applied sciences

MDPI

*Article*

# Machine Learning for COVID-19 and Influenza Classification during Coexisting Outbreaks

**Iris Viana dos Santos Santana [1,†], Álvaro Sobrinho [2,*,†], Leandro Dias da Silva [1,†] and Angelo Perkusich [3,†]**

[1] Computing Institute, Federal University of Alagoas, Maceió 57072-260, Brazil; ivss@ic.ufal.br (I.V.d.S.S.); leandrodias@ic.ufal.br (L.D.d.S.)
[2] Center for Technological Innovation and Entrepreneurship, Federal University of the Agreste of Pernambuco, Garanhuns 55292-270, Brazil
[3] Virtus Research, Development and Innovation Center, Federal University of Campina Grande, Campina Grande 58429-000, Brazil; perkusic@virtus.ufcg.edu.br
* Correspondence: alvaro.alvares@ufape.edu.br; Tel.: +55-87-9-8149-3955
† These authors contributed equally to this work.

**Abstract:** This study compares the performance of machine learning models for selecting COVID-19 and influenza tests during coexisting outbreaks in Brazil, avoiding the waste of resources in healthcare units. We used COVID-19 and influenza datasets from Brazil to train the Decision Tree (DT), Multilayer Perceptron (MLP), Gradient Boosting Machine (GBM), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), K-Nearest Neighbors, Support Vector Machine (SVM), and Logistic Regression algorithms. Moreover, we tested the models using the 10-fold cross-validation method to increase confidence in the results. During the experiments, the GBM, DT, RF, XGBoost, and SVM models showed the best performances, with similar results. The high performance of tree-based models is relevant for the classification of COVID-19 and influenza because they are usually easier to interpret, positively impacting the decision-making of health professionals.

**Keywords:** COVID-19; influenza; machine learning



**Citation:** Viana dos Santos Santana, I.; Sobrinho, Á.; Silva, L.D.d.; Perkusich, A. Machine Learning for COVID-19 and Influenza Classification during Coexisting Outbreaks. *Appl. Sci.* **2023**, *13*, 11518. https://doi.org/10.3390/app132011518

Academic Editor: Stefano Silvestri and Francesco Gargiulo

Received: 22 August 2023
Revised: 16 October 2023
Accepted: 18 October 2023
Published: 20 October 2023



## 1. Introduction

The control of outbreaks of viral infectious diseases in Brazil presents a pertinent challenge, given the size of the population, population density, social habits, and constrained testing strategies with limited test availability [1]. This challenge is further amplified when simultaneous outbreaks of diseases occur, such as COVID-19 and influenza [2]. Therefore, studies are needed to assist in mitigating issues associated with concurrent outbreaks of such diseases. Due to their constrained testing resources, test prioritization is a pertinent public health strategy for low- and middle-income countries.

Machine Learning (ML) models can be a foundation for developing eHealth and mHealth systems [3–5]. These systems can support healthcare professionals and policymakers in test prioritization. To facilitate real-world clinical application and integration into the current clinical workflow, classification models can be made accessible, and attribute relevance information can be leveraged through web services for consumption by a healthcare system. These eHealth and mHealth systems should provide classification results to healthcare professionals through clear and concise graphical user interfaces. Thus, the direct interpretability of the ML models is crucial to enhance the confidence of healthcare professionals in the classification results [6]. For instance, healthcare systems of Brazilian public healthcare units can reuse models deployed by web services to prioritize scarce testing resources.

Amidst the COVID-19 pandemic, the challenge of testing resource scarcity became evident in numerous countries, such as Brazil [7]. Brazil's first COVID-19 case occurred

in March 2020, and over an extended period, the Ministry of Health reported a consistent increase in confirmed cases and fatalities. The Brazilian government reported over 30 million COVID-19 cases and over 664,000 fatalities. Considering Brazil's most recent data from the Ministry of Health, from January to September 2023, the unfortunate toll of COVID-19 resulted in over 12,000 individuals (https://covid.saude.gov.br/, accessed on 10 October 2023 ).

The limited number of tests becomes even more critical when concurrent outbreaks of viral infectious diseases (concomitant outbreaks) occur. During the most challenging pandemic phases, the Brazilian population faced at least two coexisting outbreaks of viral infectious diseases: COVID-19 and influenza. Therefore, it is pertinent to assist policymakers in formulating solutions to address concurrent outbreaks of viral infectious diseases (present and future).

This article extends our previous research [8] by presenting ML models to assist in test prioritization based on symptoms during concurrent outbreaks of COVID-19 and influenza in Brazil. To our knowledge, no prior studies are experimenting with Brazilian datasets for COVID-19 and influenza classification in this context. In the clinical scenario we envision, symptomatic patients present themselves at the hospital's testing site during a coexisting outbreak of COVID-19 and influenza. Before conducting tests for COVID-19 or influenza, which can be limited resources in certain countries, healthcare professionals can gather patient information as input data for an ML model. This approach can empower healthcare providers to make more informed decisions about which test to administer to specific patients, optimizing resource allocation and patient care.

We implemented ML models that rely on supervised learning, employing the following algorithms: Decision Tree (DT), Multilayer Perceptron (MLP), Gradient Boosting Machine (GBM), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Logistic Regression (LR). To address our multi-class problem, which involves distinguishing between COVID-19, non-COVID-19, influenza, and non-influenza cases, we conducted training and testing of algorithms using datasets containing demographic attributes and symptom information. This approach reduces the necessity for expensive exams (e.g., computed tomography scans).

The remainder of this article is structured as follows. Section 2 discusses the materials and methods, encompassing data collection, preprocessing, attribute selection, validation procedures, and details about the algorithms employed. Section 3 presents and discusses our results. Section 4 concludes from our findings and presents future research directions.

## 2. Related Works

Studies on viral infection outbreaks are relevant for public administration (for instance, in the context of surveillance systems) from a diagnostic standpoint. For instance, Son et al. [9] used a South Korean time series of influenza incidence to detect initial outbreaks, aiming to assist in policy formulation for control. Indeed, a pandemic presents a challenging scenario. In another study, Kumar [10] analyzed and enhanced the monitoring of COVID-19 in India by cluster analysis, offering insights into how the disease affected Indian states. The authors considered a total of twenty-eight states and eight Indian territories.

The existing literature also offers studies that explore the concurrent outbreaks of COVID-19 and influenza. For instance, Aftab et al. [11] experimented with deep-learning models for COVID-19 and influenza classification during coexisting outbreaks based on chest X-ray images. Li et al. [12] applied an XGBoost model to detect patients with both COVID-19 and influenza. The authors used clinical data, including laboratory test results, to train the XGBoost and baseline algorithms.

In another study, Zhou et al. [13] used data from patients at Zhongnan Hospital of Wuhan University to implement an XGBoost model. The authors also considered symptoms and laboratory test results when training the algorithm.

Furthermore, Elbasi et al. [14] addressed the classification of influenza and COVID-19 by employing the Bayes network, naive Bayes, locally weighted learning, MLP, and RF

algorithms. Like the studies above, the authors used demographic data, symptoms, risk factors, and laboratory test results for training the ML models.

Phu et al. [15] compared the XGBoost and RF algorithms for classifying influenza and COVID-19, considering symptoms and laboratory test results. Their findings indicated that the XGBoost model outperformed RF.

Nevertheless, the need for expensive laboratory test results could limit the practical application of the previously proposed models, particularly when considering low- and middle-income populations. Thus, using Brazilian datasets, our study investigates the performance of ML models with different characteristics (e.g., tree-based and distance-based approaches) by only considering demographic data and reported symptoms.

## 3. Materials and Methods

This study considers data preprocessing, creation of new datasets, attribute selection, 10-fold cross-validation, statistical comparisons, and attribute importance. The following sections present an overview of these steps.

### *3.1. Data Collection and Preprocessing*

Data collection is not considered a contribution of this study. The raw data used in this study was collected by the public health agency of Campina Grande, located in the Paraíba State of Northeast Brazil. This public agency receives information from all COVID-19 tests conducted in Campina Grande [8]. To ensure patient privacy, the agency's staff removed patient identification details, and the de-identified data were then made available for reuse in this study. The raw dataset contains various categorical features, including information about health professionals, security professionals, ethnicity, test types, symptoms (e.g., fever, sore throat, dyspnea, olfactory disorders, cough, coryza, taste disorders, and headache), additional symptoms, test results, comorbidities, test status, and symptom descriptions. Therefore, raw data from 55,676 Brazilian individuals were preprocessed to establish new datasets containing information on symptomatic patients tested for COVID-19.

Additionally, we gathered data from a sample of 14,570 Brazilian individuals, which included information about symptomatic patients who underwent testing for influenza. As for COVID-19, the tests encompassed Reverse Transcription Polymerase Chain Reaction (RT-PCR) and rapid tests (antibody and antigen). We sourced information from the OpenDatasus platform (https://opendatasus.saude.gov.br/dataset/srag-2021-a-2023, accessed on 10 October 2023) the Brazilian Ministry of Health provided for this dataset.

We performed data preprocessing using the Python programming language. During this preprocessing stage, we applied string-matching algorithms to address inconsistencies in the raw dataset. One notable inconsistency we addressed was the presence of empty columns related to symptoms, as the same symptoms were available within a column designated for general symptom descriptions. We also excluded certain instances from the initial samples based on our predefined exclusion criteria.

For the COVID-19 dataset, these criteria encompassed patients with incomplete tests or undefined final classifications (12,929 instances), duplicated instances (12,929 instances), outliers resulting from input errors (10,408 instances), test types that were not RT-PCR or rapid (771 instances), individuals with undefined gender (27 instances), and asymptomatic patients (11,269 instances). The exclusion of asymptomatic patients was necessary because the algorithms relied on demographic characteristics and symptom data for accurate processing. We applied the same criteria to the influenza dataset. For example, the raw dataset, which was unbalanced, contained 5954 instances. After filtering and removal, the dataset was reduced to 4212 instances.

We filtered the data to generate new datasets, including patients tested for COVID-19 and influenza. This process ensures that a patient tested positive for COVID-19 also tested negative for influenza. However, we acknowledge that a patient can be infected with both diseases simultaneously, even though this scenario is more uncommon. We did not consider

this scenario in our study. The following attributes were considered: respiratory distress, vomiting, saturation, fatigue, diarrhea, abdominal pain, gender, healthcare professional status, fever, sore throat, dyspnea, olfactory disorder, cough, runny nose, taste disorder, and headache.

We formulated six preprocessed datasets by combining COVID-19 and influenza data: unbalanced RT-PCR, balanced RT-PCR, unbalanced rapid, balanced rapid, both unbalanced, and both balanced. Thus, during the preprocessing, we oversampled the data by applying the near-miss technique [16]. In the datasets, the numbers 0 and 1 denote positive and negative results for COVID-19, while 2 and 3 represent positive and negative outcomes for influenza. In the balanced RT-PCR datasets, each class contains 916 samples, the rapid test datasets have 646 samples for each class, and the combined datasets of both tests contain 1564 samples. For the unbalanced sets, the RT-PCR dataset consists of 0 (916 samples), 1 (1863 samples), 2 (1502 samples), and 3 (1423 samples). The rapid test dataset includes class 0 (648 samples), 1 (16,594 samples), 2 (691 samples), and 3 (646 samples). In the combined dataset of both tests, class 0 has 1564 samples, class 1 has 18,457 samples, class 2 has 2106 samples, and class 3 has 2148 samples.

### 3.2. Attribute Selection

We applied the chi-squared test to the new datasets to assist in attribute selection with a $p < 0.01$ threshold, examining attribute relevance for classification tasks through dependence and independence relationships [17]. The chi-squared test for independence was employed to compare four variables within a contingency table, determining whether they are related.

### 3.3. Validation Method and Algorithms

We used the 10-fold cross-validation method with five repetitions to validate the ML models MLP, GBM, DT, RF, XGBoost, KNN, SVM, and LR (weak/strong regularization). An MLP is a feedforward neural network, meaning data travels in a single direction [18]. This model comprises one (or more) hidden neurons between the layers related to inputs and outputs.

In contrast, the GBM is a decision tree with a fixed size that uses a boosting strategy [19]. This algorithm features integrated attribute selection, producing an estimation, approximation, or the function denoted as $F^*(x)$, which maps the input $x$ to the output $y$ while minimizing the expected value by using a loss function $L(y, F(x))$ across the joint distribution [20].

A DT algorithm commonly applies a divide-and-conquer strategy to construct a directed acyclic graph, where rule splitting is determined by maximizing information gain [21]. DT algorithms such as C4.5 include internal attribute selection, and the information gain is influenced by the concept of entropy, which quantifies the uncertainty or randomness associated with a discrete random variable. DT offers the advantage of straightforward result interpretation by following the decision rules of a single tree [22–24].

Similarly, the RF relies on classification and regression tree principles while following specific guidelines for tree growth, combination, self-testing, and post-processing [25]. The algorithm includes an embedded attribute selection mechanism, evaluated using the Gini impurity criterion index. RF also facilitates a straightforward interpretation of results from the individual trees within the ensemble.

Relying on a different approach, KNN is a distance-based algorithm that classifies new instances using the distance from neighbor instances [26]. Given an instance as a point in space, KNN computes the distance between two points.

SVM is an algorithm designed to handle binary data using a linear separator to maximize the distance between data points. The algorithm considers concepts such as the separation hyperplane, maximum margin hyperplane, and soft margin [18].

Lastly, LR extends linear regression, assessing the connections between variables in probabilistic classifications. The algorithm employs a sigmoid function to model these

relationships and predict the probability of class membership [27]. Regularization can also be employed to prevent overfitting. We used the LR algorithm as a baseline to compare the linear model with the above ML approaches.

Moreover, we calculated the mean results for classification metrics, including precision, accuracy, recall, Brier score, Receiver Operating Characteristic Curve (ROC), and Area Under the Precision-Recall (PR) curve. We used the random search method to fine-tune the hyperparameters of the algorithms, aiming to enhance their performance. Furthermore, we analyzed recall-related outcomes using the Friedman and Nemenyi statistical tests to improve the comparisons among the ML models [28]. The Friedman test is a valuable statistical method for identifying differences between models. The Nemenyi test is pertinent for grouping classification models based on assessing differences through many comparisons. We determined the Critical Difference (CD) between the ML models by employing the Nemenyi test, with a significance level set at $\alpha = 0.1$. If the performance differences among models are within an interval more minor than the CD, it suggests that the models are indistinguishable from each other [8].

### 3.4. Attribute Importance

Finally, we performed attribute ranking for each ML model with the best performance through the permutation-based attribute importance method, which provided mean importance and Standard Deviation (SD) as the evaluation criteria. The attribute ranking relied on the permutation feature importance method to gauge the importance of an attribute by evaluating the decrease in the model's score, thereby assessing the degree of reliance of the model on that particular attribute [29]. Our discussions center around the top five attributes with the highest importance.

## 4. Results and Discussion

The COVID-19 and influenza datasets were merged to form a unified dataset for implementing the models. As mentioned in Section 2, the numbers 0 and 1 indicated positive and negative cases of COVID-19, while 2 and 3 indicated positive and negative cases of influenza. The balanced RT-PCR datasets contained 916 instances for each class, the rapid test datasets contained 646 instances for each class, and both tests combined included 1564 instances for each class. The unbalanced RT-PCR datasets contained class 1 (1863), 2 (1502), 3 (1423), and 0 (916); the rapid test datasets contained class 1 (16,594), 2 (691), 0 (648), and 3 (646); and both tests combined included class 1 (18,457), 3 (2148), 2 (2106), and 0 (1564).

The tree-based models, considering the combination of COVID-19 and influenza, are among those that exhibited superior outcomes. Table 1 presents the results of the 10-fold cross-validation. We computed precision, recall, accuracy, and Brier score. LR and LRR denote weak and strong regularization models, respectively.

Tables 2 and 3 present the mean importance and SD for attributes considering the tree-based models and MLP and SVM, respectively. Table 4 displays the top five most significant attributes for test prioritization. Such analyses consider the imbalanced datasets.

Throughout the experiments, attributes were not removed based on the chi-squared test results, as all attributes exhibited dependence. To provide a more detailed illustration of the outcomes, for instance, in Figures 1 and 2, the average results for the ROC and PR curves using cross-validation for the decision tree model are depicted, employing both the balanced and unbalanced datasets of both tests, respectively. As mentioned, we conducted a 10-fold cross-validation five times to enhance our confidence in the results. We presented the results for each of the four classes of our multi-class problem.

**Table 1.** Performance of the classification models.

| Database | Model | Precision | Recall | Accuracy | Brier Score |
|---|---|---|---|---|---|
| PCR Imbalanced (Balanced) | MLP | 82.44 (82.72) | 82.72 (82.26) | 82.37 (82.25) | 0.088 (0.088) |
| | GBM | 82.59 (82.48) | 82.92 (82.14) | 82.54 (82.14) | 0.08 (0.089) |
| | RF | 83.04 (82.96) | 83.12 (82.47) | 82.84 (82.47) | 0.085 (0.087) |
| | DT | 82.46 (82.37) | 82.60 (81.76) | 82.27 (81.76) | 0.088 (0.091) |
| | XGBoost | 82.73 (82.67) | 82.99 (82.27) | 82.65 (82.27) | 0.086 (0.088) |
| | KNN | 82.16 (82.35) | 82.22 (81.51) | 82.07 (81.61) | 0.089 (0.092) |
| | SVM | 82.23 (81.93) | 82.55 (81.57) | 82.27 (81.57) | 0.088 (0.092) |
| | LRR | 71.76 (71.65) | 70.13 (70.78) | 71.37 (70.78) | 0.143 (0.145) |
| | LR | 72.29 (72.29) | 70.85 (71.38) | 72.02 (71.38) | 0.139 (0.143) |
| Rapid Imbalanced (Balanced) | MLP | 85.30 (81.40) | 79.41 (81.40) | 96.44 (81.40) | 0.171 (0.092) |
| | GBM | 85.61 (81.14) | 79.40 (81.15) | 96.46 (81.15) | 0.017 (0.094) |
| | RF | 86.38 (81.50) | 79.25 (81.56) | 96.53 (81.56) | 0.017 (0.092) |
| | DT | 85.42 (80.99) | 79.61 (80.93) | 96.53 (80.93) | 0.017 (0.095) |
| | XGBoost | 85.76 (80.89) | 79.32 (80.96) | 96.47 (80.96) | 0.017 (0.095) |
| | KNN | 87.81 (80.52) | 78.93 (79.90) | 96.56 (79.90) | 0.017 (0.100) |
| | SVM | 86.12 (80.97) | 79.11 (81.11) | 96.49 (81.11) | 0.017 (0.094) |
| | LRR | 73.32 (68.78) | 55.60 (68.98) | 93.24 (68.98) | 0.033 (0.155) |
| | LR | 72.42 (70.70) | 58.37 (70.24) | 93.50 (70.24) | 0.032 (0.148) |
| Both imbalanced (Balanced) | MLP | 85.93 (80.95) | 71.11 (80.47) | 90.73 (80.95) | 0.046 (0.097) |
| | GBM | 85.89 (81.14) | 71.16 (80.64) | 90.70 (80.64) | 0.046 (0.096) |
| | RF | 86.85 (81.50) | 71.63 (80.95) | 90.96 (80.95) | 0.045 (0.095) |
| | DT | 85.60 (80.74) | 71.66 (80.13) | 90.76 (80.13) | 0.046 (0.099) |
| | XGBoost | 86.07 (81.28) | 71.39 (80.77) | 90.79 (80.77) | 0.046 (0.096) |
| | KNN | 82.23 (80.33) | 66.15 (79.58) | 88.98 (79.58) | 0.055 (0.102) |
| | SVM | 86.43 (80.54) | 70.95 (80.11) | 90.77 (80.11) | 0.046 (0.099) |
| | LRR | 68.87 (70.16) | 52.82 (70.10) | 85.54 (70.10) | 0.072 (0.149) |
| | LR | 71.16 (71.24) | 55.39 (71.22) | 86.29 (71.22) | 0.068 (0.143) |

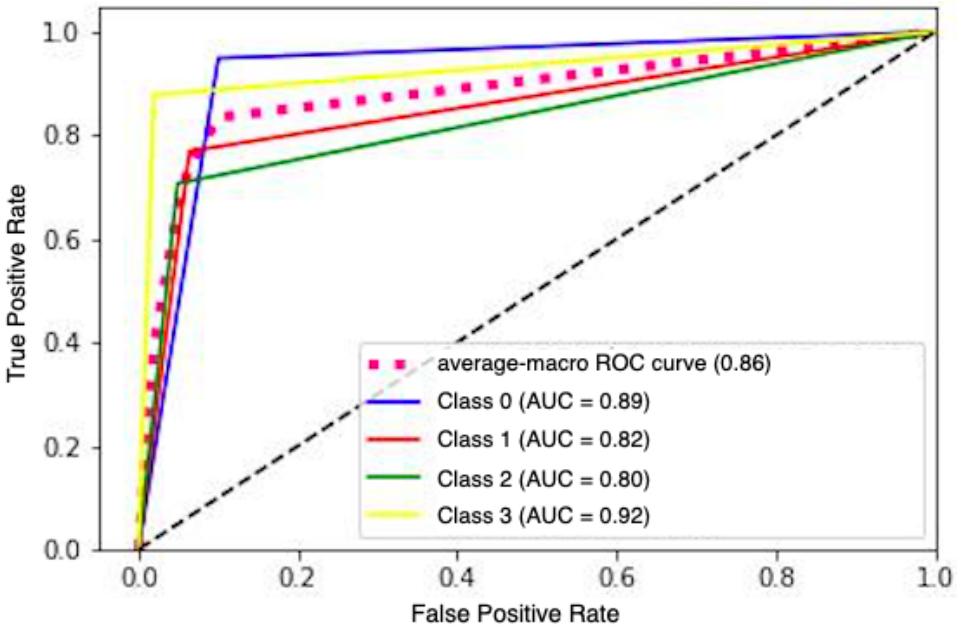


**Figure 1.** Average ROC curve for each class of the DT model using both balanced tests.

**Table 2.** Mean importance and SD for attributes in tree-based classification models using the imbalanced datasets.

| Dataset | Feature | GBM | DT | RF | XGBoost |
|---|---|---|---|---|---|
| PCR | Respiratory distress | 0.141 (0.008) | 0.155 (0.008) | 0.156 (0.008) | 0.144 (0.008) |
| | Vomit | 0.037 (0.004) | 0.036 (0.004) | 0.035 (0.004) | 0.037 (0.004) |
| | Saturation | 0.149 (0.007) | 0.166 (0.008) | 0.161 (0.007) | 0.154 (0.008) |
| | Fatigue | 0.055 (0.006) | 0.047 (0.005) | 0.057 (0.006) | 0.058 (0.006) |
| | Diarrhea | 0.028 (0.003) | 0.024 (0.003) | 0.027 (0.003) | 0.028 (0.003) |
| | Abdominal pain | 0.013 (0.003) | 0.006 (0.002) | 0.011 (0.002) | 0.013 (0.002) |
| | Gender | 0.138 (0.007) | 0.133 (0.007) | 0.137 (0.007) | 0.137 (0.007) |
| | Health professional | 0.029 (0.003) | 0.026 (0.002) | 0.029 (0.003) | 0.031 (0.003) |
| | Fever | 0.235 (0.009) | 0.230 (0.008) | 0.234 (0.008) | 0.230 (0.008) |
| | Sore throat | 0.077 (0.006) | 0.073 (0.005) | 0.075 (0.005) | 0.079 (0.005) |
| | Dyspnoea | 0.098 (0.007) | 0.095 (0.006) | 0.093 (0.007) | 0.092 (0.007) |
| | Smell disorder | 0.009 (0.002) | 0.017 (0.002) | 0.004 (0.002) | 0.010 (0.002) |
| | Cough | 0.105 (0.007) | 0.101 (0.007) | 0.102 (0.007) | 0.102 (0.007) |
| | Runny nose | 0.007 (0.002) | 0.006 (0.002) | 0.007 (0.002) | 0.005 (0.002) |
| | Taste disorder | 0.019 (0.002) | 0.014 (0.002) | 0.009 (0.002) | 0.010 (0.002) |
| | Headache | 0.012 (0.002) | 0.012 (0.002) | 0.013 (0.002) | 0.008 (0.002) |
| Rapid | Respiratory distress | 0.109 (0.007) | 0.113 (0.007) | 0.156 (0.008) | 0.121 (0.008) |
| | Vomit | 0.019 (0.004) | 0.017 (0.003) | 0.015 (0.004) | 0.020 (0.004) |
| | Saturation | 0.140 (0.008) | 0.194 (0.008) | 0.165 (0.008) | 0.154 (0.007) |
| | Fatigue | 0.056 (0.005) | 0.072 (0.006) | 0.055 (0.005) | 0.055 (0.005) |
| | Diarrhea | 0.013 (0.003) | 0.012 (0.003) | 0.009 (0.002) | 0.015 (0.004) |
| | Abdominal pain | 0.004 (0.002) | 0.008 (0.002) | 0.003 (0.001) | 0.007 (0.002) |
| | Gender | 0.096 (0.009) | 0.096 (0.010) | 0.100 (0.009) | 0.095 (0.008) |
| | Health professional | 0.013 (0.003) | 0.014 (0.002) | 0.015 (0.002) | 0.018 (0.003) |
| | Fever | 0.148 (0.008) | 0.158 (0.009) | 0.148 (0.007) | 0.152 (0.007) |
| | Sore throat | 0.080 (0.006) | 0.073 (0.006) | 0.072 (0.006) | 0.075 (0.006) |
| | Dyspnoea | 0.130 (0.008) | 0.144 (0.009) | 0.136 (0.008) | 0.136 (0.008) |
| | Smell disorder | 0.083 (0.005) | 0.096 (0.006) | 0.087 (0.006) | 0.084 (0.005) |
| | Cough | 0.075 (0.007) | 0.086 (0.008) | 0.072 (0.007) | 0.067 (0.008) |
| | Runny nose | 0.054 (0.004) | 0.046 (0.003) | 0.039 (0.003) | 0.048 (0.003) |
| | Taste disorder | 0.042 (0.004) | 0.036 (0.004) | 0.023 (0.004) | 0.039 (0.004) |
| | Headache | 0.047 (0.004) | 0.053 (0.004) | 0.045 (0.004) | 0.055 (0.004) |
| Both | Respiratory distress | 0.140 (0.004) | 0.148 (0.004) | 0.147 (0.004) | 0.143 (0.005) |
| | Vomit | 0.035 (0.003) | 0.041 (0.004) | 0.039 (0.004) | 0.038 (0.003) |
| | Saturation | 0.171 (0.006) | 0.191 (0.005) | 0.190 (0.006) | 0.184 (0.006) |
| | Fatigue | 0.053 (0.003) | 0.062 (0.003) | 0.057 (0.003) | 0.055 (0.003) |
| | Diarrhea | 0.017 (0.002) | 0.017 (0.002) | 0.015 (0.002) | 0.018 (0.002) |
| | Abdominal pain | 0.018 (0.002) | 0.010 (0.002) | 0.010 (0.002) | 0.018 (0.002) |
| | Gender | 0.123 (0.005) | 0.126 (0.006) | 0.121 (0.005) | 0.119 (0.005) |
| | Health professional | 0.020 (0.002) | 0.016 (0.002) | 0.019 (0.002) | 0.020 (0.002) |
| | Fever | 0.182 (0.006) | 0.187 (0.006) | 0.183 (0.007) | 0.179 (0.007) |
| | Sore throat | 0.084 (0.004) | 0.087 ( 0.004) | 0.082 (0.004) | 0.083 (0.004) |
| | Dyspnoea | 0.103 (0.005) | 0.108 (0.005) | 0.103 (0.005) | 0.107 (0.005) |
| | Smell disorder | 0.047 (0.003) | 0.053 (0.003) | 0.039 (0.003) | 0.048 (0.003) |
| | Cough | 0.086 (0.004) | 0.088 (0.004) | 0.087 (0.005) | 0.084 (0.004) |
| | Runny nose | 0.017 (0.001) | 0.016 (0.001) | 0.017 (0.002) | 0.016 (0.001) |
| | Taste disorder | 0.024 (0.003) | 0.027 (0.003) | 0.019 (0.003) | 0.024 (0.003) |
| | Headache | 0.035 (0.003) | 0.033 (0.002) | 0.028 (0.003) | 0.031 (0.003) |

**Table 3.** Mean importance and SD for attributes in MLP and SVM models using the imbalanced datasets.

| Dataset | Feature | MLP | SVM |
|---|---|---|---|
| RT-PCR | Respiratory distress | 0.139 (0.009) | 0.135 (0.008) |
| | Vomit | 0.041 (0.005) | 0.039 (0.004) |
| | Saturation | 0.146 (0.008) | 0.152 (0.007) |
| | Fatigue | 0.056 (0.005) | 0.059 (0.005) |
| | Diarrhea | 0.025 (0.003) | 0.028 (0.003) |
| | Abdominal pain | 0.015 (0.003) | 0.015 (0.003) |
| | Gender | 0.137 (0.007) | 0.125 (0.006) |
| | Health professional | 0.043 (0.003) | 0.021 (0.002) |
| | Fever | 0.230 (0.008) | 0.219 (0.008) |
| | Sore throat | 0.081 (0.007) | 0.069 (0.006) |
| | Dyspnoea | 0.100 (0.006) | 0.080 (0.006) |
| | Smell disorder | 0.013 (0.003) | 0.010 (0.002) |
| | Cough | 0.105 (0.007) | 0.098 (0.007) |
| | Runny nose | 0.010 (0.002) | 0.003 (0.002) |
| | Taste disorder | 0.018 (0.002) | 0.014 (0.002) |
| | Headache | 0.020 (0.002) | 0.007 (0.002) |
| Rapid | Respiratory distress | 0.090 (0.007) | 0.092 (0.007) |
| | Vomit | 0.019 (0.004) | 0.019 (0.004) |
| | Saturation | 0.132 (0.007) | 0.125 (0.007) |
| | Fatigue | 0.067 (0.005) | 0.061 (0.006) |
| | Diarrhea | 0.018 (0.003) | 0.017 (0.004) |
| | Abdominal pain | 0.011 (0.003) | 0.007 (0.002) |
| | Gender | 0.090 (0.009) | 0.076 (0.007) |
| | Health professional | 0.017 (0.003) | 0.016 (0.003) |
| | Fever | 0.141 (0.007) | 0.143 (0.007) |
| | Sore throat | 0.064 (0.006) | 0.062 (0.006) |
| | Dyspnoea | 0.129 (0.008) | 0.118 (0.007) |
| | Smell disorder | 0.074 (0.005) | 0.075 (0.005) |
| | Cough | 0.071 (0.007) | 0.068 (0.008) |
| | Runny nose | 0.041 (0.003) | 0.047 (0.003) |
| | Taste disorder | 0.032 (0.004) | 0.034 (0.003) |
| | Headache | 0.049 (0.005) | 0.044 (0.004) |
| Both | Respiratory distress | 0.123 (0.005) | 0.123 (0.005) |
| | Vomit | 0.038 (0.003) | 0.035 (0.004) |
| | Saturation | 0.146 (0.005) | 0.157 (0.006) |
| | Fatigue | 0.052 (0.003) | 0.047 (0.003) |
| | Diarrhea | 0.017 (0.002) | 0.016 (0.002) |
| | Abdominal pain | 0.020 (0.002) | 0.018 (0.002) |
| | Gender | 0.122 (0.005) | 0.109 (0.005) |
| | Health professional | 0.025 (0.002) | 0.018 (0.002) |
| | Fever | 0.180 (0.006) | 0.173 (0.006) |
| | Sore throat | 0.083 (0.004) | 0.083 (0.004) |
| | Dyspnoea | 0.108 (0.005) | 0.092 (0.005) |
| | Smell disorder | 0.058 (0.003) | 0.040 (0.002) |
| | Cough | 0.089 (0.005) | 0.079 (0.005) |
| | Runny nose | 0.023 (0.02) | 0.017 (0.001) |
| | Taste disorder | 0.030 (0.003) | 0.022 (0.003) |
| | Headache | 0.048 (0.003) | 0.028 (0.003) |

**Table 4.** The five most relevant attributes for test prioritization when dealing with imbalanced datasets.

| Dataset | Model | Top 1 | Top 2 | Top 3 | Top 4 | Top 5 |
|---|---|---|---|---|---|---|
| PCR | MLP | Fever | Saturation | Respiratory distress | Gender | Cough |
| | GBM | Fever | Saturation | Respiratory distress | Gender | Cough |
| | RF | Fever | Saturation | Gender | Respiratory distress | Cough |
| | DT | Saturation | Fever | Gender | Cough | Respiratory distress |
| | XGBoost | Saturation | Health professional | Respiratory distress | Fever | Fatigue |
| | SVM | Fever | Saturation | Respiratory distress | Gender | Cough |
| Rapid | MLP | Fever | Saturation | Dyspnoea | Respiratory distress | Gender |
| | GBM | Saturation | Dyspnoea | Fever | Smell disorder | Respiratory distress |
| | RF | Saturation | Fever | Dyspnoea | Fatigue | Gender |
| | DT | Fever | Saturation | Gender | Respiratory distress | Smell disorder |
| | XGBoost | Saturation | Smell disorder | Runny nose | Headache | Respiratory distress |
| | SVM | Fever | Saturation | Dyspnoea | Respiratory distress | Gender |
| Both | MLP | Fever | Saturation | Respiratory distress | Gender | Dyspnoea |
| | GBM | Diarrhea | Abdominal pain | Taste disorder | Headache | Runny nose |
| | RF | Abdominal pain | Diarrhea | Runny nose | Headache | Health professional |
| | DT | Health professional | Abdominal pain | Headache | Taste disorder | Runny nose |
| | XGBoost | Taste disorder | Cough | Gender | Dyspnoea | Sore throat |
| | SVM | Fever | Saturation | Respiratory distress | Gender | Dyspnoea |

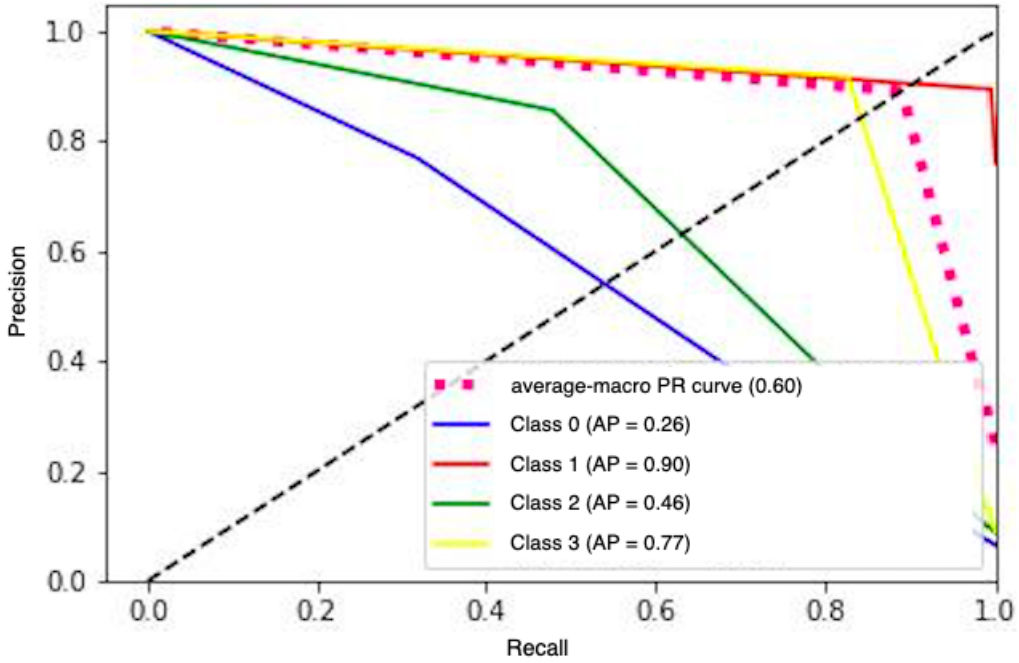


**Figure 2.** Average precision-recall curve for each class of the DT model using both unbalanced tests.

The ROC and PR curves were computed using five random folds for each class. The tree-based models exhibited the highest Average Precision (AP) values, ranging from 58% to 86%, using the balanced dataset with both tests. Additionally, AP values ranged from 30% to 92% using the unbalanced dataset with both types of tests.

Afterward, using the Friedman and Nemenyi tests increased confidence in validating the ML models. They were compared using the six COVID-19 and influenza datasets. The comparison predominantly concentrates on recall outcomes due to the significant adverse effects of false negatives in COVID-19 and influenza applications. Figure 3 depicts the recall results for the employed datasets.

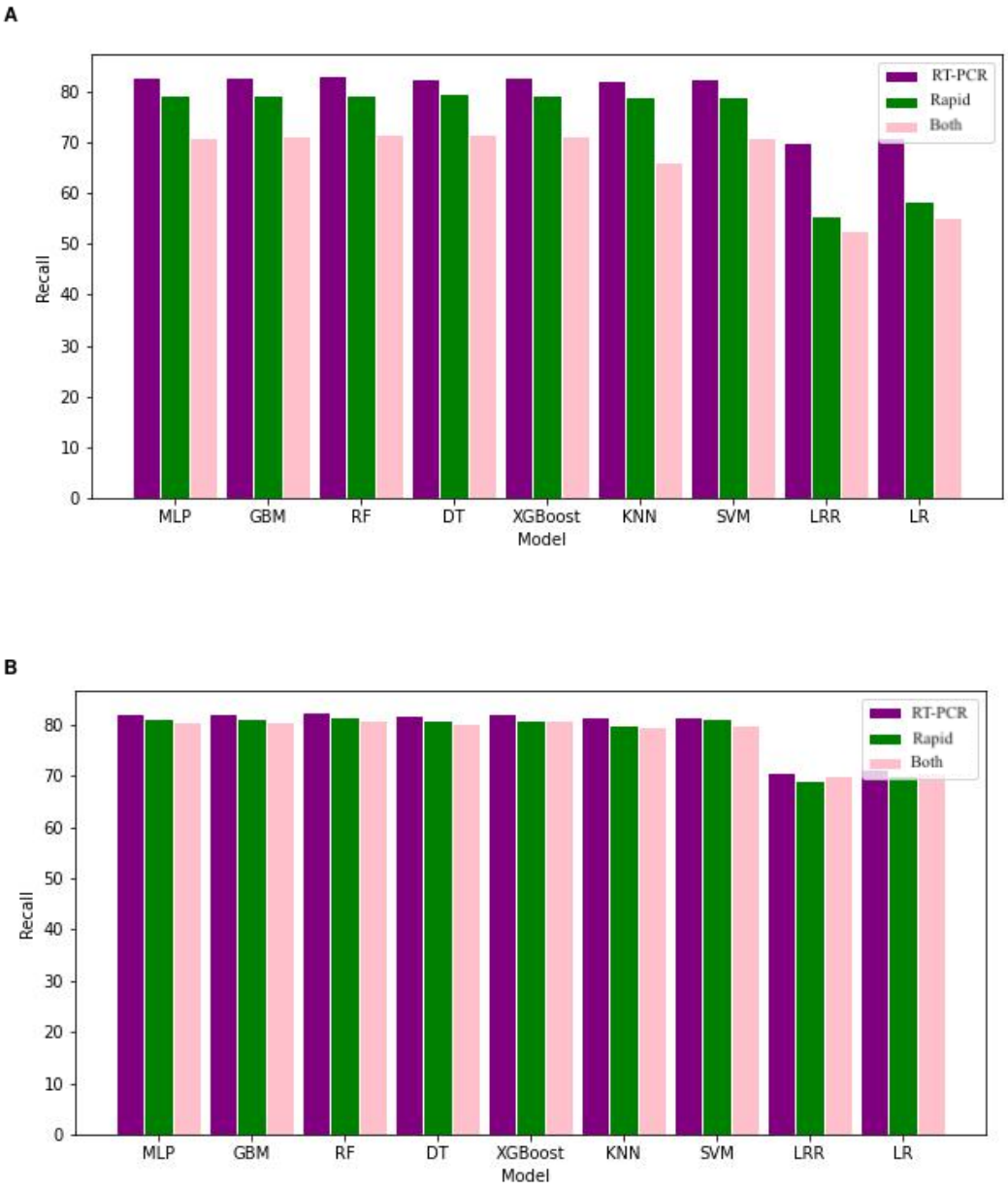


**Figure 3.** (**A**) Mean recall for the models using unbalanced datasets for RT-PCR, rapid, and both types. (**B**) Mean recall for the models using balanced datasets for RT-PCR, rapid, and both types.

The null hypothesis and the results were as follows: for unbalanced RT-PCR (t = 217.363), balanced RT-PCR (t = 231.788), unbalanced rapid (t = 234.098), balanced rapid (t = 188.243), unbalanced both datasets (t = 221.810), and balanced both datasets (t = 253.074). The results suggest that the difference in mean recall was likely statistically significant ($p < 0.05$). Additionally, depending on the dataset, MLP and GBM appeared to be statistically indistinguishable, as were DT, RF, XGBoost, KNN, and SVM.

Using the permutation-based attribute importance method across the six datasets, we ranked the most significant five attributes for the ML models, demonstrating the highest performance. Fever and oxygen saturation symptoms displayed higher mean importance values in the case of the balanced datasets for RT-PCR and rapid tests. However, the symptoms with mean importance values appeared more diverse when both tests were balanced.

As a result, the preprocessing of raw datasets facilitated the implementation, validation, and comparison of classification models with diverse characteristics, including using neural layers, tree ensembles, and data distance computation. This preprocessing also led

to the public availability of patient data, including individuals tested as symptomatic using RT-PCR and rapid tests, as referenced in [8].

We conducted training and testing of the algorithms using both unbalanced and balanced datasets to improve data representativeness. When considering test-type grouping, the best classification metric results were achieved in both unbalanced and balanced scenarios for RT-PCR and rapid tests. While the classification model performances were similar for RT-PCR and rapid test scenarios, the RT-PCR testing scenario holds greater clinical relevance due to the high confidence associated with RT-PCR testing. The accuracy of RT-PCR testing enhances diagnostic confidence even if the patient has been tested in the early days after the onset of symptoms.

The recall metric is relevant in our context because of the adverse consequences of false negatives in clinical practice. We improved the quality of comparisons between ML models using the Friedman and Nemenyi tests, which relied on the recall performance across the six datasets.

The tree-based classification models examined in this study demonstrated superior performance and were grouped based on their classification metric outcomes and statistical test results. This observation is of particular importance because tree-based models are highly interpretable, which can positively influence the decision-making process of healthcare professionals. In clinical practice, the acceptance of ML-based systems increases when healthcare practitioners can easily comprehend and interpret the outputs of classification models to understand the decision-making logic, as referenced in [30].

Additionally, it is important to acknowledge that in a real-world scenario, the presence of asymptomatic patients may be seen as a limitation in the applicability of ML models [31]. However, in the context of this study, its relevance persists due to the presence of symptomatic cases that demand the attention of healthcare professionals and government authorities. Evaluating symptomatic patients remains crucial to avoid the inadvertent overuse of testing resources, especially in the face of concurrent disease outbreaks in Brazil caused by other viral infections (e.g., COVID-19 and influenza). Certain viral infections can present with similar symptoms, making it challenging for healthcare professionals to decide the appropriate type of testing needed.

A symptomatic patient with limited symptoms can pose challenges for ML models. However, attribute ranking and additional information, such as whether the patient has had contact with infected individuals, are valuable factors to supplement ML models. They provide additional context and data that can aid healthcare professionals and policymakers in making informed decisions.

Another limitation of this study is the number of ML models experimented with. However, we addressed this limitation by considering a set of well-established algorithms that cover various approaches, including tree-based models, linear regression, statistical learning, distance-based methods, and neural concepts.

## 5. Conclusions

The results emphasize the importance of employing ML models for test prioritization in Brazil during coexisting COVID-19 and influenza outbreaks, mainly focusing on non-expensive input data. The elevated performance of tree-based ML models holds significance for the healthcare domain due to their high interpretability reported in recent literature, which positively influences the final decision-making process of healthcare professionals.

Therefore, tree-based models have been identified as the most suitable ML models when considering the ease of interpretation and performance criterion. They can effectively aid in prioritizing the testing of symptomatic patients. The relevance of utilizing symptoms that do not require costly tests is evident, for instance, in underserved and hard-to-reach communities. These communities usually depend on public services to conduct expensive exams, which may only sometimes be promptly available.

Our experiments have demonstrated the viability of employing ML models to aid in prioritizing testing when concurrent outbreaks of COVID-19 and influenza occur. These

ML models can be seamlessly integrated into the clinical practice workflow at testing sites, enhancing the efficiency and effectiveness of the testing process. However, it is important to note that one limitation of our model is that it does not account for the scenario where a patient could be simultaneously infected with both COVID-19 and influenza. We recognize this as a significant aspect for future research and further development of our model.

Moreover, the solution proposed in this study holds the potential for scalability to decision support systems, considering the high number of existing viral infectious diseases. As a future endeavor, the intention is to develop a clinical decision support system based on the proposed approach, utilizing web technologies. Additionally, usability tests are planned, adhering to established standards in the literature, for the developed system to assess user-friendliness and perception, considering the potential diverse target audience for this type of system.

**Author Contributions:** Conceptualization, I.V.d.S.S., Á.S., L.D.d.S. and A.P.; Methodology, I.V.d.S.S.; Validation, I.V.d.S.S.; Investigation, I.V.d.S.S.; Writing—original draft, I.V.d.S.S., Á.S., L.D.d.S. and A.P.; Writing—review & editing, Á.S., L.D.d.S. and A.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** The APC was funded by the VIRTUS Research, Development & Innovation Center, Federal University of Campina Grande (Paraíba/Brazil).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available in OpenDatasus at [https://opendatasus.saude.gov.br/dataset/srag-2021-a-2023] and Mendeley Data at [https://doi.org/10.17632/b7zcgmmwx4.3].

**Acknowledgments:** We would like to thank VIRTUS Research, Development & Innovation Center, Federal University of Campina Grande (Paraíba/Brazil) and Alagoas Research Foundation for supporting this research.

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

## References

1. Monteiro de Oliveira, M.; Fuller, T.L.; Brasil, P.; Gabaglia, C.R.; Nielsen-Saines, K. Controlling the COVID-19 pandemic in Brazil: A challenge of continental proportions. *Nat. Med.* **2020**, *26*, 1505–1506. [CrossRef] [PubMed]
2. Konala, V.M.; Adapa, S.; Gayam, V.; Naramala, S.; Daggubati, S.R.; Kammari, C.B.; Chenna, A. Co-infection with Influenza A and COVID-19. *Eur. J. Case Rep. Intern. Med.* **2020**, *7*, 001656. [CrossRef] [PubMed]
3. Istepanian, R.S.H.; Al-Anzi, T. m-Health 2.0: New perspectives on mobile health, machine learning and big data analytics. *Methods* **2018**, *151*, 34–40. [CrossRef] [PubMed]
4. da Silveira, A.C.M.; Sobrinho, Á.; Dias da Silva, L.; de Barros Costa, E.; Pinheiro, M.E.; Perkusich, A. Exploring Early Prediction of Chronic Kidney Disease Using Machine Learning Algorithms for Small and Imbalanced Datasets. *Appl. Sci.* **2022**, *12*, 3673. [CrossRef]
5. Sobrinho, A.; Queiroz, A.C.M.D.S.; Silva, L.D.D.; Costa, E.D.B.; Pinheiro, M.E.; Perkusich, A. Computer-Aided Diagnosis of Chronic Kidney Disease in Developing Countries: A Comparative Analysis of Machine Learning Techniques. *IEEE Access* **2020**, *8*, 25407–25419. [CrossRef]
6. Kar, K.; Kornblith, S.; Fedorenko, E. Interpretability of artificial neural network models in artificial intelligence versus neuroscience. *Nat. Mach. Intell.* **2022**, *4*, 1065–1067. [CrossRef]
7. Belard, A.; Buchman, T.; Forsberg, J.; Potter, B.K.; Dente, C.J.; Kirk, A.; Elster, E. Precision diagnosis: A view of the clinical decision support systems (CDSS) landscape through the lens of critical care. *J. Clin. Monit. Comput.* **2017**, *31*, 261–271. [CrossRef]
8. Viana Dos Santos Santana, Í.; Cm da Silveira, A.; Sobrinho, Á.; Chaves E Silva, L.; Dias da Silva, L.; Santos, D.F.S.; Gurjão, E.C.; Perkusich, A. Classification Models for COVID-19 Test Prioritization in Brazil: Machine Learning Approach. *J. Med. Internet Res.* **2021**, *8*, e27293. [CrossRef]
9. Son, W.S.; Park, J.E.; Kwon, O. Early detection of influenza outbreak using time derivative of incidence. *EPJ Data Sci.* **2020**, *9*, 28. [CrossRef]
10. Kumar, S. Monitoring Novel Corona Virus (COVID-19) Infections in India by Cluster Analysis. *Ann. Data Sci.* **2020**, *7*, 417–425. [CrossRef]
11. Aftab, M.; Amin, R.; Koundal, D.; Aldabbas, H.; Alouffi, B.; Iqbal, Z. Classification of COVID-19 and Influenza Patients Using Deep Learning. *Contrast Media Mol. Imaging* **2022**, *2022*, 8549707. [CrossRef] [PubMed]

12. Li, W.; Ma, J.; Shende, N.; Castaneda, G.; Chakladar, J.; Tsai, J.C.; Apostol, L.; Honda, C.O.; Xu, J.; Wong, L.M.; et al. Using machine learning of clinical data to diagnose COVID-19: A systematic review and meta-analysis. *BMC Med. Inform. Decis. Mak.* **2020** *20*, 247. [CrossRef]
13. Zhou, X.; Wang, Z.; Li, S.; Liu, T.; Wang, X.; Xia, J.; Zhao, Y. Machine Learning-Based Decision Model to Distinguish Between COVID-19 and Influenza: A Retrospective, Two-Centered, Diagnostic Study. *Risk Manag. Healthc. Policy* **2021**, *14*, 595–604 [CrossRef]
14. Elbasi, E.; Zreikat, A.; Mathew, S.; Topcu, A.E. Classification of influenza H1N1 and COVID-19 patient data using machine learning. In Proceedings of the 44th International Conference on Telecommunications and Signal Processing (TSP), Brno, Czech Republic, 26–28 July 2021; pp. 278–282.
15. Phu, D.N.; Vinh, P.C.; Quoc, N.K. Enhanced Diagnosis of Influenza and COVID-19 Using Machine Learning. *EAI Endorsed Trans. Context Aware Syst. App. [Internet]* **2023**, *9*, 1–6 [CrossRef]
16. Shilaskar, S.; Ghatol, A. Diagnosis system for imbalanced multi-minority medical dataset. *Soft Comput.* **2018**, *23*, 4789–4799. [CrossRef]
17. Chatterjee, A.; Gerdes, M.W.; Martinez, S.G. Identification of Risk Factors Associated with Obesity and Overweight—A Machine Learning Overview. *Sensors* **2020**, *20*, 2734. [CrossRef]
18. Almansour, N.A.; Syed, H.F.; Khayat, N.R.; Altheeb, R.K.; Juri, R.E.; Alhiyafi, J.; Alrashed, S.; Olatunji, S.O. Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study. *Comput. Biol. Med.* **2019**, *109*, 101–111. [CrossRef]
19. Biau, G.; Cadre, B.; Rouvière, L. Accelerated gradient boosting. *Mach. Learn.* **2019**, *108*, 971–992. [CrossRef]
20. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *5*, 1189–1232. [CrossRef]
21. Xing, W.; Bei, Y. Medical Health Big Data Classification Based on KNN Classification Algorithm. *IEEE Access* **2020**, *8*, 28808–28819. [CrossRef]
22. Valdes, G.; Luna, J.; Eaton, E.; Simone, C.B., II; Ungar, L.H.; Solberg, T.D. MediBoost: A Patient Stratification Tool for Interpretable Decision Making in the Era of Precision Medicine. *Sci. Rep.* **2016**, *6*, 37854. [CrossRef] [PubMed]
23. Gao, X.; Alam, S.; Shi, P.; Dexter, F.; Kong, N. Interpretable machine learning models for hospital readmission prediction: A two-step extracted regression tree approach. *BMC Med. Inform. Decis. Mak.* **2023**, *23*, 104. [CrossRef] [PubMed]
24. Joyce, D.W.; Kormilitzin, A.; Smith, K.A.; Cipriani, A. Explainable artificial intelligence for mental health through transparency and interpretability for understandability. *NPJ Digit. Med.* **2023**, *6*, 6. [CrossRef]
25. Ahamad, M.M.; Aktar, S.; Rashed-Al-Mahfuz, M.; Uddin, S.; Liò, P.; Xu, H.; Summers, M.A.; Quinn, J.M.W.; Moni, M.A. A machine learning model to identify early stage symptoms of SARS-Cov-2 infected patients. *Expert Syst. Appl.* **2020**, *160*, 113661. [CrossRef] [PubMed]
26. Sarica, A.; Cerasa, A.; Quattrone, A. Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. *Front. Aging Neurosci.* **2017**, *9*, 329. [CrossRef] [PubMed]
27. Schober, P.; Vetter, T.R. Logistic Regression in Medical Research. *Anesth. Analg.* **2021**, *132*, 365–366. [CrossRef]
28. Demsar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
29. Choi, S.; Park, J.; Park, S.; Byon, I.; Choi, H.Y. Establishment of a prediction tool for ocular trauma patients with machine learning algorithm. *Int. J. Ophthalmol.* **2021**, *14*, 1941–1949. [CrossRef]
30. The Lancet Respiratory Medicine. Opening the black box of machine learning. *Lancet Respir. Med.* **2018**, *6*, 801. [CrossRef]
31. Boyton, R.J.; Altmann, D.M. The immunology of asymptomatic SARS-CoV-2 infection: What are the key questions? *Nat. Rev. Immunol.* **2021**, *21*, 762–768. [CrossRef]