

Article

Supervised Machine Learning Methods for Seasonal Influenza Diagnosis

Edna Marquez ^{1,*}, Eira Valeria Barrón-Palma ¹, Katya Rodríguez ², Jesus Savage ³ and Ana Laura Sanchez-Sandoval ¹¹ Genomic Medicine Department, General Hospital of México “Dr. Eduardo Liceaga”, Mexico City 06726, Mexico; valeirabarron@gmail.com (E.V.B.-P.)² Institute for Research in Applied Mathematics and Systems, National Autonomous University of Mexico, Mexico City 04510, Mexico; katya.rodriguez@iimas.unam.mx³ Signal Processing Department, Engineering School, National Autonomous University of Mexico, Mexico City 04510, Mexico; robotssavage@gmail.com

* Correspondence: cednam@gmail.com

Abstract: Influenza has been a stationary disease in Mexico since 2009, and this causes a high cost for the national public health system, including its detection using RT-qPCR tests, treatments, and absenteeism in the workplace. Despite influenza’s relevance, the main clinical features to detect the disease defined by international institutions like the World Health Organization (WHO) and the United States Centers for Disease Control and Prevention (CDC) do not follow the same pattern in all populations. The aim of this work is to find a machine learning method to facilitate decision making in the clinical differentiation between positive and negative influenza patients, based on their symptoms and demographic features. The research sample consisted of 15480 records, including clinical and demographic data of patients with a positive/negative RT-qPCR influenza tests, from 2010 to 2020 in the public healthcare institutions of Mexico City. The performance of the methods for classifying influenza cases were evaluated with indices like accuracy, specificity, sensitivity, precision, the f1-measure and the area under the curve (AUC). Results indicate that random forest and bagging classifiers were the best supervised methods; they showed promise in supporting clinical diagnosis, especially in places where performing molecular tests might be challenging or not feasible.

Keywords: machine learning; decision support system; medical diagnosis; influenza; artificial intelligence



Citation: Marquez, E.; Barrón-Palma, E.V.; Rodríguez, K.; Savage, J.; Sanchez-Sandoval, A.L. Supervised Machine Learning Methods for Seasonal Influenza Diagnosis.

Diagnostics **2023**, *13*, 3352. <https://doi.org/10.3390/diagnostics13213352>

Academic Editors: Dechang Chen and Mugahed A. Al-antari

Received: 22 August 2023

Revised: 24 October 2023

Accepted: 25 October 2023

Published: 31 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Influenza is a respiratory disease that can increase the incidence of pneumonia and cause a high number of hospitalizations [1]. In March 2009, Mexico, the United States and Canada were the focus of international attention when the influenza A H1N1 virus burst onto the epidemiological scene [2]. In June of that same year, the World Health Organization (WHO) declared an influenza pandemic of moderate severity. Since 2009, respiratory diseases due to influenza have recurred in numerous nations during the colder months annually, thus acquiring the category of seasonal influenza. There are four types of influenza viruses: A, B, C and D. Influenza A and B viruses cause seasonal epidemics of disease, and have been responsible for thousands of deaths worldwide, despite the annual vaccination campaigns [3]. In Mexico, between 2020 and 2021, the incidence of influenza decreased substantially due to the Coronavirus disease (COVID-19) caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2); however, influenza cases increased again in 2022 [4]. Before the COVID-19 pandemic, an estimated 291,000 to 646,000 respiratory deaths occurred worldwide each year due to seasonal influenza [5,6]. In Mexico, information regarding influenza infections has been registered since 2010 in the Influenza Epidemiological Surveillance System to identify its behavior and be able to predict how the next influenza season will develop [7]. The data in this system are obtained from symptomatic patients treated at healthcare centers and who had undergone

a quantitative reverse transcription polymerase chain reaction (RT-qPCR) test to detect the presence of influenza viral RNA.

The Centers for Disease Control and Prevention (CDC) specify the common symptoms experienced by influenza patients, such as fever or feeling feverish/chills, cough, sore throat, runny or stuffy nose, muscle or body aches, headaches, and fatigue (tiredness) [8]. The presence of these symptoms is not a guarantee of having been infected by the influenza virus; moreover, they vary among the population. Distinguishing the causal agent of the illness between the influenza virus and other viral or bacterial agents proves to be challenging through clinical evaluation alone. Therefore, other tests should be applied to confirm the diagnosis of influenza, RT-qPCR being the most successful test for the molecular diagnosis [9]. However, in developing countries, this test is not routinely performed due to high costs and the limited availability of testing facilities. Not all hospitals and clinics have the necessary equipment and supplies to perform these tests, leading to potentially lengthy turnaround times [10].

We propose to use alternative methods to facilitate the diagnosis of influenza, like methods based on Artificial Intelligence (AI), as they could serve as tools to assist in medical attention for diagnosis prior to RT-qPCR tests or can be applied in locations with difficult access to molecular analysis. In Figure 1, the workflow applied in this work to find the best Machine Learning method to diagnose influenza is shown.

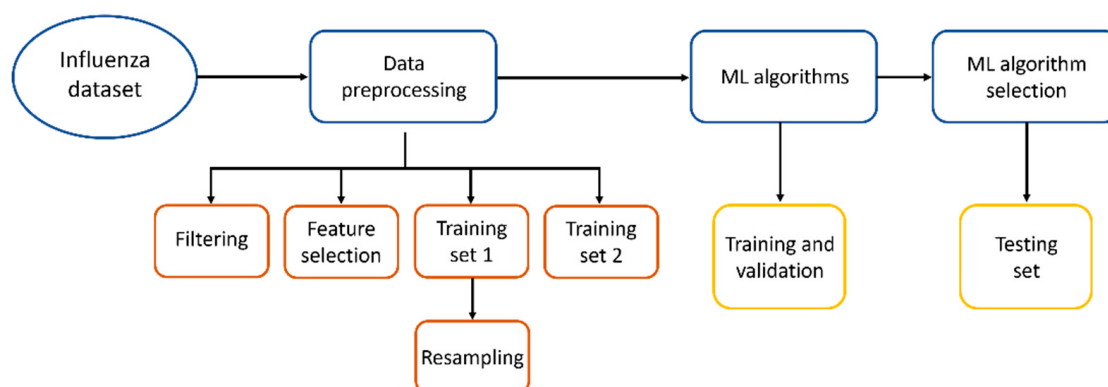


Figure 1. Workflow to find the best ML method to diagnose influenza.

Machine Learning (ML) is a component of AI that encompasses a set of techniques which enable the implementation of adaptive algorithms to make predictions and self-organize input data based on common characteristics. When ML is trained with a correct data set and the algorithm is standardized to accurately respond to all inputs, it is called a supervised ML [11].

Since the 1970s, interest in applying AI-ML in the health sector has grown [12]. In the medical field, a significant amount of patient data is managed, including sociodemographic and epidemiological information, results from physical examinations, diagnostic support test outcomes, and procedures performed, among others [10,13–47]. Because ML can effectively handle a large number of attributes (features), and due to its ability to identify and leverage the interactions among these numerous attributes, it becomes a particularly compelling tool in this domain [20]. ML algorithms have found extensive application across numerous medical specialties, serving purposes in prevention, diagnosis, treatment, and survival analysis alike.

In the case of influenza, ML has been applied to achieve several objectives, one of which is predicting the incidence of cases in the upcoming influenza seasons [38–40], including predicting the most prevalent types of influenza viruses for the season [41–43]. In the diagnostic stage, studies have demonstrated how metabolomic data from patients can be used to infer whether they are positive or negative for influenza [44]. There is even a report in which open access data were employed to develop a classifier for influenza

diagnosis; however, not all included patients had a RT-qPCR result to confirm the diagnosis and validate the classifier's functionality [10]. ML has also been applied to forecast the efficacy of influenza vaccines [45–47].

2. Materials and Methods

2.1. Data Set

In this study, a clinical data set comprising 19,160 patient records from Mexico City was used. The database was made to track influenza's seasonal behavior and make prognosis for the next season. Data excluded patients' names, home addresses and hospital registration numbers. The data set was exported to the authorized researcher for this retrospective study, which was reviewed and approved by the institutional ethical committee (D1/19/501-T/03/096).

We applied three exclusion criteria: (1) age < 7 years, (2) patients with a negative influenza test but positive for another respiratory virus, and (3) no RT-qPCR result record. After these criteria, the study included 15,480 records of patients aged between 7 and 119 years old. Figure 2 shows the data distribution. The age ranges were 7–19 (41.5%), 20–39 (20.1%), 40–59 (21.2%), and age ≥ 60 (17.2%); according to the RT-qPCR test, 11,268 (72.8%) were negative and 4212 (27.2%) were positive for influenza virus, and the distribution by sex was 7710 (49.8%) men, and 7770 (50.2%) women. The data set consisted of 24 attributes encompassing clinical and demographic information collected from patients upon arrival at healthcare institutions for clinical examination and before the sample taking (nasopharyngeal and/or oropharyngeal exudate) for a RT-qPCR test.

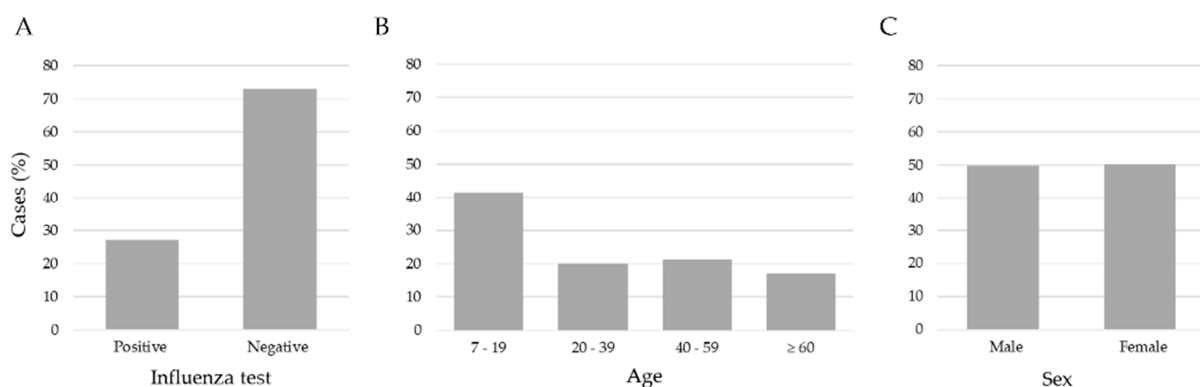


Figure 2. Distribution of influenza cases, ages and sex in the sample of 15,480 patients. (A) The number of negative (72.8) and positive (27.2) cases of influenza, (B) the ages separated into age ranges; (C) the % of feminine (50.2) and masculine (49.8) samples were very similar.

The set of symptoms and demographic features selected was the vector. The data were subjected to manual labeling by a clinician who assessed each patient and assigned values of 'yes' or 'no' to denote the presence or absence of symptoms, and 'unknown' otherwise.

2.2. Data Preprocessing

The symptoms and demographic data were labeled by binary numbers to indicate the presence or absence of a feature (1, 0, respectively); an unknown case was labeled as 0. Age range was mapped into the range {0...1}, to normalize the data.

In this study, two-step approaches were used to select the main features for training and testing the supervised ML methods. In the first step, Spearman's correlation was used to determine the correlation coefficient between features, as these are categorical variables. We selected them with a weak correlation ($r < 0.75$). In the second approach, chi-squared was computed to analyze the association between independent variables and influenza. The features selected had a strong association ($p < 0.001$).

To examine ML models and evaluate their performance, the data set was randomly split into 80% for the training set and 20% for the testing set. The models were evaluated

with 10-fold cross-validation to select the best one. Python 3 functions were applied to create the k-Fold distribution and stratification.

The original data set was unbalanced, with the majority of the cases being negative for influenza (72.8%) and a minority of cases with a positive influenza test (27.2%). In this study, the target classes of the training set were balanced 50:50, and the skew was eliminated to obtain a most appropriate performance of the ML methods [48]. The minority oversampling technique used was Random Oversampling (ROS), which increases the size of the data set by randomly resampling the original minority class without creating new samples or changing the sample variability [49]. All samples from the majority class (negative for influenza) were used, and through ROS, data were added to the minority class (positive for influenza), obtaining an equal number of samples in both classes.

2.3. Machine Learning Algorithms

ML methods are an automatic and objective way to classify the samples into two classes, positive or negative for influenza, using records with inputs and outputs for the process, features and their classification [50–52]. In this way, we tried to find patterns in the known data in order to apply them to the new unclassified data.

We had the pair (X,Y) in all cases, $X = \{x_1, x_2, \dots, x_n\}$ and $Y = [\text{positive} \mid \text{negative}]$ for influenza, according to the RT-qPCR test.

The set of signs, symptoms and demographic features selected are represented by the vector X, and the data are binary numbers that indicate the absence or presence of the feature. The age was normalized to a range $\{0 \dots 1\}$.

The aim was to find a model F of ML to represent the approximation between inputs and outputs.

$$F: X \rightarrow Y \quad (1)$$

For this study, 10 popular supervised ML methods through python.sklearn libraries for binary classification were used [53,54]: Adda Boost, Decision Tree, Bagging classification, Gradient Boosting Classifier (GBC), Random Forest (RF), K-nearest neighbors (Neighbors KNN), Naive Bayes (NB), Support Vector Machine (SVM), Logistic Regression (LR) and Discriminant Analysis. Taking advantage of the implementation of supervised ML models available in the python 3.7 programming language, the tests were carried out with several algorithms to evaluate their performance and be able to select the most accurate for this case study, and not only with the classical ML algorithms such as SVM, decision tree, KNN and NB.

Adaptive Boosting, named the AdaBoost algorithm, is a ML Meta-algorithm that can be used with many other ML algorithms to enhance its performance [55]. AdaBoost is sometimes denominated as the strongest out-of-the-box classifier for the so-called weak learners [56].

Decision trees used to predict categorical variables are called classification trees and decision trees [49]. The decision tree classifier is a flowchart-like tree structure; each internal node represents a test on an attribute, each branch represents an outcome of the test, and the class label is represented by each leaf nodes (or terminal nodes). Decision trees can be transformed into classification rules [57]. This ML algorithm is used to create the ensemble ML methods.

Random forest is a set of decision tree classifiers; in this ML model, each decision tree depends on a random vector of the training data set. They vote independently for the most popular class, and their classification is ensembled to give the final output using the classes given by each model [58]. The random forest algorithm is a special type of ensemble method. A random forest consists of many small classification decision or regression trees. Each tree, individually, is a weak learner; however, all the decision trees together can build a strong learner. It is random because (a) when building trees, a random sampling of training data sets is followed; and (b) when splitting nodes, a random subset of features is considered [59].

Bagging is a classifier which generates different subsets of the training data set by selecting data points randomly and with replacement. It can select the same instance multiple times. It is also called bootstrap aggregation and was created before the random forest. Given that a small change in the data can bring diverse effects in the model, the structure of the tree can completely change each tree to randomly sample the data set with a replacement, results on different trees [60].

The random forest algorithm is considered an extension of the bagging method. The difference is the number of features used in the decision tree construction: in bagging, all attributes are used for every decision tree, whereas in random forest, the decision trees have a random sample of attributes. Both ML methods are based on the decision tree method. The decision tree method works only with one tree to represent all samples and can be overfitting. Bagging and random forest work with several trees to represent different types of samples for each one.

Gradient boosting is a class of ensemble algorithms for machine learning that is used for regression or classification prediction modeling problems. It combines several sequential classifiers [61]. At an iteration, trees are added to the ensemble to fit correctly to the prediction errors made by prior models (boosting) and model fittings, using any arbitrary differentiable loss function and gradient descent optimization algorithm. The techniques is known as gradient boosting (Gboost) [62].

In the case of the KNN classifier, the main goal is to predict the closest value using distance as a basis. The Euclidean distance is a widely used technique [48]. The classification of the input data is based mainly on the selection of the majority class among its nearest neighbors [63].

The Naive Bayes classifier is considered a powerful probabilistic algorithm, based on Bayes theorem: the word “naïve” indicates interdependencies between characteristics. The version Bernoulli Naive Bayes (BNB) is used for Boolean variables as predictors [64,65].

Support Vector Machine (SVM) is a popular machine learning tool, which offers solutions to problems in classification and regression [66]. SVM separates classes and finds the hyperplane that best separates the data into different classes with a maximal margin between the classes. Initially using a linear decision boundary called a hyperplane to classify the data, Vapnik introduced a way of building a nonlinear classifier by using kernel functions [67]; it is placed at a location that maximizes the distance between the hyperplane and instances [68].

On the other hand, Discriminant Analysis aims to classify objects, by a set of independent variables, into one of two or more mutually exclusive and exhaustive categories. Discriminant analysis can be used only for classification (i.e., with a categorical target variable), not for regression. The target variable may have two or more categorical data by the use of multivariate information from the samples studied [69]. There are two methods: linear and quadratic discrimination. The first is the most widely used where classes are linearly separated. When a multi-classes analysis is needed, the two-groups method is used repeatedly in the analysis of pairs of data and the separation is linearity. Quadratic discrimination is used with nonlinearly separable classes.

2.4. Validation

The model performances were evaluated with a k-fold cross-validation method, which is an objective way to find the most robust ML algorithm, and we used the contingency table with the classification results [70].

In k-fold cross validation, the entire data set was divided into k equal parts, with k-1 subsets used for training while the remaining set was for testing. Each algorithm was trained and tested k times and the model output for each sample configuration obtained using cross validation was averaged to provide the global performance output of the model. The partition of the set with the folds in the subsets was arbitrary and with equal numbers of positive and negative cases in the training.

The confusion matrix helps to compare the classification result, and it has 4 values: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The columns in the matrix are the results obtained, while in the rows are the expected and results.

$$\begin{pmatrix} TP & FP \\ FN & TF \end{pmatrix}$$

We compared the performance of the methods in cross validation with the results using the following metrics:

$$\text{Accuracy, Acc} = \frac{TP + TF}{TN + FP + FN + TP} \quad (2)$$

This measure is for the samples correctly classified.

$$\text{Precision, Prec} = \frac{TP}{TP + FP} \quad (3)$$

measures the positive samples correctly classified vs. only positive samples.

$$\text{Recall, Rec} = \frac{TP}{TP + FN} \quad (4)$$

calculates the positive samples correctly classified vs. the samples expected to be positive. This is also called sensitivity.

$$\text{Specificity, Spec} = \frac{TN}{TN + FP} \quad (5)$$

measures the fraction of negative samples classified as negative.

$$\text{F1-score, F1} = \frac{2TP}{2TP + FP + FN} \quad (6)$$

It is an average between recall and precision.

Additionally, all the models were evaluated with the area under the curve (AUC) score in ROC analysis. This measure uses the ROC curve that shows the ability to make the difference between 2 classes with a graphical method using recall and specificity, and the AUC summarizes the performance of the classifiers in the training.

3. Results

At the beginning, the database had 24 attributes. According to Spearman's correlation, the feature arthralgia was highly correlated with myalgia ($r = 0.87$); hence, only myalgia was selected in the analysis, as it had fewer missing values. In the chi-squared test (p -value in Table 1), factors like sex, diarrhea and vaccination presented a low association with influenza ($p > 0.01$); therefore, these factors were also dismissed. Finally, 20 features were selected, many more than the 8 main symptoms of influenza indicated by the CDC: fever or feeling feverish/chills, cough, sore throat, runny or stuffy nose, muscle or body aches, headaches, fatigue (tiredness). Age data were used like one factor normalized in the range $\{0 \dots 1\}$.

Table 1. Attributes from influenza database of Mexico City.

Attributes	All Patients <i>n</i> = 15,480 <i>n</i> (%)	Positive <i>n</i> = 4212 <i>n</i> (%)	Negative <i>n</i> = 11,268 <i>n</i> (%)	<i>p</i> -Value
<i>Demographic information</i>				
Sex—Feminine	7770 (50.2)	2162 (51.3)	5608 (49.8)	0.087
Sex—Masculine	7710 (49.8)	2050 (48.7)	5660 (50.2)	
Hospitalized	10516 (67.9)	2449 (58.1)	8067 (71.6)	<0.001
Contact influenza-patients	2012 (13.0)	715 (17.0)	1297 (11.5)	<0.001
Vaccinated for influenza	2096 (13.5)	534 (12.7)	1562 (13.9)	0.059
Age 7–19 years	6417 (41.5)	1511 (35.9)	4906 (43.5)	<0.001
Age 20–39 years	3111 (20.1)	967 (23.0)	2144 (19.0)	
Age 40–59 years	3283 (21.2)	1056 (25.0)	2227 (19.8)	
Age ≥ 60 years	2669 (17.2)	678 (16.1)	1991 (17.7)	
<i>Symptoms</i>				
Fever	13,112 (84.7)	3853 (84.2)	9259 (82.2)	<0.001
Cough	13,953 (90.1)	3918 (85.7)	10,035 (89.1)	<0.001
Chest pain	3750 (24.2)	1160 (25.4)	2590 (23.0)	<0.001
Dyspnea	8642 (55.8)	2079 (45.5)	6563 (58.2)	<0.001
Irritability	4688 (30.3)	1159 (25.3)	3529 (31.3)	<0.001
Diarrhea	1833 (11.8)	492 (10.8)	1341 (11.9)	0.727
Shaking chills	5738 (37.1)	2003 (43.8)	3735 (33.1)	<0.001
Headache	8692 (56.1)	2896 (63.3)	5796 (51.4)	<0.001
Myalgia	6255 (40.4)	2279 (49.8)	3976 (35.3)	<0.001
Arthralgia	5539 (35.8)	2014 (44.0)	3525 (31.3)	<0.001
Malaise	9826 (63.5)	2947 (64.4)	6879 (61.0)	<0.001
Rhinorrhea	9277 (59.9)	2817 (61.6)	6460 (57.3)	<0.001
Polypnea	4602 (29.7)	1073 (23.5)	3529 (31.3)	<0.001
Vomiting	1958 (12.6)	606 (13.2)	1352 (12.0)	<0.001
Abdominal pain	2114 (13.7)	683 (14.9)	1431 (12.7)	<0.001
Sore throat	5321 (34.4)	1850 (40.4)	3471 (30.8)	<0.001
Conjunctivitis	3074 (19.9)	1104 (24.1)	1970 (17.5)	<0.001
Cyanosis	1703 (11.0)	395 (8.6)	1308 (11.6)	<0.001

The *p*-value corresponds to chi-squared. For the study, the attributes sex, vaccinated for influenza, and diarrhea were excluded with *p*-value > 0.001, and the others were selected.

The number of samples to test with ML was 15,840. This database was unbalanced with 11,268 (72.8%) negative and 4212 (27.2%) positive for influenza, according to the RT-qPCR tests. In this work, Random Over Sampling (ROS) was used to increase the number of positive records in the training set to improve the method performance with an equal number of samples in positive and negative classes.

The features sex, vaccination and diarrhea were not significant in the chi-squared test, with the dependent variable influenza: these features showed small variation between the positive and the negative classes.

Finally, our balanced training set had 7918 of positive and the same number of negative rows and 20 columns of features for training and testing 10 supervised ML algorithms,

validated with k-fold cross validation ($k = 10$). In Table 2 are the results of ML methods with cross-validation. RF had the best evaluation ($AUC = 0.94$, $Acc = 0.86$, $Rec = 0.91$ and $Spec = 0.88$ were the best); in second place was the bagging classifier, which works similarly to RF. With the resampling technique, the number of the minor class (positive for influenza) increased, and the scores reflected the equilibrium. Figure 3 shows the ROC curves of the four best ML methods in the 10-fold cross-validation.

Table 2. Results of supervised machine learning algorithms.

Algorithm	AUC	Acc	Rec	Prec	Spec	F1
Random Forest	0.94	0.86	0.91	0.82	0.88	0.86
Bagging	0.93	0.85	0.90	0.82	0.87	0.85
Decision Tree	0.85	0.70	0.71	0.73	0.73	0.72
Kneighbors (7)	0.73	0.63	0.67	0.63	0.60	0.63
Gradient Boosting	0.69	0.62	0.69	0.61	0.56	0.62
SVM rbf	0.67	0.62	0.65	0.61	0.59	0.62
Quadratic Discriminant	0.66	0.62	0.70	0.60	0.54	0.62
Ada Boost	0.66	0.62	0.62	0.61	0.61	0.62
Linear Discriminant *	0.65	0.61	0.62	0.61	0.61	0.61
Linear SVM *	0.65	0.61	0.62	0.61	0.61	0.61
Logistic Regression	0.65	0.61	0.62	0.61	0.61	0.61
BernoulliNB	0.65	0.61	0.59	0.61	0.62	0.61

* Discriminant Analysis and SVM were used twice with different parameters. AUC, area under the curve; Acc, accuracy; Rec, recall; Prec, precision; Spec, specificity; D1, F1-score.

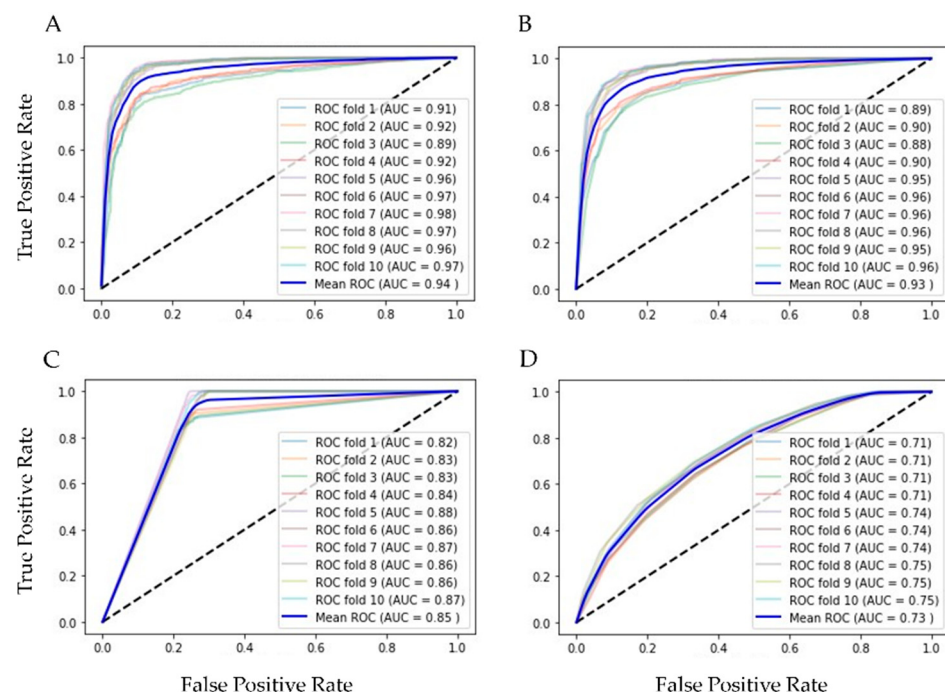


Figure 3. ROC graphics of the best ML algorithms with training set. (A) Random Forest, $AUC = 0.94 \pm 0.004$; (B) Bagging, $AUC = 0.93 \pm 0.004$; (C) Decision Tree, $AUC = 0.85 \pm 0.006$; and (D) Kneighbors (7), $AUC = 0.73 \pm 0.014$.

Random Forest, Bagging, and Decision Tree are supervised learning algorithms that can effectively handle categorical or binary features like the ones we have. The applied resampling seemed to have benefited the sensitivity results as it increased the number of

positive samples in the training set. Bagging and Random Forest are ensemble techniques, based on Decision Tree, which ranked third, and along with Kneighbors, are algorithms capable of capturing non-linear relationships in the data.

The performance of the top four ML methods is shown in Figure 4, and their respective ROC plots are shown in Figure 5, both showing the results obtained with the test set. The RF and bagging methods demonstrated the highest scores when applied to the independent samples in the test set. However, it is important to note a substantial disparity between the sensitivity and specificity results in the test set. random forest achieved a sensitivity (rec) of 0.30 and a specificity (spec) of 0.90, while bagging had a sensitivity of 0.29 and a specificity of 0.88. Additionally, the significant differences observed between the results during the training and testing phases highlight certain limitations in this study. One potential factor contributing to these limitations is the sensitivity of machine learning models to class imbalance. Even after implementing random oversampling (ROS) on the training set to address the issue of imbalance, especially considering the considerably higher proportion of negative samples compared to positive samples, the desired variability was not effectively introduced into the training data. Another contributing factor might be attributed to the nature of the data set itself. In this study, we utilized binary data to represent the presence or absence of specific characteristics, with the exception of age, which was represented as a continuous variable. In contrast, other studies in the medical field that have achieved superior results have not only incorporated binary variables but have also integrated continuous variables obtained from laboratory tests and biometric measurements to assess patients' conditions [19,20,32,37]. An illustrative case from [21] involved the transformation of continuous data into binary values, but this approach also yielded unsatisfactory results. It is plausible that, in our case, the lack of relevant information and the use of subjective values to evaluate the health status of patients may have led to weaker associations between these characteristics and the occurrence of influenza.

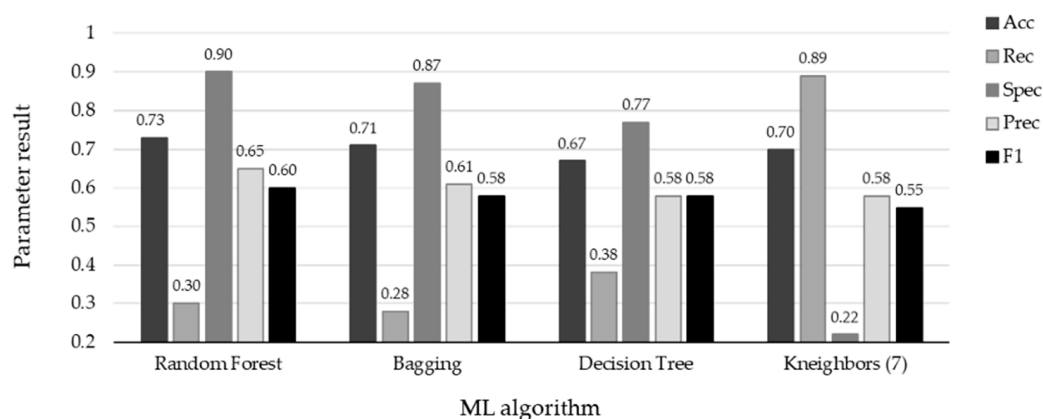


Figure 4. Results with test set using the four ML algorithms. The top scores are associated with specificity for the four algorithms: RF (spec = 0.90), Bagging (spec = 0.87), DT (spec = 0.77), and KNN (spec = 0.89). Conversely, the remaining metrics indicate the misclassification of positive influenza cases.

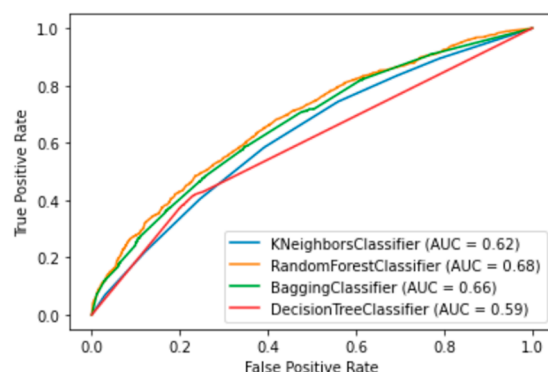


Figure 5. ROC graphics of the four best ML algorithms. Random Forest, Bagging, Decision Tree and Kneighbors (7).

4. Discussion

The use of artificial intelligence techniques through ML has increased its possibility as an alternative or powered tool in the diagnosis of infectious diseases [71–73]. With this idea, in this work we searched for an ML method as an alternative to the PCR test to perform diagnostics of influenza. Here, 19 category features and age were used, collected when the patient arrived with symptoms of influenza. In order to eliminate the unbalanced data between the positive and negative influenza cases of 15840 samples and to reduce the skew, we applied a resampling technique, random over sampling to positives. With resampling techniques, the Ensemble ML methods like Random Forest and Bagging could be favored—RF with AUC = 0.94, Acc = 0.86, Spec = 0.88 and, sensibility = 0.91, and Bagging with AUC = 0.93, Acc = 0.85, Spec = 0.87 and sensibility = 0.90—in the cross-validation. In other works [74,75] RF and Bagging were also combined with resampling techniques and showed good performance.

Nevertheless, it is important to note that these ML methods may serve as valuable screening tools to assist medical practitioners in distinguishing between positive and negative influenza cases, yielding promising results that could aid in decision making. This is particularly relevant for scenarios where the RT-qPCR test results are expected to be negative, potentially leading to reduced costs associated with testing.

Our research findings were based exclusively on data obtained from Mexican patients. This approach was chosen due to the unique health conditions prevalent in Mexico, which may differ significantly from those in other countries. It is important to consider that COVID-19 has changed the patterns of respiratory diseases [76,77]. Even though vaccines are applied every year, many people around the world are infected with influenza, causing a large number of deaths [4,78,79].

The potential advantage highlighted in our study is the use of an alternative decision-support tool, particularly relevant to regions where healthcare providers or patients, armed solely with basic questioning information, can assess the necessity for treatment and the conduction of PCR tests for the influenza disease.

4.1. Limitations of Work

Our results show problems in the prediction of positive influenza cases, maybe because the data set is imbalanced, and the binary features lose representativeness of the patient's health status. ML techniques hold potential in diverse applications; however, it is crucial to acknowledge that, in this study, these methods play a limited role as detection tools. They should not be perceived as a complete substitute for clinical diagnosis. RT-qPCR tests retain their indispensable status for precise influenza results, and therefore, machine learning models should be considered as complementary rather than as a complete replacement for conventional diagnostic approaches. Our findings indicate challenges in predicting positive influenza cases, possibly due to data imbalance and the diminished representativeness of binary features concerning a patients' health status.

It is possible that the low prediction values could be improved with our data set through several avenues. One approach could involve grouping individuals according to age, as symptoms may exhibit more pronounced patterns within specific age groups. Exploring alternative machine learning models is also a worthwhile consideration in our quest for improved predictions. Additionally, expanding the data set by including more positive cases from different Mexican regions could enhance the models' performance.

4.2. Future Work

This research has the potential for ongoing improvement and broader application. Comprehensive ablation studies can provide deeper insights into the algorithm's capabilities, allowing for a clearer grasp of its strengths and weaknesses. These studies encompass various facets, including feature selection, the incorporation of additional continuous data to enhance patient health assessment, the adoption of class balancing techniques, and the use of advanced machine learning models like convolutional neural networks to handle larger data sets and continuous data. Furthermore, it is essential to explore the creation of comprehensive models that effectively differentiate between COVID-19 and influenza cases.

5. Conclusions

In this study, machine learning models showcased a notably higher specificity compared to sensitivity, suggesting their potential utility in the identification of negative cases. This capability could help minimize the number of unnecessary molecular tests for individuals presenting with symptoms resembling influenza. This aspect is particularly pertinent in Mexico, where, for epidemiological reasons, during the influenza season around 10% of the population with symptoms resembling influenza are randomly selected for RT-qPCR testing, with approximately 70% of those cases turning out to be negative. By incorporating a tool akin to the one outlined in this study, clinicians can make more informed decisions about which patients require PCR testing, ultimately enhancing data quality for national-level decision making. Furthermore, given the limited availability of RT-qPCR testing facilities in certain areas, this tool can serve as valuable support for healthcare practitioners, aiding them in determining the necessity of conducting tests. This approach has the potential to reduce costs for patients and ease the burden on the healthcare sector.

Author Contributions: Conceptualization and methodology, E.M. and E.V.B.-P.; validation, E.V.B.-P. and J.S.; investigation, E.M. and K.R.; data curation, E.M.; writing—original draft preparation, E.M. and K.R.; writing—review and editing, E.V.B.-P. and A.L.S.-S.; supervision, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The work follows the principles of the Declaration of Helsinki and the Research Ethics Committee of the General Hospital of Mexico “Dr. Eduardo Liceaga” institutional ethical committee reviewed and approved the study design (D1/19/501-T/03/096).

Informed Consent Statement: Informed consent was not required by patients because this is a retrospective study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Centro Nacional de Programas de Control y Preventivos de Enfermedades. Manual de Atención a la Salud Ante Emergencias. Available online: <https://epidemiologia.salud.gob.mx/gobmx/salud/documentos/manuales/> (accessed on 1 February 2022).
2. LaRussa, P. Pandemic novel 2009 H1N1 influenza: What have we learned? *Semin. Respir. Crit. Care Med.* **2011**, *32*, 393–399. [CrossRef] [PubMed]
3. Gordon, A.; Reingold, A. The Burden of Influenza: A Complex Problem. *Curr. Epidemiol. Rep.* **2018**, *5*, 1–9. [CrossRef] [PubMed]

4. Epidemiology General Vigilance of Mexico, Informe Semanal de Vigilancia Epidemiológica. Available online: https://www.gob.mx/cms/uploads/attachment/data/file/737555/INFLUENZA_OVR_SE26_2022.pdf (accessed on 8 July 2022).
5. Krammer, F.; Smith, G.J.D.; Fouchier, R.A.M.; Peiris, M.; Kedzierska, K.; Doherty, P.C.; Palese, P.; Shaw, M.L.; Treanor, J.; Webster, R.G.; et al. Influenza. *Nat. Rev. Dis. Primers* **2018**, *4*, 3. [CrossRef]
6. Chow, E.J.; Doyle, J.D.; Uyeki, T.M. Influenza virus-related critical illness: Prevention, diagnosis, treatment. *Crit. Care* **2019**, *23*, 214. [CrossRef]
7. Ruiz-Matus, C.; Kuri-Morales, P.; Narro-Robles, J. Comportamiento de las temporadas de influenza en México de 2010 a 2016, análisis y prospectiva. *Gac. Med. Mex.* **2017**, *153*, 205–213.
8. Centers for Disease Control and Prevention. Flu Symptoms and Complications. Available online: <https://www.cdc.gov/flu/symptoms/symptoms.htm> (accessed on 13 July 2022).
9. World Health Organization. Influenza (Seasonal). Available online: [https://www.who.int/news-room/fact-sheets/detail/influenza-\(seasonal\)](https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)) (accessed on 13 July 2022).
10. López-Pineda, A.; Ye, Y.; Visweswaran, S.; Cooper, G.F.; Wagner, M.M.; Tsui, F.R. Comparison of machine learning classifiers for influenza detection from emergency department free-text reports. *J. Biomed. Inform.* **2015**, *58*, 60–69. [CrossRef]
11. Bonaccorso, G. *Machine Learning Algorithms*, 1st. ed.; Pack Publishing: Birmingham, UK, 2017; pp. 8–16.
12. Pandya, S.; Thakur, A.; Saxena, S.; Jassal, N.; Patel, C.; Modi, K.; Shah, P.; Joshi, R.; Gonge, S.; Kadam, K.; et al. A Study of the Recent Trends of Immunology: Key Challenges, Domains, Applications, Datasets, and Future Directions. *Sensors* **2021**, *21*, 7786. [CrossRef]
13. Vijayan, V.V.; Anjali, C. Prediction and diagnosis of diabetes mellitus—A machine learning approach. In Proceedings of the IEEE Recent Advances in Intelligent Computational Systems (RAICS), Trivandrum, India, 20 April 2021; pp. 122–127. [CrossRef]
14. Pecht, M.G.; Kang, M. Machine Learning: Diagnostics and Prognostics. In *Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things*; John Wiley and Sons Ltd.: Hoboken, NJ, USA, 2019; pp. 163–191. [CrossRef]
15. Shigueoka, L.S.; de Vasconcellos, J.P.C.; Schimiti, R.B.; Reis, A.S.C.; de Oliveira, G.O.; Gomi, E.S.; Vianna, J.A.R.; Lisboa, R.D.d.R.; Medeiros, F.A.; Costa, V.P. Automated algorithms combining structure and function outperform general ophthalmologists in diagnosing glaucoma. *PLoS ONE* **2018**, *13*, e0207784. [CrossRef]
16. Ullah, R.; Khan, S.; Ali, H.; Chaudhary, I.I.; Bilal, M.; Ahmad, I. A comparative study of machine learning classifiers for risk prediction of asthma disease. *Photodiagn. Photodyn. Ther.* **2019**, *28*, 292–296. [CrossRef]
17. Akazawa, M.; Hashimoto, K. Artificial Intelligence in Ovarian Cancer Diagnosis. *Anticancer Res.* **2020**, *40*, 4795–4800. [CrossRef]
18. Lu, W.; Tong, Y.; Yu, Y.; Xing, Y.; Chen, C.; Shen, Y. Applications of Artificial Intelligence in Ophthalmology: General Overview. *J. Ophthalmol.* **2018**, *2018*, 5278196. [CrossRef] [PubMed]
19. Silva, F.R.; Vidotti, V.G.; Cremasco, F.; Días, M.; Gomi, E.S.; Costa, V.P. Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using Spectral Domain OCT and standard automated perimetry. *Arq. Bras. Oftalmol.* **2013**, *76*, 170–174. [CrossRef] [PubMed]
20. Guncar, G.; Kukar, M.; Notar, M.; Brvar, M.; Cernelc, P. An application of machine learning to haematological diagnosis. *Sci. Rep.* **2018**, *8*, 411. [CrossRef] [PubMed]
21. Smith, J.P.; Milligan, K.; McCarthy, K.D.; Mchembere, W.; Okeyo, E.; Musau, S.K.; Okumu, A.; Song, R.; Click, E.S.; Cain, K.P. Machine learning to predict bacteriologic confirmation of *Mycobacterium tuberculosis* in infants and very young children. *PLoS Digit. Health* **2023**, *2*, 249. [CrossRef]
22. Peng, B.; Gong, H.; Tian, H.; Zhuang, Q.; Li, J.; Cheng, K.; Ming, Y. The study of the association between immune monitoring and pneumonia in kidney transplant recipients through machine learning models. *J. Transl. Med.* **2020**, *18*, 370. [CrossRef]
23. Saybani, M.R.; Shamshirband, S.; Hormozi, S.G.; Wah, T.Y.; Aghabozorgi, S.; Pourhoseingholi, M.A.; Olariu, T. Diagnosing tuberculosis with a novel support vector machine-based artificial immune recognition system. *Iran. Red Crescent Med. J.* **2015**, *17*, e24557. [CrossRef]
24. Melendez, J.; Sánchez, C.I.; Philipsen, R.H.H.M.; Maduskar, P.; Dawson, R.; Theron, G.; Dheda, K.; van Ginneken, B. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Sci. Rep.* **2016**, *6*, 25265. [CrossRef]
25. Er, O.; Temurtas, F.; Tanrikulu, A.Ç. Tuberculosis disease diagnosis using artificial neural networks. *J. Med. Syst.* **2010**, *34*, 299–302. [CrossRef]
26. e Souza, J.B.D.O.; Sanchez, M.; de Seixas, J.M.; Maidantchik, C.; Galliez, R.; Moreira, A.D.S.R.; da Costa, P.A.; Oliveira, M.M.; Harries, A.D.; Kritski, A.L. Screening for active pulmonary tuberculosis: Development and applicability of artificial neural network models. *Tuberculosis* **2018**, *111*, 94–101. [CrossRef]
27. Revuelta-Zamorano, P.; Sánchez, A.; Rojo-Álvarez, J.L.; Álvarez-Rodríguez, J.; Ramos-López, J.; Soguero-Ruiz, C. Prediction of healthcare associated infections in an intensive care unit using machine learning and big data tools. In Proceedings of the XIV Mediterranean Conference on Medical and Biological Engineering and Computing, Paphos, Cyprus, 31 March–2 April 2016; Springer International Publishing: Cham, Switzerland, 2016; Volume 57, p. 840e5. [CrossRef]
28. Hernandez, B.; Herrero, P.; Rawson, T.M.; Moore, L.S.P.; Evans, B.; Toumazou, C.; Holmes, A.H.; Georgiou, P. Supervised learning for infection risk inference using pathology data. *BMC Med. Inform. Decis. Mak.* **2017**, *17*, 168. [CrossRef]

29. Van Steenkiste, T.; Ruyssinck, J.; De Baets, L.; Decruyenaere, J.; De Turck, F.; Ongenae, F.; Dhaene, T. Accurate prediction of blood culture outcome in the intensive care unit using long short-term memory neural networks. *Artif. Intell. Med.* **2018**, *97*, 38–43. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Ke, C.; Jin, Y.; Evans, H.; Lober, B.; Qian, X.; Liu, J.; Huang, S. Prognostics of surgical site infections using dynamic health data. *J. Biomed. Inform.* **2017**, *65*, 22–33. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Horng, S.; Sontag, D.A.; Halpern, Y.; Jernite, Y.; Shapiro, N.I.; Nathanson, L.A. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS ONE* **2017**, *12*, e0174708. [\[CrossRef\]](#)
32. Taylor, R.A.; Moore, C.L.; Cheung, K.; Brandt, C. Predicting urinary tract infections in the emergency department with machine learning. *PLoS ONE* **2018**, *13*, e0194085. [\[CrossRef\]](#)
33. Revett, K.; Gorunescu, F.; Ene, M. A machine learning approach to differentiating bacterial from viral meningitis. In Proceedings of the IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing, Sofia, Bulgaria, 3–6 October 2006; pp. 155–162. [\[CrossRef\]](#)
34. D'Angelo, G.; Pilla, R.; Tascini, C.; Rampone, S. A proposal for distinguishing between bacterial and viral meningitis using genetic programming and decision trees. *Soft Comput.* **2019**, *23*, 11775–11791. [\[CrossRef\]](#)
35. Shamshirband, S.; Hessam, S.; Javidnia, H.; Amiribesheli, M.; Vahdat, S.; Petković, D.; Gani, A.; Kiah, L.M. Tuberculosis disease diagnosis using artificial immune recognition system. *Int. J. Med. Sci.* **2014**, *11*, 508–514. [\[CrossRef\]](#)
36. Jayatilake, S.M.; Ganegoda, G.U. Involvement of Machine Learning Tools in Healthcare Decision Making. *J. Healthc. Eng.* **2021**, *2021*, 6679512. [\[CrossRef\]](#)
37. Altini, N.; Brunetti, A.; Mazzoleni, S.; Moncelli, F.; Zagaria, I.; Prencipe, B.; Lorusso, E.; Buonamico, E.; Carpagnano, G.E.; Bavaro, D.F.; et al. Predictive Machine Learning Models and Survival Analysis for COVID-19 Prognosis Based on Hematochemical Parameters. *Sensors* **2021**, *21*, 8503. [\[CrossRef\]](#)
38. Borkenhagen, L.K.; Allen, M.W.; Runstadler, J.A. Influenza virus genotype to phenotype predictions through machine learning: A systematic review. *Emerg. Microbes Infect.* **2021**, *10*, 1896–1907. [\[CrossRef\]](#)
39. Cheng, H.-Y.; Wu, Y.-C.; Lin, M.-H.; Liu, Y.-L.; Tsai, Y.-Y.; Wu, J.-H.; Pan, K.-H.; Ke, C.-J.; Chen, C.-M.; Liu, D.-P.; et al. Applying Machine Learning Models with An Ensemble Approach for Accurate Real-Time Influenza Forecasting in Taiwan: Development and Validation Study. *J. Med. Internet Res.* **2020**, *22*, e15394. [\[CrossRef\]](#)
40. Reich, N.G.; McGowan, C.J.; Yamana, T.K.; Tushar, A.; Ray, E.L.; Osthus, D.; Kandula, S.; Brooks, L.C.; Crawford-Crudell, W.; Gibson, G.C.; et al. Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. *PLoS Comput. Biol.* **2019**, *15*, e1007486. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Hayati, M.; Biller, P.; Colijn, C. Predicting the short-term success of human influenza virus variants with machine learning. *Proc. Biol. Sci.* **2020**, *287*, 20200319. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Kwon, E.; Cho, M.; Kim, H.; Son, S. A Study on Host Tropism Determinants of Influenza Virus Using Machine Learning. *Curr. Bioinform.* **2020**, *15*, 121–134. [\[CrossRef\]](#)
43. Al Dalbhi, S.; Alshahrani, H.A.; Almadi, A.; Busaleh, H.; Alotaibi, M.; Almutairi, W.; Almukhrq, Z. Prevalence and mortality due to acute kidney injuries in patients with influenza A (H1N1) viral infection: A systemic narrative review. *Int. J. Health Sci.* **2019**, *13*, 56–62.
44. Hogan, C.A.; Rajpurkar, P.; Sowrirajan, H.; Phillips, N.A.; Le, A.T.; Wu, M.; Garamani, N.; Sahoo, M.K.; Wood, M.L.; Huang, C.; et al. Nasopharyngeal metabolomics and machine learning approach for the diagnosis of influenza. *EBioMedicine* **2021**, *71*, 103546. [\[CrossRef\]](#)
45. Fukuta, H.; Goto, T.; Wakami, K.; Kamiya, T.; Ohte, N. The effect of influenza vaccination on mortality and hospitalization in patients with heart failure: A systematic review and meta-analysis. *Heart Fail. Rev.* **2019**, *24*, 109–114. [\[CrossRef\]](#)
46. Tomic, A.; Tomic, I.; Rosenberg-Hasson, Y.; Dekker, C.L.; Maecker, H.T.; Davis, M.M. SIMON, an Automated Machine Learning System, Reveals Immune Signatures of Influenza Vaccine Responses. *J. Immunol.* **2019**, *203*, 749–759. [\[CrossRef\]](#)
47. Wolk, D.M.; Lanyado, A.; Tice, A.M.; Shermohammed, M.; Kinar, Y.; Goren, A.; Chabris, C.F.; Meyer, M.N.; Shoshan, A.; Abedi, V. Prediction of Influenza Complications: Development and Validation of a Machine Learning Prediction Model to Improve and Expand the Identification of Vaccine-Hesitant Patients at Risk of Severe Influenza Complications. *J. Clin. Med.* **2022**, *11*, 4342. [\[CrossRef\]](#)
48. López, V.; Fernández, A.; García, S.; Herrera, F. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Inf. Sci.* **2013**, *250*, 113–141. [\[CrossRef\]](#)
49. Chawla, N.V. Data Mining for Imbalanced Datasets: An Overview. In *Data Mining and Knowledge Discovery Handbook*; Springer: Boston, MA, USA, 2005.
50. Kamal, K.H.; Ritesh, K.J.; Kamlesh, L.; Ruchi, D. *Machine Learning: Master Supervised and Unsupervised Learning Algorithms with Real Examples*; BPB Publications: Noida, India, 2021.
51. Alpaydin, E. *Machine Learning: The New AI*; MIT Press Essential Knowledge Series; MIT Press: London, UK, 2016.
52. Singh, A.; Thakur, N.; Sharma, A. A review of supervised machine learning algorithms. In Proceedings of the 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016; pp. 1310–1315.
53. Raschka, S. *Python Machine Learning*, 2nd ed.; Packt Publishing: Birmingham, UK, 2017.
54. Müller, A.C.; Guido, S. *Introduction to Machine Learning with Python: A Guide for Data Scientists*, 1st ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2016.

55. Mathanker, S.K.; Weckler, P.R.; Bowser, T.J.; Wang, N.; Maness, N.O. AdaBoost classifiers for pecan defect classification. *Comput. Electron. Agric.* **2011**, *77*, 60–68. [\[CrossRef\]](#)
56. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [\[CrossRef\]](#)
57. Decision-Trees. Available online: <https://www.ibm.com/topics/decision-trees> (accessed on 5 October 2022).
58. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *5*, 5–32. [\[CrossRef\]](#)
59. Satpathy, R.; Choudhury, T.; Satpathy, S.; Mohanty, S.; Zhang, X. Introduction to supervised learning. In *Data Analytics in Bioinformatics: A Machine Learning Perspective*, 1st ed.; John Wiley & Sons: Hoboken, NJ, USA, 2021; pp. 18–20.
60. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [\[CrossRef\]](#)
61. Marsland, S. Boosting. In *Machine Learning: An Algorithmic Perspective*, 2nd ed.; Taylor and Francis Group: Boca Raton, FL, USA; CRC Press: Oxford, NY, USA, 2015; pp. 268–273.
62. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurobot.* **2013**, *7*, 21. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Taunk, K.; Verma, S.; Swetapadma, A. A Brief Review of Nearest Neighbor Algorithm for Learning and Classification. In Proceedings of the International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 15–17 May 2019. [\[CrossRef\]](#)
64. Bafjaish, S. Comparative Analysis of Naive Bayesian Techniques in Health-Related for Classification Task. *J. Soft Comput. Data Min.* **2020**, *1*, 1–10. Available online: <https://penerbit.uthm.edu.my/ojs/index.php/jscdm/article/view/7144> (accessed on 24 October 2023).
65. Ranjitha, K.V. Classification and optimization scheme for text data using machine learning Naïve Bayes classifier. In Proceedings of the 2018 IEEE World Symposium on Communication Engineering (WSCE), Singapore, 28–30 December 2018; pp. 33–36.
66. Sulaiman, M.A. Evaluating Data Mining Classification Methods Performance in Internet of Things Applications. *J. Soft Comput. Data Min.* **2020**, *1*, 11–25. Available online: <https://penerbit.uthm.edu.my/ojs/index.php/jscdm/article/view/7127> (accessed on 13 August 2021).
67. Ghosh, S.; Dasgupta, A.; Swetapadma, A. A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification. In Proceedings of the 2019 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 21–22 February 2019; pp. 24–28.
68. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [\[CrossRef\]](#)
69. Alayande, S.; Adekunle, B. An overview and application of discriminant analysis in data analysis. *IOSR J. Math.* **2015**, *11*, 12–15.
70. Varoquaux, G.; Colliot, O. Evaluating machine learning models and their diagnostic value. In *Machine Learning for Brain Disorders*; Springer: New York, NY, USA, 2022; pp. 1–30.
71. Chiu, H.-Y.R.; Hwang, C.-K.; Chen, S.-Y.; Shih, F.-Y.; Han, H.-C.; King, C.-C.; Gilbert, J.R.; Fang, C.-C.; Oyang, Y.-J. Machine learning for emerging infectious disease field responses. *Sci. Rep.* **2022**, *12*, 328. [\[CrossRef\]](#)
72. Mondal, M.R.; Bharati, S.; Podder, P. Diagnosis of COVID-19 Using Machine Learning and Deep Learning: A Review. *Curr. Med. Imaging* **2021**, *17*, 1403–1418. [\[CrossRef\]](#)
73. Mayrose, H.; Bairy, G.M.; Sampathila, N.; Belurkar, S.; Saravu, K. Machine Learning-Based Detection of Dengue from Blood Smear Images Utilizing Platelet and Lymphocyte Characteristics. *Diagnostics* **2023**, *13*, 220. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Kamalov, F.; Elnagar, A.; Leung, H. Ensemble Learning with Resampling for Imbalanced Data. In *ICIC 2021: Intelligent Computing Theories and Application*; Lecture Notes in Computer Science; Huang, D.S., Jo, K.H., Li, J., Gribova, V., Hussain, A., Eds.; Springer: Cham, Switzerland, 2021; Volume 12837. [\[CrossRef\]](#)
75. Geetha, R.; Sivasubramanian, S.; Kaliappan, M.; Vimal, S.; Annamalai, S. Cervical Cancer Identification with Synthetic Minority Oversampling Technique and PCA Analysis using Random Forest Classifier. *J. Med. Syst.* **2019**, *43*, 286. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Chow, E.J.; Uyeki, T.M.; Chu, H.Y. The effects of the COVID-19 pandemic on community respiratory virus activity. *Nat. Rev. Microbiol.* **2023**, *21*, 195–210. [\[CrossRef\]](#)
77. Aloui, K.; Hamza, C.; Mefteh, K.; Hanen, S. Epidemiologic changes of Respiratory syncytial virus in the COVID-19 Era. *Med. Mal. Infect. Form.* **2022**, *1*, 109. [\[CrossRef\]](#)
78. Kandeel, A.; Fahim, M.; Deghedy, O.; Roshdy, W.H.; Khalifa, M.K.; El Shesheny, R.; Kandeil, A.; Naguib, A.; Afifi, S.; Mohsen, A.; et al. Resurgence of influenza and respiratory syncytial virus in Egypt following two years of decline during the COVID-19 pandemic: Outpatient clinic survey of infants and children. *BMC Public Health* **2022**, *23*, 1067. [\[CrossRef\]](#)
79. Barraza, M.F.O.; Fasce, R.A.; Nogareda, F.; Marcenac, P.; Mallegas, N.V.; Alister, P.B.; Loayza, S.; Chard, A.N.; Arriola, C.S.; Couto, P.; et al. Influenza Incidence and Vaccine Effectiveness During the Southern Hemisphere Influenza Season—Chile. *MMWR Morb. Mortal. Wkly.* **2022**, *71*, 1353–1358. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.