predictive power of these models in anticipating vaccine acceptance. The findings contribute to the knowledge of vaccine hesitancy during the pandemic and provide valuable implications for public health communication strategies to enhance vaccination rates.

Queena Cheong et al. [2021] [20] Study advances the field of public health by leveraging machine learning to forecast vaccination uptake. The insights gained from their work have implications for designing effective interventions that encourage vaccine adoption and contribute to achieving higher immunization coverage rates across diverse communities in the United States. The research not only provides a comprehensive analysis of vaccination trends but also highlights the versatility of machine learning in addressing complex public health challenges. The findings hold significance beyond the specific case of COVID-19, offering a framework for predicting and improving vaccination rates for other preventable diseases as well.

Saloni Kumari et al. [2021][21] The research by Saloni Kumari et al. titled "An Ensemble Approach for Classification and Prediction of Diabetes Mellitus Using Soft Voting Classifier" presents a significant contribution to the realm of healthcare and medical forecasting. The research focuses on developing an ensemble methodology to enhance the classification and forecasting of diabetes mellitus. Diabetes is a prevalent chronic condition with substantial health implications, and accurate prediction plays a crucial role in early intervention and management. The resulting model's performance is likely evaluated through metrics like accuracy, precision, recall, and F1-score, providing insights into its effectiveness in predicting diabetes.

Zaidi, S.A.J et al. [2021][22] his study aimed to categorize public feelings regarding COVID-19 vaccination patterns. It's critical to remember that a sizable proportion of individuals received vaccinations, which led to some reported side effects; however, most recipients observed positive outcomes. Numerous research efforts as mining data, information technology, semi-supervised training, and robotics employing tools like WEKA, to forecast COVID-19 vaccination trends. Through a series of experiments, the study revealed that while vaccination could potentially induce minor side effects, the probability of such occurrences was remarkably low and rare. The public held diverse perspectives on vaccination. This research prognosticated the classification of trends linked to COVID-19 vaccines. The dataset was trained across five distinct classifiers to predict trends, evaluated using criteria including area under the curve, ability to recall information, F1 score, and logarithmic loss. Ultimately, a voting classifier amalgamated the outcomes of all classifiers, culminating in a comprehensive accuracy assessment.

Gyebi et al. [2023][23] study compared the effectiveness of various machine learning classifiers in order to predict measles patients. The usefulness of several algorithms in identifying measles cases was assessed by the researchers using a comparative methodology. They probably used a dataset with pertinent attributes and cases that were labeled to represent individuals who had the disease and those who did not. The findings of this study will probably help identify the machine learning classifier(s) that predicted measles with the greatest accuracy, precision, recall, or F1score. This comparison study may help in the selection of suitable algorithms for comparable disease prediction tasks by providing insightful information on how machine learning can be used in real-world healthcare settings.

In a study conducted by Bashir, Qamar, Khan, and Javed (2017), an ensemble model comprising CART, ID3, and C4.5 achieved an accuracy of 76.5%. According to studies, the random forest technique works better than previous algorithms (Soltani & Jafarian, 2016). AdaBoost and machine learning-based reduction strategies, with J48 as the cornerstone, play a vital role in diabetes prediction was covered by the author in 2017. Based on diabetes risk variables, it correctly distinguishes diabetic and non-diabetic people. According to M. Fatima and Pasha (2017), the J48 algorithm and bagging are outperformed by the AdaBoost learning technique.

## III. METHODOLOGY

The goal of the current study aimed to increase the precision of individual vaccinations, particularly for H1N1 and seasonal vaccines. The authors proposed a binary classification approach, categorizing diseases as positive or negative, and made use of a Hard voting classifier to execute a number of machine learning methods. Data preparation and data enrichment were carried out prior to data entry into the computational framework. The suggested technique produced greater results when compared to existing methods, with the best top three models for the defined dataset. This study aimed to establish a systematic tool for predicting public trends towards vaccination.

The proposed model's objective has been used to determine individual vaccinations status particularly for H1N1 and seasonal vaccines. The recommended ensemble method with hard voting classifier structure diagram is explained in **Fig 1**. The present study makes use of a broad variety of classification algorithms as the foundation for the Voting Classifier, which combines their predictions to make a collective conclusion. The methodology Collecting information, preliminary processing, and designing features make up the three steps, in which relevant qualities are chosen to influence the model's predictive capability. Following that, a variety of separate classifiers, trained on historical vaccination data, learn connections between numerous demographic and healthcare-related parameters and vaccination outcomes. The importance of this study comes from its ability to aid in informed decision-making in public health measures. Accurate estimates of vaccination status on an individual level enable focused interventions to increase immunization rates, reduce disease spread, and improve general community well-being.

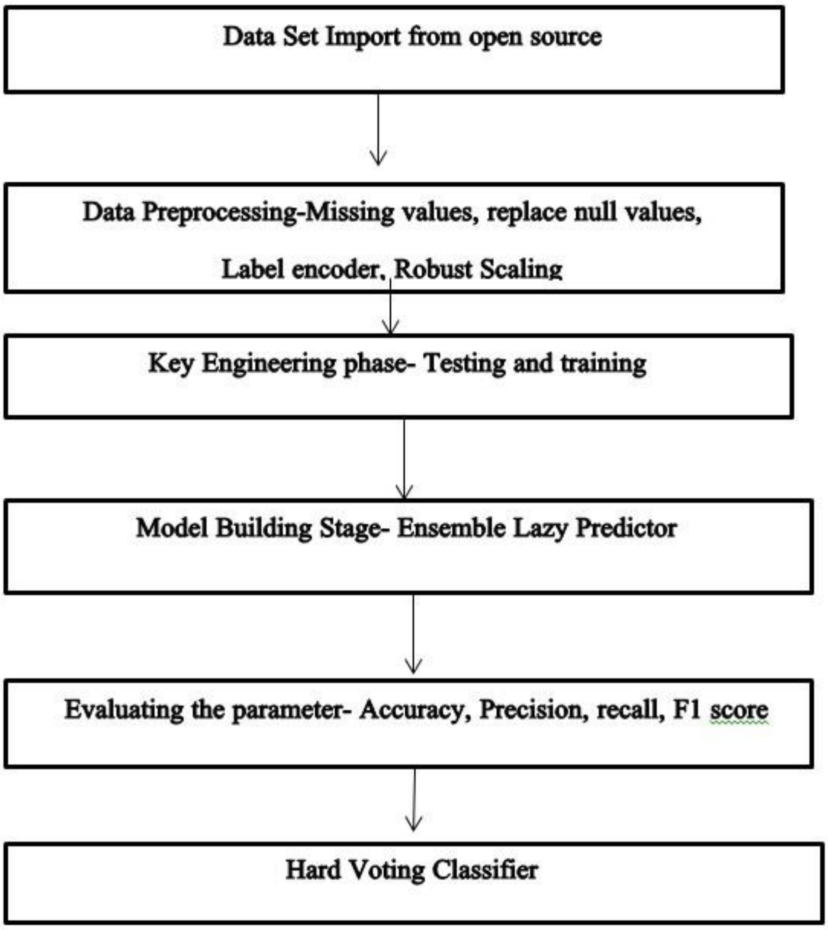


**Fig 1.** Block Diagram of Proposed Method Architecture.

*Data Set*

The purpose of this research project is to determine the public's views about seasonal and H1N1 immunization. The dataset employed in the present research for this objective depended on feedback that users gathered from Kaggle and supplied information on the total number of vaccinations given along with the total number of recoveries. Consequently, the database includes survey information concerning the vaccine. It initially utilized a comma-separated values (CSV) document. A Pandas Data Frame was loaded from a CSV file containing fields and data for the sector report. The report was generated using the Pandas Profiling tool. The information about vaccinations used to assess their efficacy is contained within a single CSV file of various classification algorithms in confirming whether individuals had received the vaccine or not, In different countries and number of recoveries. **Fig 2** shows Original Data Set Report from Kaggle. **Fig 3** shows Overview of Preprocessed Data set Report.

| respondent_id | h1n1_concern | h1n1_knowledge | behavioral_antiviral_meds | behavioral_avoidance | behavioral_face_mask | behavioral_wash_hands | behavioral_large_gatherings | behavioral_outside_home | behavioral_touch_face | rent_or_own | employment_status | hhs_geo_region | census_msa | household_adults | household_children | employment_industry | employment_occupation | h1n1_vaccine | seasonal_vaccine |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | Own | Not in Labor Force | oxchjgsf | Non-MSA | 0 | 0 | | | 0 | 0 |
| 1 | 3 | 2 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | Rent | Employed | bhuqouqj | Principle City | 0 | 0 | pxcmvdjn | xgwztkwe | 0 | 1 |
| 2 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | Own | Employed | qufhixun | Principle | 2 | 0 | rucpziij | xtkaffoo | 0 | 0 |
| 3 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | Rent | Not in Labor | lrircsnp | MSA, Principle | 0 | 0 | | | 0 | 1 |
| 4 | 2 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | Own | Employed | qufhixun | MSA, Not Principle | 1 | 0 | wxleyezf | emcorrxb | 0 | 0 |

**Fig 2.** Original Data Set Report from Kaggle.

| Dataset statistics | |
|---|---|
| Number of variables | 38 |
| Number of observations | 26707 |
| Missing cells | 60762 |
| Missing cells (%) | 6.0% |
| Duplicate rows | 0 |
| Duplicate rows (%) | 0.0% |
| Total size in memory | 7.7 MiB |
| Average record size in memory | 304.0 B |

| Variable types | |
|---|---|
| Numeric | 1 |
| Categorical | 37 |

**Fig 3.** Overview of Preprocessed Data set Report.

*Hard Voting Classifier for The Proposed Ensemble*

A meta-classifier, this classifier for combining Machine learning models that share conceptual similarity or dissimilarity can be combined for prediction using a majority vote.

A voting classifier employs both hard and soft voting techniques. With hard voting, the class prediction that is most commonly observed among the base models garners the majority of the votes, while that category prediction becomes the final prediction made. Voting classifier outperforms other Foundational models in terms of overall performance. On a separate testing dataset, the HARD Voting Classifier's performance is evaluated using measures like as accuracy, precision, recall, F1-score, and confusion matrix. These measures show how well the ensemble predicts a person's vaccination status.

Finally, the suggested HARD Voting Classifier is a robust ensemble technique that predicts an individual's vaccination status by leveraging the collective decisions of varied base classifiers. The ensemble improves accuracy and stability by aggregating forecasts by majority voting, making it a helpful tool for accurate vaccination status classification and public health decision making. **Fig 4** shows Model Ensemble Hard Voting Classifier.

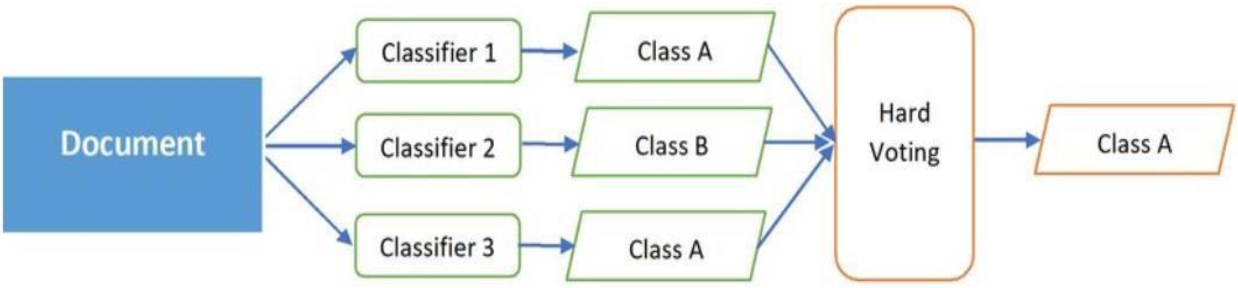


**Fig 4.** Model Ensemble Hard Voting Classifier.

The hard voting classifier works by counting the votes for each class label from the individual classifiers and selecting the class label with the highest vote count. This can be summarized in the following formula:

$$y=\mathbf{argmax}cj\in C\sum i=\mathbf{1}nI(\boldsymbol{hi}(x)=cj) \tag{1}$$

where:

- $y$ is the final predicted class.
- $CC$ is the set of all possible class labels.
- $nn$ is the number of base classifiers.
- $hi(x)hi(x)$ is the prediction of the $ii$-th classifier for input $xx$.
- $I(\cdot)I(\cdot)$ is the indicator function.

The core of the hard voting classifier in an ensemble learning setting is captured in this formula.

*Working of Ensemble Model*

Python programming was used to complete the preprocessing. Data processing is a critical step that converts the data into a format that is useful and efficient. so that the machine learning algorithm can use it. This method is used to conduct data that has linear movement transforms. The attribute values for this process, which is also known as Min-Max normalization, vary from [0,1]. So all the output by substituting 0 and 1 for the string values in the output variable, class is established. The dataset contained a large number of missing values for different attributes. The median for each attribute was used to fill in each blank value. The term "replacement by median" is also used to describe this data-preprocessing method. Upon addition of data from the Data Frame, the data underwent preprocessing to identify problematic data and handle missing values. Problematic data were removed from the data source, while missing data in certain columns were replaced with "0" in one context and "1" in another context. In this scenario, "0" denoted "TRUE" or "FALSE," while "1" represented ratings in the column. After this assignment, any remaining null values were replaced. In cases where null values couldn't be replaced, the respective columns were also dropped from the data. Ultimately, no data was missing in the columns. Multiple data access was performed using the correlation heat map. Text-related columns, along with their values, were checked and converted to numerical format. In this process. Label encoding is the next pre-processing method employed. The dependent variable is subjected to this procedure. Utilising the processed data Additionally, study conducted sophisticated exploratory data analyses (EDAs) to gain valuable insights. **Fig 5** shows Heat Map Correlation Matrix.

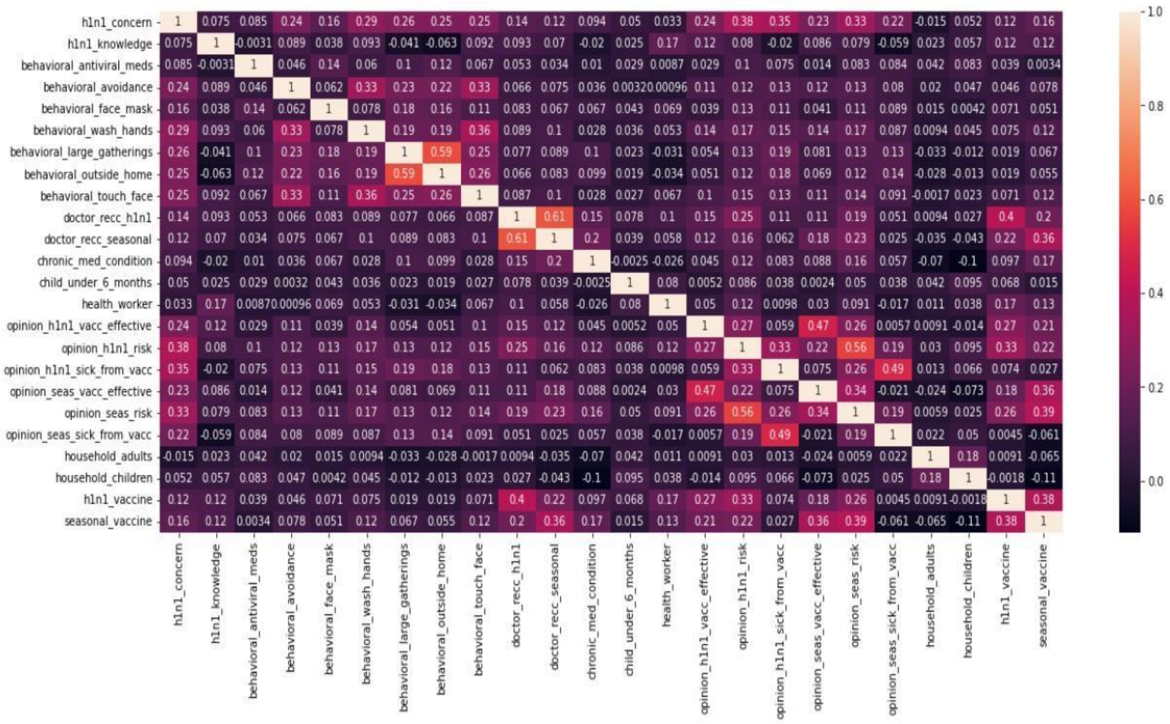


**Fig 5.** Heat Map Correlation Matrix.

*Algorithm*

Step 1: Dataset Prepared and divide it into test and training sets.

(Testing phase: 30%; training phase: 70%)

Step 2: The base classifiers defined (in this case, Decision Tree, Logistic Regression, and Support Vector Machine or models).

Step 3: Create the Voting Classifier using the list of tuples and specify voting='hard'.

**voting_classifier = VotingClassifier(estimators=classifiers, voting='hard')**

Step 4: Train the Voting Classifier on the training data.

Step 5: Utilize the trained Voting Classifier to make predictions on the testing dataset.

Step 6: Compute and display the accuracy of the ensemble classifier.

**accuracy = accuracy_score(y_test, predictions)**

Step 7: Assess the model's performance by computing metrics such as accuracy, F1 score, recall, sensitivity, and specificity.

## IV. RESULT AND DISCUSSION

The ensemble hard voting classifier using H1N1 and seasonal vaccine the dataset has the highest Accuracy, Precision, F1 score, and Recall. in all the three models in comparison to other machine learning methods and the final research was concluded with the Confusion matrix heat map of various models.

*Final Data*

The aim of the study was to determine the probability of individuals will get their H1N1 and seasonal flu shots. Specifically, this prediction aimed to provide two chances: one for the seasonal vaccine and one for the h1n1 vaccine. The dataset, and its processed state, was subsequently partitioned into two segments: the testing set (comprising 30% o) and the training set (encompassing 70%). Robust scaling was utilized as a standard scaler. The ensemble strategy that was suggested used a hard voting classifier.

*Model Training Phase*

The primary evaluation criteria utilized to assess the algorithms' strength and effectiveness were accuracy, precision, recall, and F1 score. The expected and actual class values in true positive (tp) situations were both 1. Instances where both the anticipated and actual class values were 0 were denoted as true negative (TN) cases, while false negatives (FN) and false positives (FP) arose when the predicted class contradicted the actual class.

Accuracy, as a key metric, indicated the proportion of properly predicted observed to total amount of data. The formulas for precision, recall, and F1 score were utilized to compute these metrics. The Lazy Classifier was used to predict the optimal model among multiple datasets within the provided dataset. Utilizing the Lazy Classifier, models were trained and tested. After selecting the top three models through a voting classifier, they were combined.

*Testing Matrix*

Examining a model's accuracy and efficiency is known as testing it. Each trained model in this study was assessed and put to the test using a variety of parameters, including Precision, recall, F-measure, logarithmic loss (LL), and area under the curve (AUC) are all metrics to consider. All of the outcomes from the various classifiers were combined at the conclusion of the study to forecast accuracy using the voting classifier.

$$Precision = \text{tp / tp+fp} \tag{2}$$

$$Accuracy = \frac{tp + tn}{tn + tp + \text{fp} + \text{fn}} \tag{3}$$

$$Recall = tp\,/tp + fn \tag{4}$$

$$\text{F1score} = \frac{2 \times P\,recision \times recall}{P\,recision + recall} \tag{5}$$

*H1N1 Vaccine*

In the context of the H1N1 vaccine, the Lazy Predictor was utilized to predict the top three models. A **Table1** displays all the best models, showcasing their respective maximum accuracy, precision, F1 score, and recall. Among these metrics, the top three models were selected: Ensemble ,AdaBoostClassifier,LGBMClassifier , LogisticRegression for further prediction of confusion Heat matrix .

The confusion matrix for the H1N1 vaccination model was constructed solely from the accurate and inaccurate predictions of the proposed ensemble hard voting classifier, aligned with the true labels and predicted labels. Within the H1N1 vaccination model, there were 1559 actual positives and 1293 false positives. **Fig 6** shows Confusion Matrix of Voting Classifier H1N1 Vaccine**.**

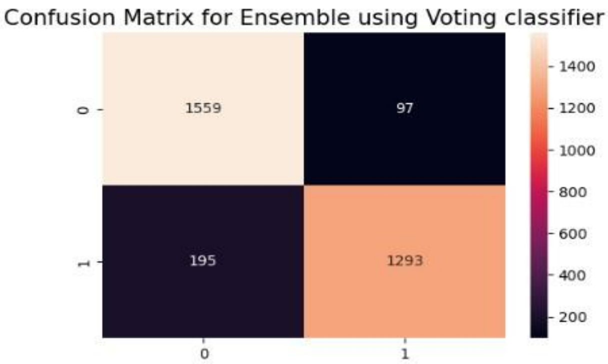


**Fig 6.** Confusion Matrix of Voting Classifier H1N1 Vaccine**.**

**Table 1.** Comparison of Machine Learning Models H1N1 Vaccine

| H1N1 vaccine | Accuracy | Precision | Recall | f1 | Sensitivity | Specificity |
|---|---|---|---|---|---|---|
| Logistic Regression | 0.8197 | 0.7949 | 0.8647 | 0.8283 | 0.8647 | 0.7740 |
| LGBM classifier | 0.8394 | 0.8138 | 0.8841 | 0.8475 | 0.8841 | 0.7938 |
| AdaBoost classifier | 0.8874 | 0.8751 | 0.9174 | 0.8958 | 0.9174 | 0.8540 |
| Ensemble | 0.9071 | 0.8888 | 0.9414 | 0.9144 | 0.9414 | 0.8690 |

**Fig 7** Demonstrate the time efficiency of the top four ML models, including Ensemble, AdaBoostClassifier, LGBMClassifier, and Logistic Regression. These models exhibit notably low time consumption, with only 0.02 seconds required, which is significanty less compared to other leading ML classifications models.

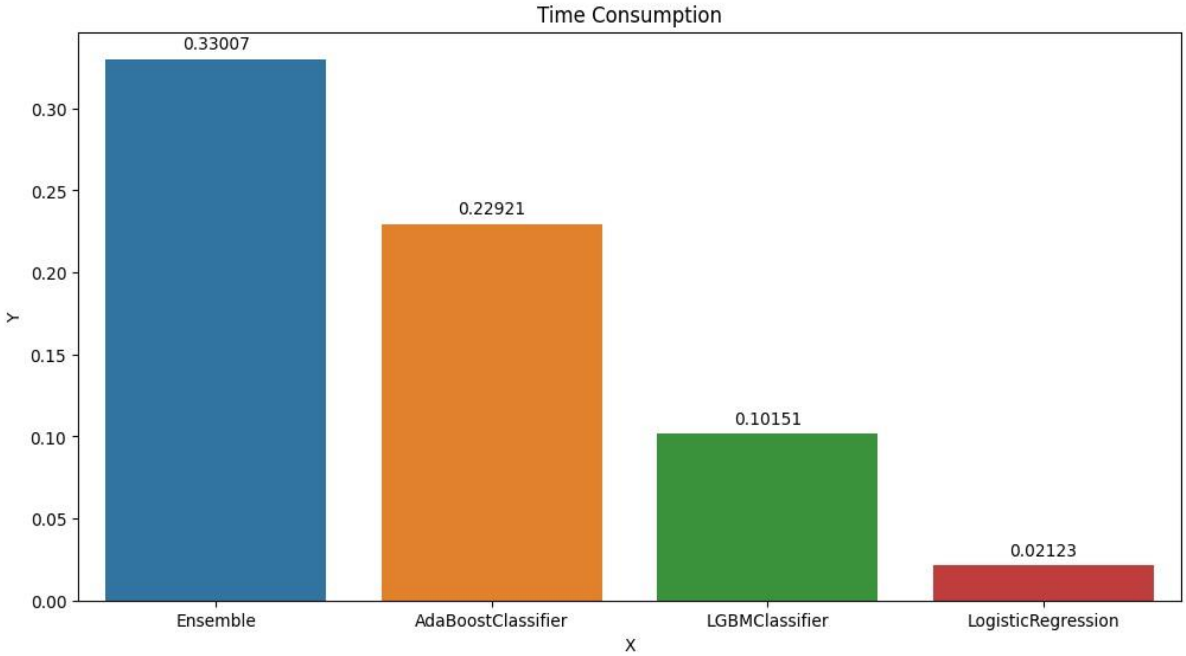


**Fig 7.** Time Complexity of Top 4 Models.

**Fig 8** illustrate the precision, recall,F1-Score, specificity and senstivity for top 4 models in Accurately identifying intruders relies on precision, which measures the proportion of correctly identified intruders among all instances identified as such. Ensemble, Adabooster, LGBM and Logical regression are 0.88,0.87,0.81,0.79 in all these model is considered as top 4 ML furthermore, the reason these models are considered for better detection lies in their association with accuracy, particularly the F1-Score, which is closely linked with accuracy. Among them, the Ensemble model showcased superior accuracy, precision, recall, and F1-Score compared to its counterparts.

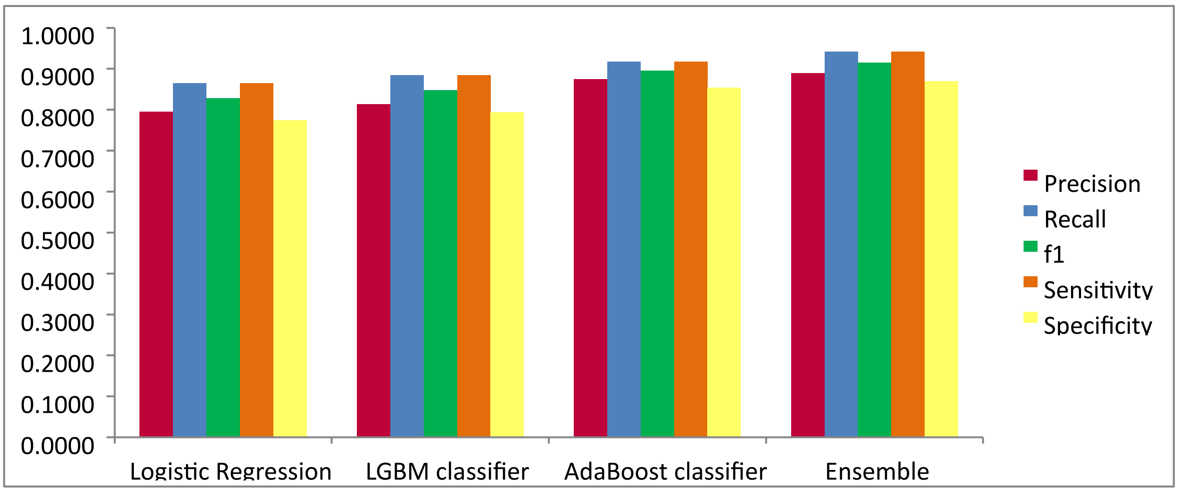


**Fig 8.** Graphical Analysis of Precision, Recall, F1 Support, Specificity and Sensitivity for Top 4 Models in Class 0**.**

**Fig 9** involves the top four accuracy-based ML models, which are Ensemble, Adabooster, LGBM, and Logistic Regression with accuracies of 0.90, 0.88, 0.83, and 0.81 respectively, it is evident that the Ensemble model achieves the highest accuracy of 0.90 for the H1N1 vaccine. Comparatively, this accuracy surpasses that of the top four ML models. Therefore, the final accuracy obtained for the H1N1 vaccine using the Ensemble model stands at 0.90.

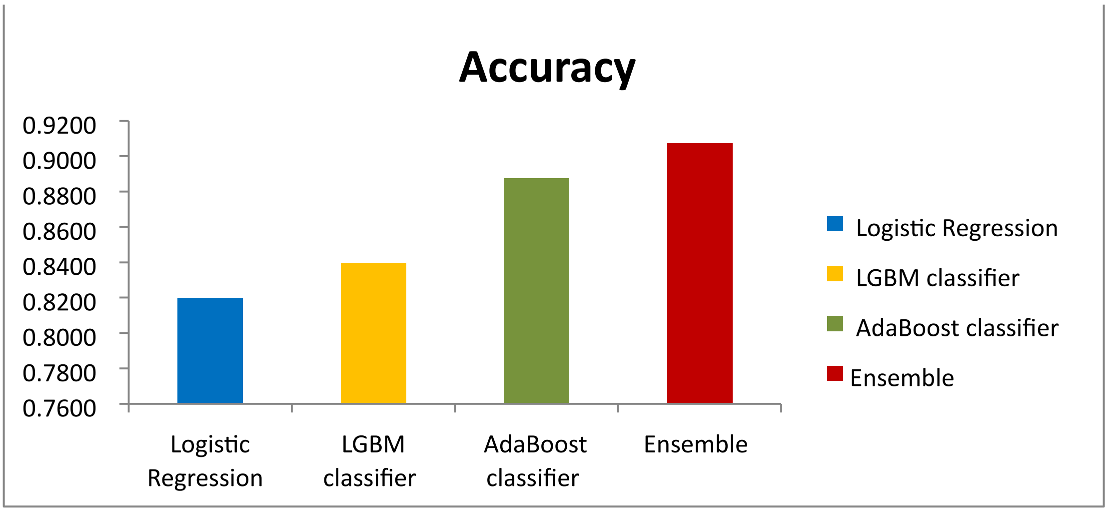


**Fig 9.** Comparative Analysis of Top 4 Models' Accuracy.

*Seasonal Vaccine*

In the context of the seasonal vaccine, the Lazy Predictor was utilized to predict the top three models. A **Table 2** displays all the best models, showcasing their respective maximum accuracy, precision, F1 score, and recall. Among these metrics, the top three models were selected:

LGBMClassifier, AdaBoostClassifier, and RandomForestClassifier.

In the Seasonal Vaccine dataset, there were 3794 true positives and 2821 true negatives in the confusion matrix. We now have a better understanding of both our forecasts and the results of our research efforts thanks to the confusion matrix. **Fig 10** shows Confusion Matrix of Voting Classifier Seasonal Vaccine.

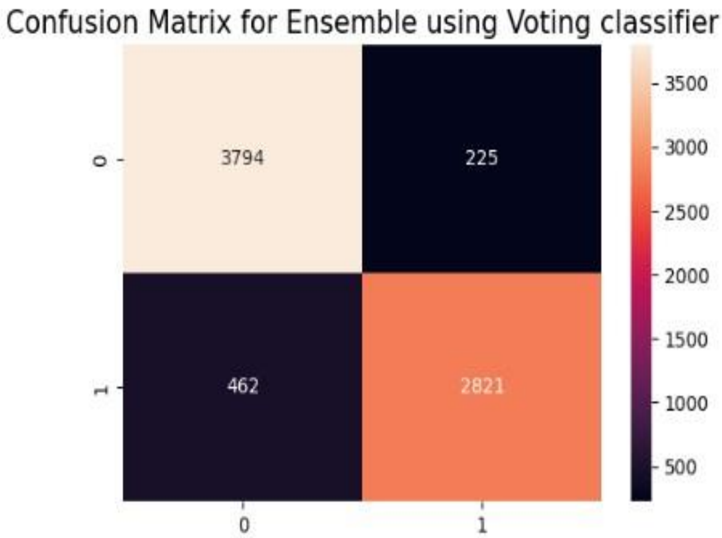


**Fig 10.** Confusion Matrix of Voting Classifier Seasonal Vaccine.

**Table 2**. Comparison of Machine learning models Seasonal Vaccine.

| Seasonal vaccine | Accuracy | Precision | Recall | f1 | Sensitivity | Specificity |
|---|---|---|---|---|---|---|
| Random forest classifier | 0.8435 | 0.8247 | 0.8966 | 0.8591 | 0.8966 | 0.7830 |
| AdaBoost classifier | 0.8707 | 0.8626 | 0.9095 | 0.8855 | 0.9095 | 0.8234 |
| LGBM classifier | 0.8974 | 0.8840 | 0.9334 | 0.9080 | 0.9334 | 0.8548 |
| Ensemble | 0.9059 | 0.8914 | 0.9440 | 0.9170 | 0.9440 | 0.8593 |

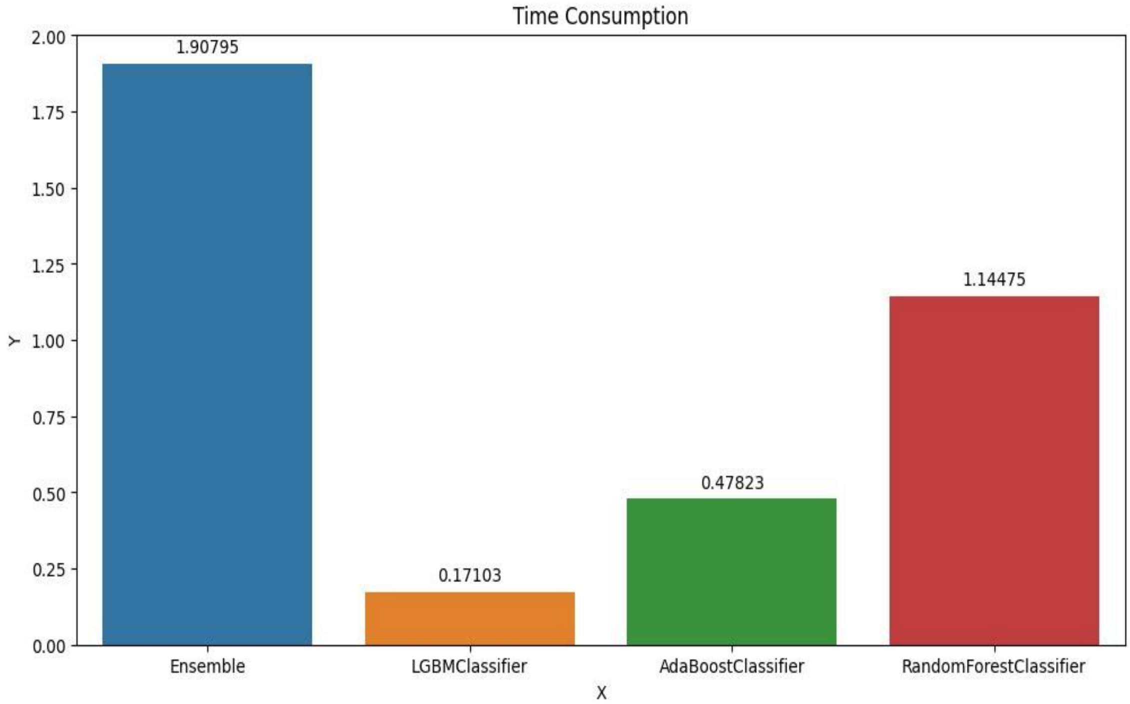


**Fig 11**. Time Complexity of Top 4 Models.

**Fig 11** The time efficiency of the top four ML models for the seasonal vaccine, including Ensemble, Adaboost Classifier, LGBM Classifier, and Random Forest Classifier, is noteworthy, with a consumption of only 0.17 seconds. This time consumption is notably lower when compared to other leading ML classification models.

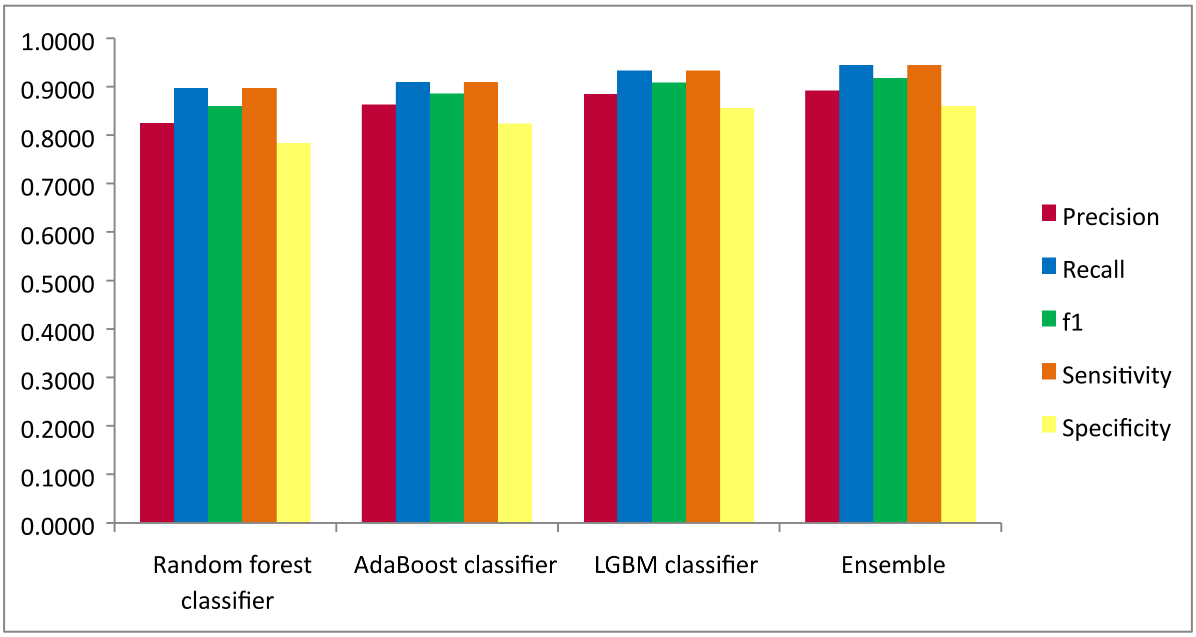


**Fig 12.** Graphical Analysis of Precision, Recall and F1 Support for Top 4 Models in Class0

**Fig 12** presents the precision, recall, F1-Score, sensitivity, and specificity for the top 4 models, which were accurately identified through precision. The precision values for Ensemble, Adabooster, Random Forest Classifier, and LGBM were 0.89, 0.86, 0.82, and 0.88 respectively, all of which positioned them as the top 4 ML models, prompting their selection for further analysis. F1-Score, being directly associated with accuracy, played a crucial role in the selection of these models. Notably, the Ensemble model demonstrated superior accuracy, precision, recall, and F1 score compared to the other models.

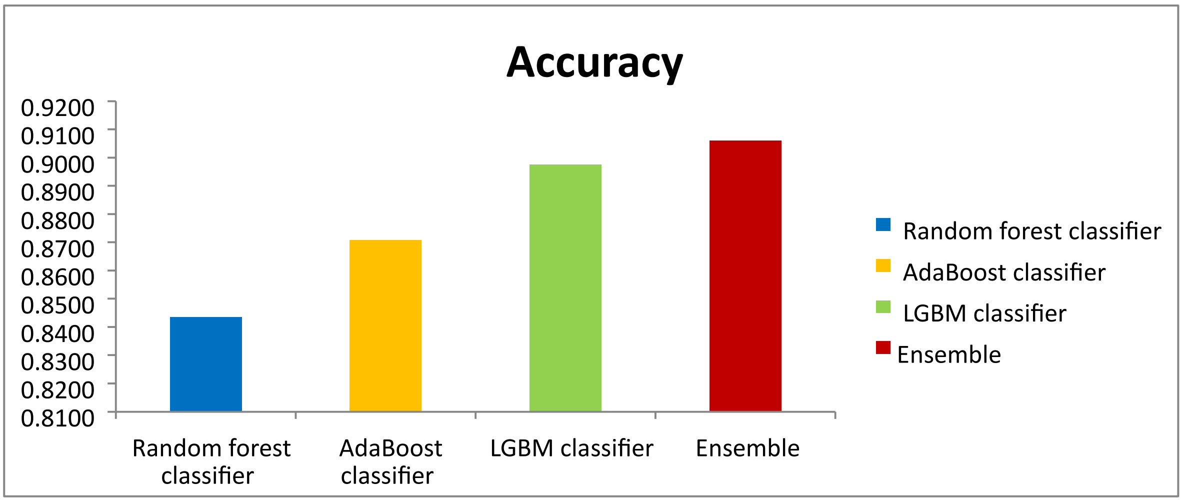


**Fig 13.** Comparative Analysis of Top 4 Models' Accuracy.

In **Fig 13**, the analysis is restricted to the top four accuracy-based ML models, namely Ensemble, Adabooster, LGBM, and Random, with accuracies of 0.90, 0.87, 0.89, and 0.84 respectively. Upon comparing these top models for the seasonal vaccine, it is evident that the Ensemble model achieves the highest accuracy of 0.90. Therefore, the final accuracy obtained for the seasonal vaccine using the Ensemble model stands at 0.90.

## V. CONCLUSION AND FUTURE WORK

The results of the research appeared more encouraging and inspiring for the random dataset. It was clear from examining the overall results using the data that was accessible that this was a favorable trend in people's attitudes toward vaccination. Additionally, the suggested system could have been applied to different datasets to predict their impact in various challenging fields. An ensemble hard voting classifier model based on a variety of machine learning approaches was proposed. Following that, the three best-performing models had been applied to both the H1N1 and seasonal vaccine datasets. The data gathered observations showed the greatest accuracy, precision, F1 score, and recall for both vaccines. Looking forward, aimed to improve this work by incorporating more efficient and effective techniques, advancing with hyper parameter tuning for ensemble model.

### Data Availability

No data was used to support this study.

### Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

### Funding

No funding agency is associated with this research.

### Competing Interests

There are no competing interests.

### References

[1]. H.-D. Emborg, T. G. Krause, A. Hviid, J. Simonsen, and K. Molbak, "Effectiveness of vaccine against pandemic influenza A/H1N1 among people with underlying chronic diseases: cohort study, Denmark, 2009-10," BMJ, vol. 344, no. jan25 4, pp. d7901–d7901, Jan. 2011, doi: 10.1136/bmj.d7901.
[2]. J. M. Zhang, M. Harman, L. Ma, and Y. Liu, "Machine Learning Testing: Survey, Landscapes and Horizons," IEEE Transactions on Software Engineering, vol. 48, no. 1, pp. 1–36, Jan. 2022, doi: 10.1109/tse.2019.2962027.
[3]. G. Bontempi, S. Ben Taieb, and Y.-A. Le Borgne, "Machine Learning Strategies for Time Series Forecasting," Business Intelligence, pp. 62–77, 2013, doi: 10.1007/978-3-642-36318-4_3.
[4]. M. M. Rahman, F. Khatun, A. Uzzaman, S. I. Sami, M. A.-A. Bhuiyan, and T. S. Kiong, "A Comprehensive Study of Artificial Intelligence and Machine Learning Approaches in Confronting the Coronavirus (COVID-19) Pandemic," International Journal of Health Services, vol. 51, no. 4, pp. 446–461, May 2021, doi: 10.1177/00207314211017469.
[5]. J. Griffith, H. Marani, and H. Monkman, "COVID-19 Vaccine Hesitancy in Canada: Content Analysis of Tweets Using the Theoretical Domains Framework," Journal of Medical Internet Research, vol. 23, no. 4, p. e26874, Apr. 2021, doi: 10.2196/26874.
[6]. A. Shaham, G. Chodick, V. Shalev, and D. Yamin, "Personal and social patterns predict influenza vaccination decision," BMC Public Health, vol. 20, no. 1, Feb. 2020, doi: 10.1186/s12889-020-8327-3.
[7]. Haldorai, B. L. R, S. Murugan, and M. Balakrishnan, "Hemorrhage Detection from Whole-Body CT Images Using Deep Learning," EAI/Springer Innovations in Communication and Computing, pp. 139–151, 2024, doi: 10.1007/978-3-031-53972-5_7.

[8]. B. Bravi, “Development and use of machine learning algorithms in vaccine target selection,” npj Vaccines, vol. 9, no. 1, Jan. 2024, doi: 10.1038/s41541-023-00795-8.
[9]. B. Arifin and T. Anas, “Lessons learned from COVID-19 vaccination in Indonesia: experiences, challenges, and opportunities,” Human Vaccines & Immunotherapeutics, vol. 17, no. 11, pp. 3898–3906, Oct. 2021, doi: 10.1080/21645515.2021.1975450.
[10]. S. Koesnoe et al., “Using Integrative Behavior Model to Predict COVID-19 Vaccination Intention among Health Care Workers in Indonesia: A Nationwide Survey,” Vaccines, vol. 10, no. 5, p. 719, May 2022, doi: 10.3390/vaccines10050719.
[11]. T. Ching et al., “Opportunities and obstacles for deep learning in biology and medicine,” Journal of The Royal Society Interface, vol. 15, no. 141, p. 20170387, Apr. 2018, doi: 10.1098/rsif.2017.0387.
[12]. C. Magazzino, M. Mele, and M. Coccia, “A machine learning algorithm to analyse the effects of vaccination on COVID-19 mortality,” Epidemiology and Infection, vol. 150, 2022, doi: 10.1017/s0950268822001418.
[13]. J. Samuel, G. G. Md. N. Ali, Md. M. Rahman, E. Esawi, and Y. Samuel, “COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification,” Information, vol. 11, no. 6, p. 314, Jun. 2020, doi: 10.3390/info11060314.
[14]. S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “Statistical and Machine Learning forecasting methods: Concerns and ways forward,” PLOS ONE, vol. 13, no. 3, p. e0194889, Mar. 2018, doi: 10.1371/journal.pone.0194889.
[15]. P. M. Dhulavvagol, S. G. Totad, P. Pratheek, R. Ostwal, S. Sudhanshu, and M. Y. Veerabhadra, “An Efficient Ensemble Based Model for Data Classification,” 2022 IEEE 7th International conference for Convergence in Technology (I2CT), vol. 545, pp. 1–5, Apr. 2022, doi: 10.1109/i2ct54291.2022.9824722.
[16]. C. N. Villavicencio, J. J. E. Macrohon, X. A. Inbaraj, J.-H. Jeng, and J.-G. Hsieh, “COVID-19 Prediction Applying Supervised Machine Learning Algorithms with Comparative Analysis Using WEKA,” Algorithms, vol. 14, no. 7, p. 201, Jun. 2021, doi: 10.3390/a14070201.
[17]. A. A. Hussain, O. Bouachir, F. Al-Turjman, and M. Aloqaily, “Notice of Retraction: AI Techniques for COVID-19,” IEEE Access, vol. 8, pp. 128776–128795, 2020, doi: 10.1109/access.2020.3007939.
[18]. R. Nistal, M. de la Sen, J. Gabirondo, S. Alonso-Quesada, A. J. Garrido, and I. Garrido, “A Study on COVID-19 Incidence in Europe through Two SEIR Epidemic Models Which Consider Mixed Contagions from Asymptomatic and Symptomatic Individuals,” Applied Sciences, vol. 11, no. 14, p. 6266, Jul. 2021, doi: 10.3390/app11146266.
[19]. L. Shmueli, “Predicting intention to receive COVID-19 vaccine among the general population using the health belief model and the theory of planned behavior model,” BMC Public Health, vol. 21, no. 1, Apr. 2021, doi: 10.1186/s12889-021-10816-7.
[20]. Q. Cheong, M. Au-yeung, S. Quon, K. Concepcion, and J. D. Kong, “Predictive Modeling of Vaccination Uptake in US Counties: A Machine Learning–Based Approach,” Journal of Medical Internet Research, vol. 23, no. 11, p. e33231, Nov. 2021, doi: 10.2196/33231.
[21]. S. Kumari, D. Kumar, and M. Mittal, “An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier,” International Journal of Cognitive Computing in Engineering, vol. 2, pp. 40–46, Jun. 2021, doi: 10.1016/j.ijcce.2021.01.001.
[22]. S. A. J. Zaidi, S. Tariq, and S. B. Belhaouari, “Future Prediction of COVID-19 Vaccine Trends Using a Voting Classifier,” Data, vol. 6, no. 11, p. 112, Nov. 2021, doi: 10.3390/data6110112.
[23]. R. Gyebi et al., “Prediction of measles patients using machine learning classifiers: a comparative study,” Bulletin of the National Research Centre, vol. 47, no. 1, Jul. 2023, doi: 10.1186/s42269-023-01079-w.