

FLU SHOT LEARNING: PREDICT H1N1 AND SEASONAL FLU VACCINES

Srividya Inampudi
Computer Science
Department
Fr. Conceicao
Rodrigues Institute of
Technology

Greshma Johnson
Computer Science and
Engineering
Saintgits College of
Engineering

Jay Jhaveri
Computer Science
Engineering
Vivekanand Education
Society's Institute of
Technology

Niranjan.S
Electronics and
Communication
Engineering
Sri Ramakrishna
Engineering College

Dr. Kuldeep Chaurasia
Assistant Professor
School of Engineering
and Applied Sciences
Bennett University

Abstract—The H1N1 Flu that came into existence in 2009 in the U.S. and spread to the rest of the world had a great impact on the lives of people around the world. It was a life-threatening season to hundreds of people mainly below 65 years old which eventually made the World Health Organization (WHO) to declare it as the greatest pandemic in more than 40 years. Since it continues to circulate seasonally worldwide vaccination should be carried out effectively. The vaccines for H1N1 were first publicly available in the United States in October 2009, when the United States government began a vaccination campaign and people took vaccination based on certain factors. So, to find out the vaccination status National 2009 H1N1 Flu Survey (NHFS) was conducted. In this paper, we use the data from the above survey to develop a model that predicts how likely people got H1N1 and seasonal flu vaccine.

I. INTRODUCTION

H1N1 or swine flu virus first emerged in 2009, spring season in Mexico and then in the United States and quickly spread across the globe. A distinctive combination or integration of influenza genes was discovered in this novel H1N1 virus which was not identified prior in humans or animals [1]. This contagious novel virus had a very powerful impact on the whole world and spread across the world like a forest fire and as a result on June 11 2009 the World health Organization (WHO) declared that a pandemic of 2009 H1N1 flu or swine flu had begun [2]. The effects of this novel H1N1 virus were more severe on people below the age of 65. There was significantly high pediatric mortality, and higher rate of hospitalizations for young adults and children, so the hospitals would soon get overwhelmed quickly as compared to non-pandemic influenza seasons [3].

According to CDC the first and foremost step in protecting oneself of this virus is a yearly flu vaccination [4]. There are various factors such as age, health perceptions of an individual and the similarities or “match” in the vaccine’s virus structure and the virus structure which is affecting the community which affects the ability of the vaccination to provide protection to the person who is vaccinated [5]. Several activities were performed using various social media platforms and broadcasting networks such as Twitter was used to track the levels of disease activity and the concern of the public towards this pandemic [6]. The social media played an important role to assess the sentiments towards vaccination. The implications for disease dynamics and control [7], etc. The popular among them is the phone survey conducted by the U.S. where they asked respondents whether they had received the H1N1 and seasonal flu vaccines, in conjunction with questions about themselves. In this study, we use the data obtained from the National 2009 H1N1 Flu Survey to predict how likely people got H1N1 and seasonal flu vaccines. Thus, we can hope for a better immunized society.

II. LITERATURE REVIEW

Mabrouk et al.[8] "A chaotic study on pandemic and classical (H1N1) using EIIP sequence indicators, "et.al: Described in their paper that the methods such as Moment invariants, correlation dimension, and largest Lyapunov exponent which were used to detect H1N1 and indicated the differences between the pandemic and classical influenza virus and Chinh et al. (2010)[9] In his paper: “A possible mutation that enables the H1N1 influenza A virus to escape antibody recognition” Explained the methods such as Phylogenetic Analysis of Pandemic Strains, Molecular docking for the predicted epitopes. And concluded that the B-cell epitope helps the virus to escape the antibody recognition, and gives the vaccine against the H1N1 influenza A virus and Huang et al. [10], In their research paper "Aptamer-modified CNTFET biosensor for detecting H1N1 virus in a droplet," suggested the

combination of APTS and SAM immersed in nanotube which gives CNTFET and thus it acts as a biosensor which is used in the detection of H1N1 virus by droplet.

M. S. Ünlü [11], In his famous paper "Optical interference for multiplexed, label-free, and dynamic biosensing: Protein, DNA and single virus detection," described Interferometric Reflectance Imaging Sensor which can be used for label-free, high throughput, high sensitivity and dynamic detection and gives detection of H1N1 virus and nanoparticles and Kamikawa et al. (2012) [12], In their journal paper Pandemic Influenza Detection by Electrically Active Magnetic Nanoparticles and Surface Plasmon Resonance. Indicated that the detection consists of several processes such as nanoparticle synthesis, glycans, polyaniline, and sensor modification by means to find H1N1 by nanoparticle and resonance and Jerald et al.[13], In his paper "Influenza virus vaccine efficacy based on conserved sequence alignment," spoke about the vital strain sequence used from NCBI and Sequence alignment which helps vaccine efficiency for influenza.

Chrysostomou, et al. (2013) [14], In their paper the "Signal-processing-based bioinformatics approach for the identification of influenza A virus subtypes in Neuraminidase genes" majorly spoke about the Neuraminidase genes, Signal Processing, F-score, Support Vector Machines are the methods used for the identification of influenza virus and Wiriyaichaiorn et al.[15], In their research paper "Rapid influenza an antigen detection using carbon nano string as the label for lateral flow immunochromatographic assay," presented a preparation of allantoic fluid infected with influenza A, Virus conjugation of CNS to antibody and about the Evaluation of CBNS-MAb using LFIA are used in the detection of nanoparticles and Ma et al.[16], In their paper "An integrated passive microfluidic device for rapid detection of influenza a (H1N1) virus by reverse transcription loop-mediated isothermal amplification (RT-LAMP)," demonstrated the loading of virus and magnetic beads, about virus capture, collection of virus-magnetic beads complexes, removal of excessive wastes, virus particle lysis, RT-LAMP reaction and the coloration are the steps to detect H1N1 virus.

Nieto-Chaupis, Huber. (2019)[17]. In their paper titled "Face To Face with Next Flu Pandemic with a Wiener-Series-Based Machine Learning: Fast Decisions to Tackle Rapid Spread" explained about the Wiener model used in order to increase optimization, efficiency and performance to find the spread of seasonal flu and Stalder et al.[18] In the famous paper "Tracking the flu pandemic by monitoring the social web" related the retrieving data from Twitter and Official health reports provides inexpensive and timely information about the epidemic and Motoyama et al.[19] in their paper "Predicting Flu Trends using Twitter Data" demonstrated the SNEFT model and twitter crawler are the methods used in Predicting the flu using twitter data.

Wong et al.[20] in their paper "Diagnosis of Response Behavioural Patterns Towards the Risk of Pandemic Flu Influenza A (H1N1) of Urban Community Based on Rasch Measurement Model" presented the source of data and data analysis methodology used for the response behavioral Patterns towards H1N1 and Bao et al.[21] in their infamous paper "Influenza-A Circulation in Vietnam through Data Analysis of Hemagglutinin Entries" provided NCBI influenza virus resources which provides many datasets (2001-2012) which is used for the analysis of influenza virus and Hu et al.[22], in their research paper "Computational Study of Interdependence Between Hemagglutinin and Neuraminidase of Pandemic 2009 H1N1" explained Sequence data and Informational Spectrum Model are used to study between Hemagglutinin and neuraminidase of H1N1.

III. DATA RESOURCES

Data is one of the most important and vital aspects of any research study. The data for our study is obtained from the National 2009 H1N1 Flu Survey (NHFS) which was held by the Centres for Disease Control and Prevention (CDC) the main aim of the survey was to monitor and evaluate H1N1 flu vaccination efforts among adults and children. The survey was conducted through telephones, twitter and with the help of various other electronic media in all the 50 states and the districts of Columbia. Various questions about flu related behaviors, opinions about flu vaccine's safety and effectiveness, medical history like recent respiratory illness and pneumococcal vaccination status were asked apart from the major question about H1N1 and seasonal flu vaccination status. This data was obtained to get a fair idea about the knowledge of people on the effectiveness and safety of flu vaccines and to learn why some people refrained from getting vaccinated against the H1N1 flu and seasonal flu. Huge amount of data was gathered through

this survey which is being commonly used for analysis and research purposes which also measures the number of children and adults nationwide who have received vaccinations. The NFS has been conducted since the 2010-11 influenza season. [10]

Here in our project and its implementation we are using the programming language Python in the Google Colab platform to develop the model for prediction. As python is a general-purpose language and Google Colab is a free online platform where using large datasets, we can train complex models on CPUs, GPUs, and TPUs we prefer to work using it. Python-based Deep Learning libraries like Tensorflow will be exploited for the development and experimentation of the project.

IV. METHODOLOGY

The main aim of our project is to find the probability that a person will receive H1N1 and seasonal Flu vaccination based on many parameters. The data obtained from the National 2009 H1N1 Flu Survey (NHFS) contains 3 CSV files namely the training set features, the training set labels, and the test set features. The data has been obtained from over 53000 people from which around 26000 observations have been considered for the training set and the rest have been considered for the testing set.

We have considered various methodologies and compared different Machine Learning and Artificial Neural Network models to predict the required probability. The Machine Learning algorithms such as Multiple Linear regression, Support Vector Regression, Random Forest Regression and Logistic Regression were used.

Artificial Neural Network (ANN) with different optimizers such as Adam, RMSprop, SGD were used to predict the probability of the test set features.

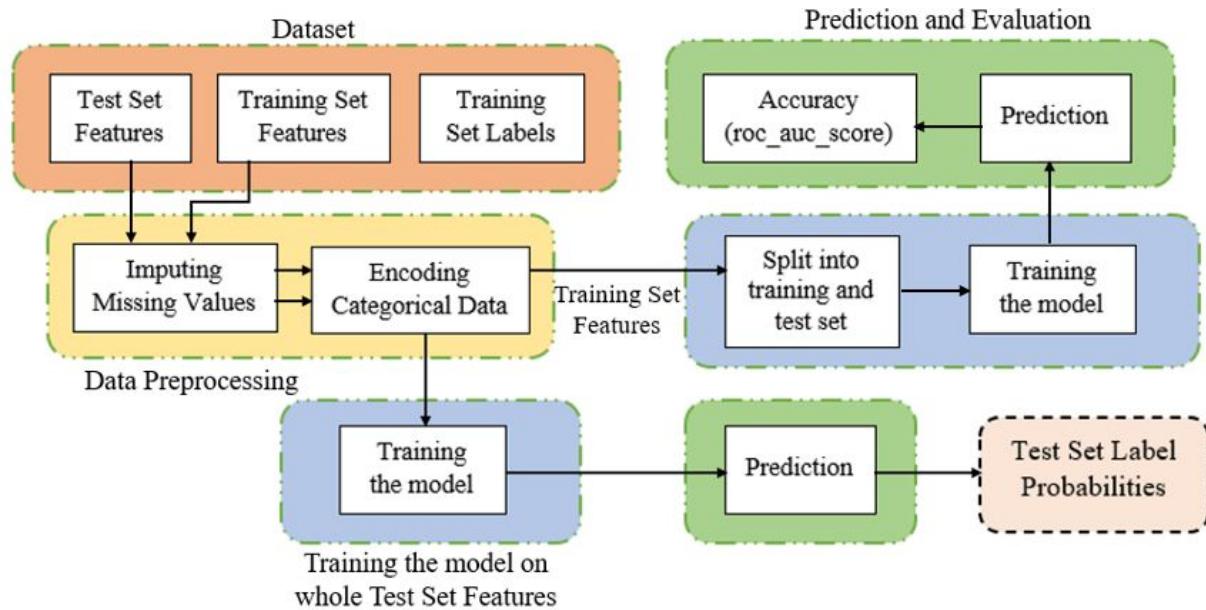


Fig. 1. System Architecture of Machine Learning Model

Taking Care of Missing Data:

The missing data values in the dataset were imputed by Univariate feature imputation using the most frequent strategy (statistics) with the help of the SimpleImputer class from the sklearn.impute module. This SimpleImputer class provides fundamental strategies for assigning some value for the missing values in the columns of the dataset [24][25].

Encoding Categorical Data:

For Encoding Categorical Data the columns are transformed separately and then the features generated during this transformation of columns are concatenated to a single feature space this process is carried out with the help of ColumnTransformer class of sklearn.compose module. Heterogeneous or columnar data is most benefited from this method as several feature extraction mechanisms can be combined together or it gets transformed into a one transformer [24]. Then The categorical data is encoded using the OneHotEncoder class of the sklearn.preprocessing module. The features are encoded using an encoding scheme. A binary column is created in this method for every category. A sparse matrix is returned or a dense array is returned depending on the sparse parameter [25].

Splitting the Dataset:

The training set features and training set labels have been split into training set (80%) and testing set (20%) using train_test_split from sklearn.model_selection. This library splits the dataset into training and testing sets.

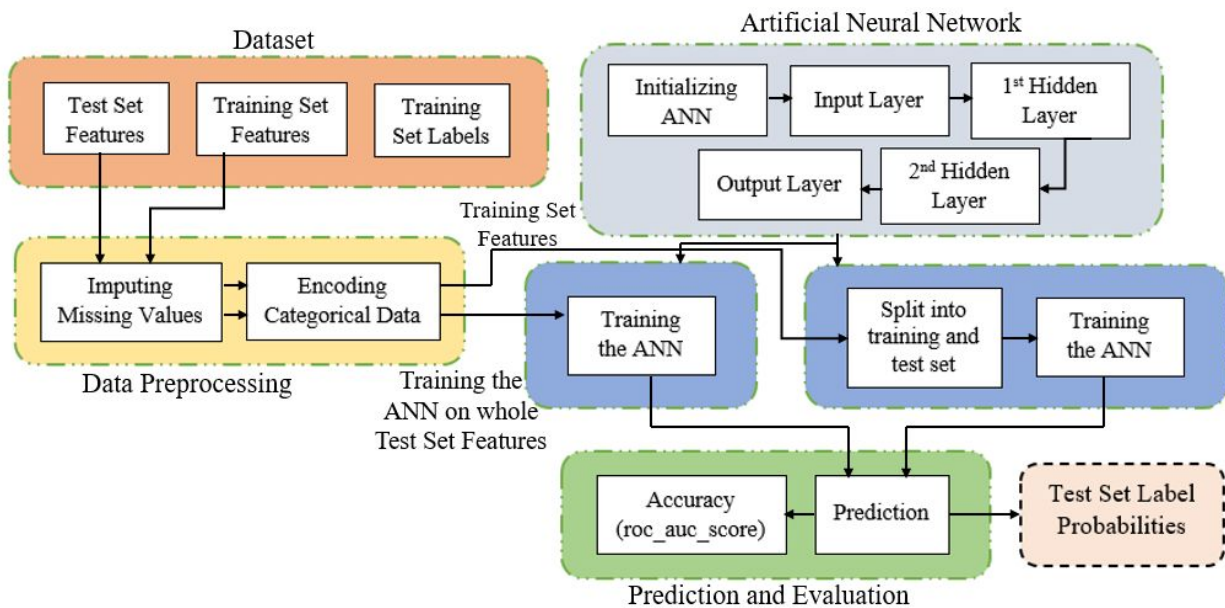


Fig. 2. System Architecture of Artificial Neural Network

Hyperparameter Tuning -

Hyperparameter Tuning is done to find the most optimal parameter for the model on which the model gives the best results. We have used various Hyperparameter tuning methods such as GridsearchCV, RandomSearchCV for our machine learning models to obtain better results. K fold cross Validation method has been used to tune hyperparameters for the Artificial Neural Network.

Training the Model on the Dataset:

An open-source Python library scikit-learn is used which using a unified interface implements many machine learning, preprocessing, cross-validation, and visualization algorithms [26]. Scikit-learn can be used for both supervised and unsupervised learning, using a consistent, task-oriented interface [27].

In scikit-learn all supervised estimators implement a fit(X,y) method to fit the model i.e. to train a model and a predict(X) is a method that gives predicted labels y for given unlabelled observations X.

Predicting the Results:

Given the training model, we have predicted the label output for the test set features using the predict function of the model, probabilities of the labels of test set features have also been predicted using the predict_proba method. In this predic_proba method, the highest probability is returned [27][28].

Evaluation of the results:

There are various evaluation methods available to measure the performance and the quality of the prediction made by the model such as roc_auc_score(), r2_score(), Confusion Matrix, etc.

In our implementation, we have used the roc_auc_score() method from the sklearn.metrics library. roc_auc_score metric is essentially defined for binary classification tasks. In this by default the positive class labeled is 1, only the positive labelled class is evaluated. Then the roc_auc_score function computes the area under the curve by this the curve information can be summarized in one number this is also denoted as AUC or AUROC [27][28].

V. RESULTS AND DISCUSSION

Results for H1N1 and Seasonal Flu Vaccination

In the default models it has been observed that the best performing method on the dataset has been the Artificial Neural Network method with 2 hidden layers and activation function being selu and the optimizer being SGD optimizer the sigmoid function is used in the output layer for activation function. The accuracy obtained with ANN is shown in Table I. Other machine learning algorithms have also yielded comparatively good results except logistic regression which has been the worst performing model with accuracy less than 70% in both H1N1 flu and seasonal flu vaccination prediction. Comparison of all the methods used during implementation has been shown in Table I.

TABLE I
Results for H1N1 flu and Seasonal flu vaccination prediction

Model	H1N1 Score	Seasonal Flu Score	Parameters
svm	0.8085	0.8596	{'kernel': 'rbf'}
random_forest	0.8154	0.8494	{'n_estimators': 100}
logistic_regression	0.6792	0.5949	{max_iter=1000}
ANN	0.8257	0.8601	{'1st hidden layer: 'units=60, activation='selu' '2nd hidden layer : 'units=30, activation='selu' output layer : 'units=1, activation='sigmoid' 'optimizer = 'SGD', loss = 'binary_crossentropy'}

The Results have also been plotted using the ROC AUC curve. In Fig 3, Fig 4, Fig 5, Fig 6 we can observe the performance of various models on the dataset and it can hence be concluded that Artificial Neural Network method has performed the best with accuracy over 82% in H1N1 flu vaccination prediction and 86% in Seasonal flu vaccination prediction.

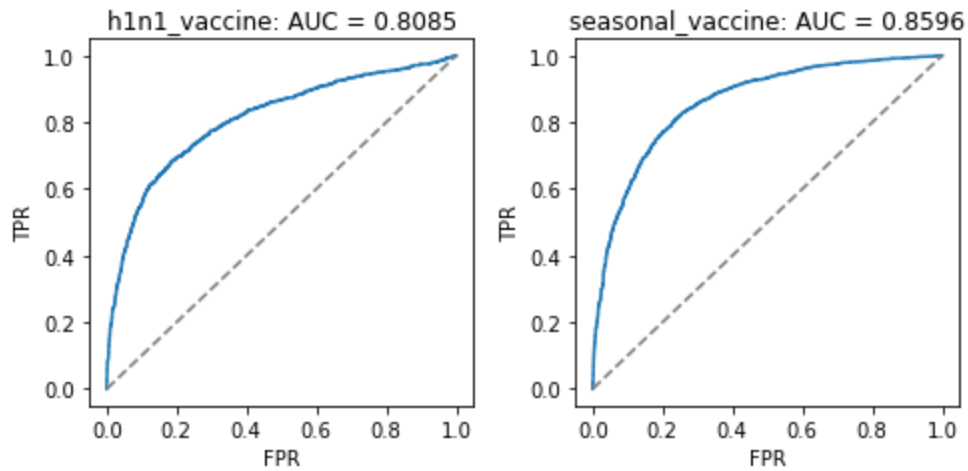


Fig. 3. ROC AUC Curve using Support Vector Machine: RBF Kernel

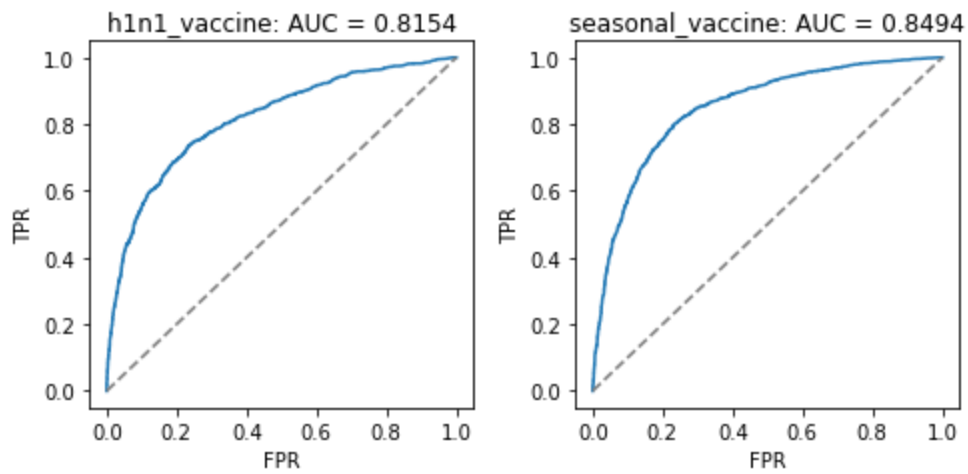


Fig. 4. ROC AUC Curve using Random Forest Regressor

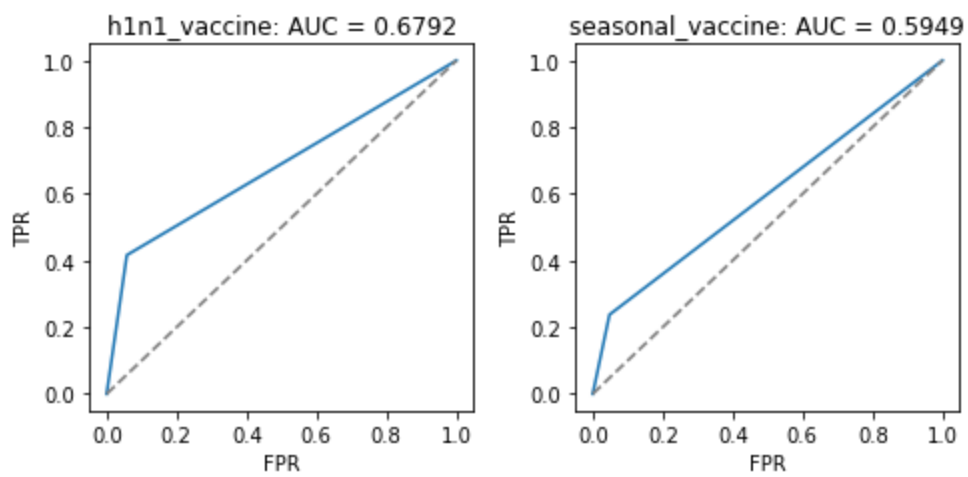


Fig. 5. ROC AUC Curve using Logistic Regression

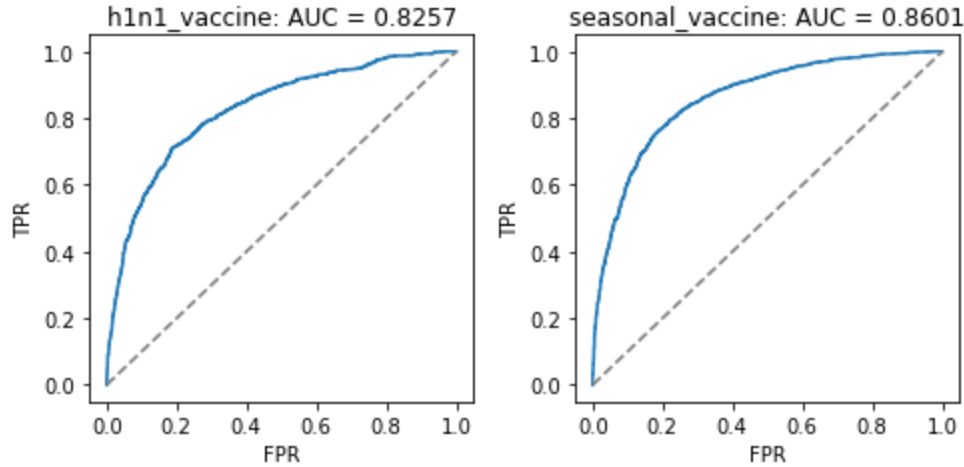


Fig. 6. ROC AUC Curve using Artificial Neural Network

Results for H1N1 and Seasonal Flu Vaccination with Hyperparameter Tuning

To obtain better results tuning of parameters has been done. Various methods such as GridSearchCV, RandomSearchCV, kfold method have been used for hyperparameter tuning of the machine learning models. It can be learned that Support Vector Machine with RBF kernel and C:20 using the GridSearchCV method yield the best result for H1N1 vaccination prediction as shown in Table II. The optimal parameters for random forest regression is training the model with '10' n_estimators, and the optimal parameters for logistic regression is C:5. All these results have been shown in tabulated form in Table II and Table III. It can also be observed that the results of Seasonal flu vaccination prediction have not been upto the mark using hyperparameter tuning, rather they have been better predicted using the default models the comparison can be seen in Table I and Table II, Table III.

TABLE II
Results with Hyperparameter tuning (GridSearchCV) for H1N1 flu vaccination prediction

Model	H1N1 Score	Seasonal flu Score	Parameters
svm	0.8397	0.7836	{ 'C':20, 'Kernel': 'rbf' }
random_forest	0.8213	0.7504	{ 'n_estimators': 10 }
logistic_regression	0.8363	0.7799	{ 'C': 5 }

TABLE III
Results with Hyperparameter tuning (RandomSearchCV) for H1N1 flu vaccination prediction

Model	H1N1 Score	Seasonal flu Score	Parameters
svm	0.8367	0.7836	{ 'Kernel': 'rbf', 'C':20 }
random_forest	0.8205	0.7467	{ 'n_estimators': 10 }
logistic_regression	0.8363	0.7799	{ 'C':5 }

Results for H1N1 and Seasonal Flu Vaccination in Artificial Neural Network with Hyperparameter Tuning

Kfold method has been used to fine tune hyperparameters in the Artificial Neural Network method. Thus the obtained results are somewhat equal to the default method but a slight increase in performance has been noted which can be clearly seen in Table IV. The most optimal parameters hence obtained for ANN with kfold method are 1st hidden layer with selu as activation function and having 60 units, the 2nd hidden layer with selu as activation function and having 3 units, and the output layer with sigmoid as activation function and having 2 units. All the results have been shown in Table IV.

TABLE IV
Results with Hyperparameter tuning (kfold method) for H1N1 flu and Seasonal vaccination prediction

Model	Best Score	Parameters
ANN	0.8323	{‘1st hidden layer: ‘units=60, activation=‘selu’ ‘2nd hidden layer : ‘units=3, activation=‘selu’ output layer : ‘units=2, activation=‘sigmoid’ ‘n_splits’=10,‘shuffle’=‘True’, ‘optimizer = ‘SGD’, loss = ‘binary_crossentropy’}

VI. CONCLUSION

In this paper, prediction of H1N1 and seasonal flu vaccination are based on the data source given by the Nation H1N1 flu survey"2009"(NHFS) and center of disease control (CDC) and prediction of H1N1 vaccination is done best by the help of SVM model with RBF kernel with the help of hyperparameter tuning using GridSearchCV which yielded an accuracy of 83.97% and seasonal flu vaccination prediction is done best with Artificial neural networks which yielded an accuracy of 86.10%. This vaccine helps in protection from the H1N1 virus and seasonal flu. Awareness was created in all kinds of social media in order to give the required information to the people regarding the importance of the H1N1 vaccine and seasonal flu vaccine in 2009. It was observed that the younger population was more affected than the population aged above 65 years. In order to immunize the society and to provide a good safeguard environment for everybody, vaccinations were provided.

VII. FUTURE RESEARCH SCOPE

Although we have achieved promising results, our study has many limitations. Mainly, the use of Twitter for data collection is not uniform across time and geography. Due to this inconsistency in the data our model's performance may vary and suffer. The accuracy difference can be noticed among the regional level and national level pertaining to the fact that people of the same region usually tend to have the same behavioural aspects hence lack of proper data could tamper the implementation. In future with advancements in technology the quality and the quantity of data could increase which could result in better performance and analysis of the issue. More information about the seasons, especially non-pandemic seasons could be very helpful for our analysis of this project. In future we also look forward to exploring more machine learning algorithms, methods and deep learning techniques so that we can obtain more optimal results.

VIII. ACKNOWLEDGEMENT

Success of our Internship project involving such high technical proficiency requires patience and massive support of guides this has been possible with commitment, passion and teamwork. We take this opportunity to express our gratitude to those who have been instrumental in the successful completion of this internship. We would like to show our appreciation to Dr. Kuldeep Chaurasia and Mr. Mayank Dixit for their tremendous support and help, without them this project would have reached nowhere. We would like to thank them for being our project coordinators and providing us with regular inputs about documentation and project timeline. A big thanks to Dr.

Madhushi Verma for all the encouragement, timely details and guidelines given to our team. We would also like to thank Dr. Deepak Garg, HOD of Computer Science Engineering department and Dr. Sudhir Chandra, Dean, School of Engineering & Applied Sciences, Bennett University for giving us the opportunity and the environment to learn and grow.

REFERENCES

- [1] 2009 H1N1 Pandemic (H1N1pdm09 virus), Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases (NCIRD), June 11, 2019. Accessed on: May 22, 2020. [Online]. Available: <https://www.cdc.gov/flu/pandemic-resources/2009-h1n1-pandemic.html>
- [2] 2009 H1N1 Flu ("Swine Flu") and You, Centres for Disease Control and Prevention, February 10, 2010. Accessed on: May 22, 2020. [Online]. Available: <https://www.cdc.gov/h1n1flu/qa.htm>
- [3] Update: Influenza Activity --- United States, 2009--10 Season, Centers for Disease Control and Prevention, July 30, 2010. Accessed on: May 22, 2020. [Online]. Available: <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5929a2.htm>
- [4] 2009 H1N1 Flu, Centres for Disease Control and Prevention, August 11, 2010. Accessed on: May 22, 2020. [Online]. Available: <https://www.cdc.gov/h1n1flu/>
- [5] Summary of the 2009-2010 Influenza Season, Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases (NCIRD), September 8, 2010. Accessed on: May 22, 2020. [Online]. Available: <https://www.cdc.gov/flu/pastseasons/0910season.htm>
- [6] Signorini A, Segre AM, Polgreen PM (2011) The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. *PLoS ONE* 6(5): e19467. Accessed on: May 22, 2020. [Online]. Available: <https://doi.org/10.1371/journal.pone.0019467>
- [7] Salathé Marcel, and Shashank Khandelwal. "Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control." *PLoS computational biology* vol. 7,10 (2011): e1002199. Accessed on: May 22, 2020. [Online]. doi:10.1371/journal.pcbi.1002199
- [8] M. S. Mabrouk and S. Y. Marzouk, "A chaotic study on pandemic and classical (H1N1) using EIIP sequence indicators," 2010 2nd International Conference on Computer Technology and Development, Cairo, 2010, pp. 218-221, doi: 10.1109/ICCTD.2010.564582.
- [9] T. T. S. Chinh, D. H. Stephanus, C. Kwoh, C. Schönbach and X. Li, "A possible mutation that enables H1N1 influenza a virus to escape antibody recognition," 2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Hong Kong, 2010, pp. 81-84, doi: 10.1109/BIBM.2010.5706541.
- [10] J. Huang, T. Lin, W. Chang and W. Hsieh, "Aptamer-modified CNTFET biosensor for detecting H1N1 virus in droplet," The 4th IEEE International NanoElectronics Conference, Tao-Yuan, 2011, pp. 1-2, doi: 10.1109/INEC.2011.5991640.
- [11] M. S. Ünlü, "Optical interference for multiplexed, label-free, and dynamic biosensing: Protein, DNA and single virus detection," 2011 XXXth URSI General Assembly and Scientific Symposium, Istanbul, 2011, pp. 1-2, doi: 10.1109/URSIGASS.2011.6050601.
- [12] Kamikawa, Tracy & Norton (Mikolajczyk), Malgorzata & Kennedy, Michael & Zhong, Lilin & Zhang, Pei & Settingington, Emma & Scott, Dorothy & Alocilja, Evangelyn. (2012). Pandemic Influenza Detection by Electrically Active Magnetic Nanoparticles and Surface Plasmon Resonance. *Nanotechnology*, IEEE Transactions on. 11. 88 - 96. 10.1109/TNANO.2011.2157936.
- [13] A. Baby Jerald and T. R. Gopalakrishnan Nair, "Influenza virus vaccine efficacy based on conserved sequence alignment," 2012 International Conference on Biomedical Engineering (ICoBE), Penang, 2012, pp. 327-329, doi: 10.1109/ICoBE.2012.6179031.
- [14] Chrysostomou, C., & Seker, H. (2013). Signal-processing-based bioinformatics approach for the identification of influenza A virus subtypes in Neuraminidase genes. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3066-3069.
- [15] N. Wiriyachaiyorn, H. Sirikett and T. Dharakul, "Rapid influenza a antigen detection using carbon nanostrings as label for lateral flow immunochromatographic assay," *2013 13th IEEE International Conference on Nanotechnology (IEEE-NANO 2013)*, Beijing, 2013, pp. 166-169, doi:10.1109/NANO.2013.6720979.
- [16] Y. Ma, W. Chang, C. Wang, H. Ma, P. Huang and G. Lee, "An integrated passive microfluidic device for rapid detection of influenza a (H1N1) virus by reverse transcription loop-mediated isothermal amplification

- (RT-LAMP)," 2017 19th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS), Kaohsiung, 2017, pp. 722-725, doi:10.1109/TRANSDUCERS.2017.7994150.
- [17] Nieto-Chaupis, Huber. (2019). Face To Face with Next Flu Pandemic with a Wiener-Series-Based Machine Learning: Fast Decisions to Tackle Rapid Spread. 0654-0658. 10.1109/CCWC.2019.8666474.
 - [18] F. Stalder and J. Hirsh, "Open source intelligence," *First Monday*, vol. 7, no. 6, 2002. 416
 - [19] M. Motoyama, B. Meeder, K. Levchenko, G. M. Voelker, and S. Savage, "Measuring online service availability using twitter." Workshop on online social networks, Boston, Massachusetts, USA, 2010.
 - [20] Wong, L. P., Sam. I. C, "Behavioral responses to the influenza A (H1N1) outbreak in Malaysia", *J Community Health*, 34:23-31, 2011.
 - [21] Bao Y., P. Bolotov, D. Dernovoy, B. Kiryutin, L. Zaslavsky, T. Tatusova, J. Ostell, and D. Lipman. The Influenza Virus Resource at the National Center for Biotechnology Information. *J. Virol.* 2008 Jan; 82(2): 596-601.
 - [22] Wei Hu, "Molecular features of highly pathogenic Avian and Human H5N1 Influenza A viruses in Asia," *Comput. Mol. Biosci.*, vol. 2, no. 2, pp. 45–59, 2012.
 - [23] Smith PJ, Wood D, Darden PM. Highlights of historical events leading to national surveillance of vaccination coverage in the United States. *Public Health Rep.* 2011;126 Suppl 2(Suppl 2):3-12. Accessed on: May 22, 2020.[Online]. doi: 10.1177/00333549111260S202
 - [24] Pedregosa, Fabian & Varoquaux, Gael & Gramfort, Alexandre & Michel, Vincent & Thirion, Bertrand & Grisel, Olivier & Blondel, Mathieu & Prettenhofer, Peter & Weiss, Ron & Dubourg, Vincent & Vanderplas, Jake & Passos, Alexandre & Cournapeau, David & Brucher, Matthieu & Perrot, Matthieu & Duchesnay, Edouard & Louppe, Gilles. (2012). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 12.
 - [25] Buitinck, Lars & Louppe, Gilles & Blondel, Mathieu & Pedregosa, Fabian & Mueller, Andreas & Grisel, Olivier & Niculae, Vlad & Prettenhofer, Peter & Gramfort, Alexandre & Grobler, Jaques & Layton, Robert & Vanderplas, Jake & Joly, Arnaud & Holt, Brian & Varoquaux, Gael. (2013). API design for machine learning software: Experiences from the scikit-learn project. *API Design for Machine Learning Software: Experiences from the Scikit-learn Project*.
 - [26] F. Dubosson, S. Bromuri and M. Schumacher, "A Python Framework for Exhaustive Machine Learning Algorithms and Features Evaluations," 2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), Crans-Montana, 2016, pp. 987-993, doi: 10.1109/AINA.2016.160.
 - [27] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, Microtome Publishing, 2011. fffhal-00650905v1f
 - [28] Virtanen, P., Gommers, R., Oliphant, T.E. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261–272 (2020). <https://doi.org/10.1038/s41592-019-0686-2>