

## Article

# A Hybrid Intelligent Approach to Predict Discharge Diagnosis in Pediatric Surgical Patients

Himer Avila-George <sup>1</sup>, Miguel De-la-Torre <sup>1</sup>, Wilson Castro <sup>2</sup>, Danny Dominguez <sup>3</sup>,  
Josué E. Turpo-Chaparro <sup>4</sup> and Jorge Sánchez-Garcés <sup>5,\*</sup>

<sup>1</sup> Departamento de Ciencias Computacionales e Ingenierías, Universidad de Guadalajara, Jalisco 46600, Mexico; himer.avila@academicos.udg.mx (H.A.-G.); miguel.dgomora@academicos.udg.mx (M.D.-I.-T.)

<sup>2</sup> Facultad de Ingeniería de Industrias Alimentarias, Universidad Nacional de Frontera, Sullana 20103, Peru; wcastro@unf.edu.pe

<sup>3</sup> Departamento Cirugía Pediátrica, Hospital Nacional San Bartolomé, Lima 15001, Peru; dandominguez@hotmail.com

<sup>4</sup> Escuela de Posgrado, Universidad Peruana Unión, Lima 15, Peru; josuetc@upeu.edu.pe

<sup>5</sup> Facultad de Ingeniería y Arquitectura, Universidad Peruana Unión, Lima 15, Peru

\* Correspondence: jasg@upeu.edu.pe



**Citation:** Avila-George, H.; De-la-Torre, M.; Castro, W.; Dominguez, D.; Turpo-Chaparro, J.E.; Sánchez-Garcés, J. A Hybrid Intelligent Approach to Predict Discharge Diagnosis in Pediatric Surgical Patients. *Appl. Sci.* **2021**, *11*, 3529. <https://doi.org/10.3390/app11083529>

Academic Editor: Donato Cascio

Received: 9 March 2021

Accepted: 9 April 2021

Published: 15 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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/>).

**Abstract:** Computer-aided diagnosis is a research area of increasing interest in third-level pediatric hospital care. The effectiveness of surgical treatments improves with accurate and timely information, and machine learning techniques have been employed to assist practitioners in making decisions. In this context, the prediction of the discharge diagnosis of new incoming patients could make a difference for successful treatments and optimal resource use. In this paper, a computer-aided diagnosis system is proposed to provide statistical information on the discharge diagnosis of a new incoming patient, based on the historical records from previously treated patients. The proposed system was trained and tested using a dataset of 1196 records; the dataset was coded according to the International Classification of Diseases, version 10 (ICD10). Among the processing steps, relevant features for classification were selected using the sequential forward selection wrapper, and outliers were removed using the density-based spatial clustering of applications with noise. Ensembles of decision trees were trained with different strategies, and the highest classification accuracy was obtained with the extreme Gradient boosting algorithm. A 10-fold cross-validation strategy was employed for system evaluation, and performance comparison was performed in terms of accuracy and F-measure. Experimental results showed an average accuracy of 84.62%, and the resulting decision tree learned from the experience in samples allowed it to visualize suitable treatments related to the historical record of patients. According to computer simulations, the proposed classification approach using XGBoost provided higher classification performance than other ensemble approaches; the resulting decision tree can be employed to inform possible paths and risks according to previous experience learned by the system. Finally, the adaptive system may learn from new cases to increase decisions' accuracy through incremental learning.

**Keywords:** computer-aided diagnosis; pediatrics; support vector machines; decision trees; CART; DBSCAN; XGBoost; AdaBoost; gradient boosting; voting ensemble; random bagging

## 1. Introduction

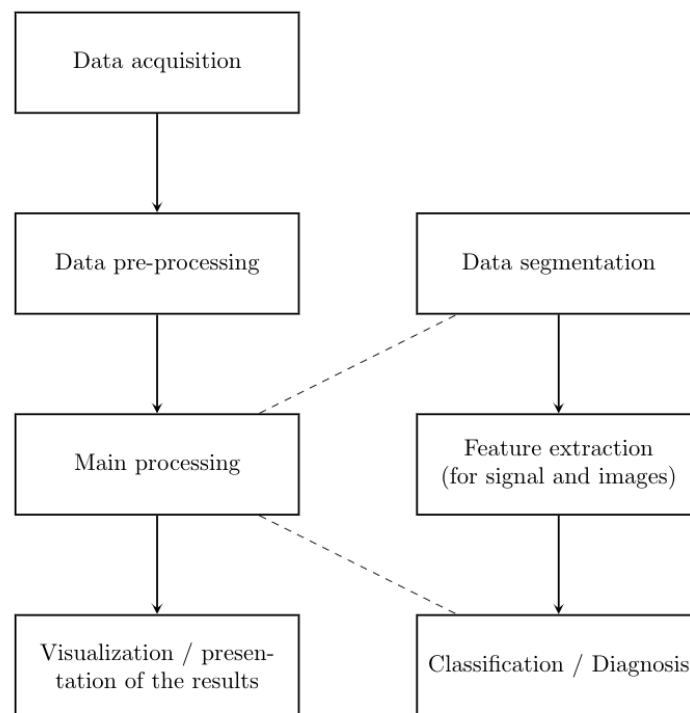
Computer-aided diagnosis systems have been proposed to solve medicine and biology problems since the late 1950s [1]. In a variety of health institutions, current clinical practices include the use of computer-based tools daily. However, such systems still present challenges either in the clinical, regulatory, and algorithmic aspects [2]. Regarding the algorithmic aspects, recent trends exhibit that studies are focusing on the use of artificial intelligence and machine learning techniques to diagnose diseases based on patients' historical records [1].

Medical treatments are complex management problems in pediatric patients due to limited historical information and the following two important factors. First, supplying antibiotics or drugs without an adequate diagnosis can lengthen treatment times and aggravate the condition of the patient. Secondly, surgical procedures in this type of patient are hazardous due to the small size of their organs, generating problems such as anastomosis [3]. From this perspective, as noted by Frongia et al. [4], a correct and timely diagnostic may help practitioners to reduce the risk involved in these procedures.

According to Ozgediz et al. [5], more than fifty percent of pediatric cases attended in hospitals correspond to readmission conditions, and many of these occurred in emergencies with surgical procedures in a critical time. Whereas some conditions may require a single intervention, in some cases, follow-up treatments are necessary due to unresolved pathologies in previous treatments. These situations may result in patients with very delicate conditions or death, which may provoke further sanctions to doctors and hospitals [5,6]. Additional complications may be caused by the lack of specialized medical equipment, supplies, and trained professionals for specialized surgical procedures, either in adults or children [5,7]. For a more accurate and prompt diagnosis, the clinical histories of patients are used to find treatment and symptom patterns over the most complex diseases. In this regard, Escobar and Caty [8] mentioned that 60 years of studies in neonatal physiology led to the creation of specialized pediatric anesthesia, critical neonatal care, and the identification of a significant part of the pathologies of the child. In that direction, the use of the *international statistical classification of diseases and related health problems* (ICD10) drives the diagnosis of complex cases, including infants; however, the diagnosis requires a subjective analysis that is highly influenced by the expertise of the physician.

Nowadays, computer scientists and medical researchers have investigated the potential of intelligent systems for computer-aided diagnosis and the planning of surgical procedures. Data from the clinical histories enable the modeling of the relationship between patient conditions and discharge diagnosis when patients are submitted to surgical procedures, using machine learning techniques [9]. An increasing tendency to use data science in medicine has been noted to provide additional information in complex cases or improve the diagnosis by reducing subjectivity [10]. Thus, Wong [11] suggested that in 2050, medicine will evolve into *precision medicine*, involving customized therapies that fit the biology of the patient and their metabolism, as well as genetically estimated drug dosing. From that perspective, pediatric procedures are not excluded, and computational models are fundamental for disease prediction and intervention, reducing time and effort for specific treatments [6,12].

The adoption of computer-aided diagnosis systems to improve the diagnosis of breast cancer [13] and lesions in organs [14] are examples of the numerous applications that are becoming more usual. Figure 1 shows the four steps, followed by a generic structure of a computer-aided diagnosis system, as suggested by Triantaphyllou [15] and Yanase and Triantaphyllou [2]. The details of each step are determined by the input data structure and the nature of the results required by the application scenario. Among the considerable amount of supervised learning approaches, one of the most well-known classifiers in computer-aided diagnosis systems is decision trees (DTs). Other classifiers commonly employed in this area include support vector machines, artificial neural networks, and k-nearest neighbors. Furthermore, it is well known that ensemble-based classifications tend to reduce bias and variance, producing more stable classification results [16–18]. Examples of general approaches include bagging and boosting, presenting a trade-off between accuracy and sensitivity to outliers. Variants of these general approaches are random forests and AdaBoost, which have been found to present comparable performance, and following the principle of no-free-lunch, the selection should follow a careful experimental trial.



**Figure 1.** The general structure of a computer-aided diagnosis system.

Table 1 presents a couple of publications that apply machine learning methods to solve classification issues. Although the results in Table 1 cannot be fairly compared due to differences in experimental settings and different classification problems, some good results are shown by distinct approaches. Among the classifiers employed in CAD systems, neural network approaches have been more often employed in medical imagery, and other approaches are spread over different applications. A relevant result obtained with AdaBoost reveals a significant reduction in the false-positive classification rate [19]. A decrease in the false detection of breast cancer is relevant to reduce the unnecessary costs derived from supplementary exams. Another interesting result is the use of XGBoost to predict the revisit, based on the historical records from patients [6]. A recent survey on the computer-aided diagnosis confirms the *no-free-lunch theorem* in CAD systems (no single algorithm can be applied to all aspects of CAD), and DTs are representative supervised approaches employed for classification [1]. Deep learning approaches are also employed for classification in CAD systems. With the drawback, these approaches require many training data to build accurate models, which is problematic for pediatric patients with a short clinical history. Despite the above, DTs' low stability to small changes in the training set makes this approach hard to tune, and the strategy employed to reduce such instability consists of using ensemble learning [16].

**Table 1.** Representative publications that report the use of machine learning approaches for medical diagnosis.

Problem	Proposal	Technique	Results	Validation	Reference
Classification of unhealthy and healthy neonates	The authors introduced a system for classifying unhealthy and healthy neonates in neonatal intensive care units using medical thermography processing and artificial neural networks (ANN).	ANN	Accuracy: 98.42%	10-fold cross validation	Savasci et al. [20]
Disease diagnosis (tuberculosis, malaria, and intestinal parasites)	The authors introduced a system based on convolutional neural networks (CNN) for disease diagnosis from microscopic images.	CNN	Accuracy: 96.05% AUC *: 0.99	Data splitting	Simon et al. [21]
Detection of congenital abnormalities of the kidney and urinary tract in children	The authors introduced a computer-aided diagnosis (CAD) system based on ultrasound imaging data, which consists of 3 components, namely: (1) kidney segmentation, (2) feature extraction, and (3) A classification model based on the support vector machines (SVM) technique.	CNN and SVM	AUCs > 0.88	10-fold cross-validation	Zheng et al. [22]
Mortality prediction	In this study, the authors developed a model based on the naïve Bayes (NB) technique to predict in-hospital mortality in patients undergoing transcatheter mitral valve repair.	NB	AUC: 0.83	Data splitting	Hernandez-Suarez et al. [23]
Detection of hypertension	The authors presented a CAD system to help clinicians in the hypertension prediction.	CNN and bagging	Accuracy: 90.9% AUC: 0.9091	10-fold cross-validation	Kandil et al. [24]
Arrhythmia classification	A system for monitoring cardiac patients using machine learning techniques like probabilistic neural network (PNN) and multilayer perceptron neural network (MLPNN).	PNN	F-measure: 91.83	22-fold cross-validation	Rajagopal [25]
Voice pathology detection	The authors presented a method combining density clustering and support vector machines for voice pathology detection.	SVM and DBScan	Accuracy: 98.00%	Cross validation	Amami and Smitib [26]
Disease diagnosis (heart disease, Parkinson's disease, and BUPA liver disorder)	A hybrid system for diseases diagnostic is proposed, which is compounded by a new method entitled k-medoids clustering-based attribute weighting (kmAW) as a data pre-processing method, and an SVM was preferred in the classification phase.	SVM and kmAW	Accuracy: 98.95%	10-fold cross-validation	Peker [9]
Predicting patient revisits	This study focuses on the predictive identification of patients frequently revisiting the University of Virginia Health System Emergency Department. The authors proposed the use of the XGBoost algorithm to predict the risk of revisit.	XGBoost	AUC: 0.754	Data splitting	Fowler et al. [6]
Identification of brain tumors	The detection of tumors is performed with the help of automatic computing technique.	CNN	Accuracy: 99.1%	Data splitting	Samikannu et al. [27]

Table 1. Cont.

Problem	Proposal	Technique	Results	Validation	Reference
Detection of breast cancer	The proposed model is based on the SVM technique, which uses a radial basis function (RBF) kernel.	SVM	Accuracy: 93.9%	10-fold cross-validation	Rahman et al. [28]
	The proposed CAD system uses ensemble learning from CNN.	CNN	Accuracy: 95.77% AUC: 0.9462	Data splitting	Moon et al. [29]
	The CAD system is based on feature selection and ensemble learning. Compared with other methods [30], the proposed method significantly reduces the false-positive classification rate.	AdaBoost	AUC: 0.9617	10-fold cross-validation	Lu et al. [19]

\* Receiver operating characteristics (ROC) is a probability curve and the area under the ROC curve (AUC) represents degree or measure of separability.

Regarding the limitations of current systems, the works reported in Table 1 provide information about the diagnosis of a given condition; some of them use images to describe the characteristics of some pathologies, and others describe demographic data. However, the works reported do not consider characteristics such as the complications that occurred during treatment and relevant factors related to the patient's comorbidities, which could have led to poor patient outcomes.

This research paper is a retrospective study in which a computer-aided diagnostic system is proposed to predict the discharge diagnosis of pediatric patients undergoing surgical procedures in a third-level pediatric hospital in Peru. According to the hospital requirements, the system predicts one of three discharge diagnoses: (1) deceased; (2) unhealthy; and (3) healthy. These three discharge diagnoses are those employed in the current input analysis made to every new patient, based on the historical record and new observations. Five modules are proposed to evaluate patients' records. In the second module, missing data are completed using the wrapper algorithm. In the third module, non-relevant records (outliers) are filtered out to reduce noisy samples. In the fourth module, the ensembles of decision trees are trained using XGBoost to classify and present the results. Finally, the system is evaluated on a dataset composed of historical records from pediatric patients, following a 10-fold cross-validation process. The proposed approach is compared to distinct classifier ensembles in terms of *Accuracy* and *F-measure*.

## 2. Materials and Methods

The context of the implementation is the Hospital Nacional San Bartolomé, in Lima, Perú, which is a category III-1 hospital, that is part of the public health service [31]. Due to the retrospective nature of the study, informed consent was waived. The sections below describe the data from clinical records and the experimental methodology employed to evaluate the performance and properties of the proposed computer-aided diagnosis system.

### 2.1. Data Acquisition

A database with 1205 medical records of pediatric patients is available, each patient was diagnosed at the hospital entrance, and their condition was coded according to the International Classification of Diseases, version 10 (ICD10) standard [32]. Additional information in patients' records includes a medical history and a detailed diagnostic before and after hospitalization.

Table 2 presents the complete list of characteristics considered for classification. Whereas ordinal features are suitable for thresholding, categorical data are restricted to a number of categories that are represented with names or tags.

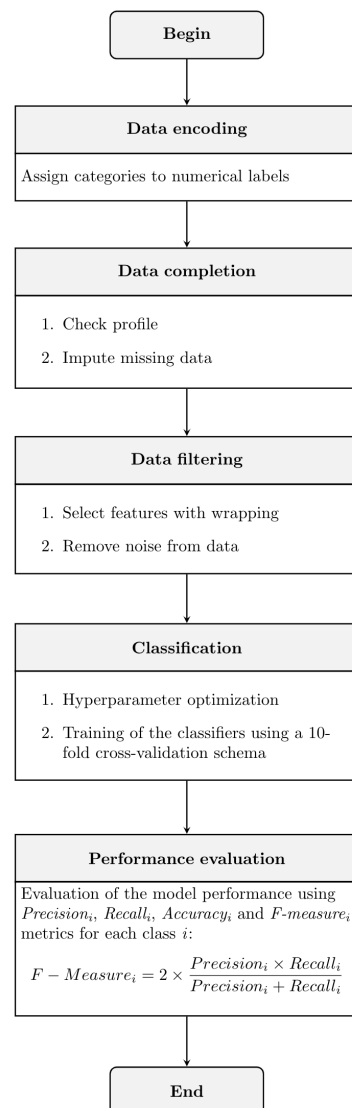
**Table 2.** Eleven data features employed to describe the present state of patient records.

Features	Description	Ordinal
Age	Patient's age	Yes
Weight	Patient's weight	Yes
Gender	Patient's gender	No
DI	Demographic information *	No
Time_qx	Surgery time	Yes
D_Hospital	Days in hospital	Yes
ICD10	Entry diagnosis	No
Medicine	Medicine supplied	No
TT_Medic	Level of medical treatment	No
Found	Treatment findings	No
Complications	Treatment complications	No

\* Residence.

## 2.2. Experimental Methodology

The process pursued to evaluate the system was summarized in the five phases shown in Figure 2; this process was based on the general structure of a computer-aided diagnostic system (see Figure 1). The first phase is the coding of the data by assigning categories to the numerical labels. Then, the missing information in the medical records must be completed using imputer methods. The third phase was to filter the data by removing noise from the data and performing a feature selection process to classify the data. The fourth phase was the classification of the data, which included optimizing the hyperparameters of the selected classification methods and their subsequent training using the cross-validation technique. Finally, an evaluation of the results was made using some performance metrics. The five phases of the methodology are detailed in the following sections.



**Figure 2.** Experimental methodology employed to evaluate machine learning algorithms for the proposed system.

### 2.2.1. Pre-Processing Data

The preprocessing phase is designed to prepare the medical records to be suitable for the central processing of the proposed system by encoding data and completing missing information.

### Data Encoding

The data coding process consists of two steps: (1) reading the input attributes and categorizing their values; (2) creating a dictionary for each attribute by assigning nominal data to the incoming categorical data. The first step consists of receiving input data in text format from which the attribute values are extracted by eliminating repetitions. In the next step, the attribute values are converted to numerical data so that the classification algorithms can process them. For example, we have an input attribute called “Provenance” with values (Piura, Tumbes, Tumbes, Piura, Comas); the result of applying the first step of the encoding process is (Piura, Tumbes, Comas); finally, applying the second step of the encoding results (Piura = 1, Tumbes = 2, Comas = 3).

### Data Completion

As a consequence of manual data retrieval, several records were found to be incomplete. Data from medical records present 8.6% of missing values among the eight numerical features and four categorical traits (see Table 2). In the pattern recognition literature, the incomplete data problem is commonly addressed by deleting incomplete cases, using models to estimate data distribution or classifier parameters, or evaluating missing data through imputation methods.

In order to complete missing information from medical records, the following two methods were selected:

1. The K-nearest neighbors imputer employs the K-nearest neighbor (KNN) classifier with Euclidean distance, as shown in Equation (1), to describe the similarity between the incomplete record ( $x_i$ ) and other records nearby ( $x_j$ ) [33,34]:

$$d_{i,j} = \text{dist}(x_i, x_j) = \sqrt{\sum_{k=1}^n (x_{i,k} - x_{j,k})^2}, \quad (1)$$

where  $\text{dist}(x_i, x_j)$  is the Euclidean distance between the encoded medical records; and  $n$  symbolizes the number of features. The estimate of the missing value at record  $i$  compared to the record  $j$  is given by

$$\hat{x}_{i,j} = \frac{\sum_{k=1}^K W_k X_k}{\sum_{k=1}^K W_k}, \quad (2)$$

where  $k$  is the number of samples selected from features;  $X_k$  is the input matrix for the  $k^{\text{th}}$  record; and  $W_k$  is the  $k^{\text{th}}$  similarity weight defined by

$$W_k = \frac{1}{d_k}, \quad (3)$$

where  $d_k \in \{d_1, d_2, \dots, d_K\}$  is the  $k^{\text{th}}$  rank distance of the neighbors.

2. The simple imputer method employed is based on mean, as described by Buuren and Groothuis-Oudshoorn [35]. The simple imputer replaces the missing values with the mean value of the missing feature, considering all records in training data, according to:

$$\hat{x}_i = \frac{1}{N} \sum_{l=1}^N x_l, \quad (4)$$

where values in matrix  $X$  are the observations of each feature in the medical record; and  $N$  is the amount of records used for training.

#### 2.2.2. Main Processing

The central processing phase was designed to produce discharge diagnosis predictions from numerical samples prepared in the preprocessing stage.



### Data Filtering

Data filtering is a two-step process that includes selecting relevant features for classification and removing noisy data. Feature selection techniques can be divided into three categories: filter, wrapper, and embedded methods. From these three categories, wrapper methods have the advantages of using feature dependencies and are developed with classifier performance [36]. Additionally, as a difference from embedded methods, the classifier can be replaced by any other available once the most relevant features are found.

In the proposed system, the sequential forward selection (SFS) wrapper method was employed to select the most relevant features evaluated [37,38]. In order to remove noisy records that are likely to affect classification performance negatively, medical records were filtered out using the Density-based spatial clustering of applications with noise (DBSCAN) clustering. The DBSCAN algorithm aimed to find the essential samples with higher density in order to expand clusters from them, finding clusters of similar density [26].

### Classification with XGBoost

XGBoost was employed to train the ensemble of decision trees for discharge diagnosis prediction. Ensembles of classifiers take advantage of the diversity of opinions between weak classifiers to produce accurate and commonly with more stable classification performance. A decision tree (DT) is a predictive model commonly used as a weak classifier in ensembles, in which the conjunctions of features are represented in the branches, and conclusions or decisions are represented in the leaves (class labels). XGBoost is commonly employed for training decision trees, where an objective function is defined for the supervised learning of a model [39]. Training consists of searching for the best parameters  $\theta$  that fit the training data  $x_i$  and the labels  $y_i$ , see Equation (5):

$$Obj(\theta) = L(\theta) + \Omega(\theta), \quad (5)$$

where  $L$  = is the training loss; and  $\Omega$  is the regularization term.  $L$  measures how predictive the model is regarding the training data, and its equation is given by Equation (6):

$$L(\theta) = \sum_i (y_i - \hat{y}_i)^2. \quad (6)$$

On the other hand, the regularization term controls the complexity of the model and is given by

$$\gamma * T + \frac{1}{2} \lambda * \sum_{j=1}^T \omega_j^2, \quad (7)$$

where  $T$  is the number of end nodes of the tree, and  $\omega$  is the vector of scores on end nodes.

The solution is provided by trees that are constructed sequentially and learn from each tree's predecessors. The learning scheme is called additive training: the functions  $f_i$  contain the structure of the tree and the end nodes' scores. The learning structure of the tree is complex because it cannot be learned from all the trees at once, and a prediction value is obtained at each step  $t$  as  $\hat{y}^t$  according to Equation (8):

$$\begin{aligned} \hat{y}^0 &= 0, \\ \hat{y}^1 &= f_1 = \hat{y}^0 + f_1(X_i), \\ \hat{y}^t &= \sum_{k=1}^t f_k(x_i) = \hat{y}^{t-1} + f_1(X_i). \end{aligned} \quad (8)$$

Optimizing performance in decision trees includes finding the maximum depth to prune the trees backward, eliminating losses, and optimizing learning. Other parameters to be considered are the number of trees, the learning rate to prevent overfitting, the percentage of samples used per tree, and the percentage of features used per tree [6,40]. The *GridSearch* algorithm was employed to automatically optimize each classifier's hyperparameters on a validation set [41]. The *GridSearch* algorithm starts by defining a limited number of values

for the hyperparameters, then the Cartesian product of these sets is evaluated through a sequential combination. Additionally, a 10-fold cross-validation process was applied for the statistical comparison of performance [42].

### 2.3. Performance Evaluation

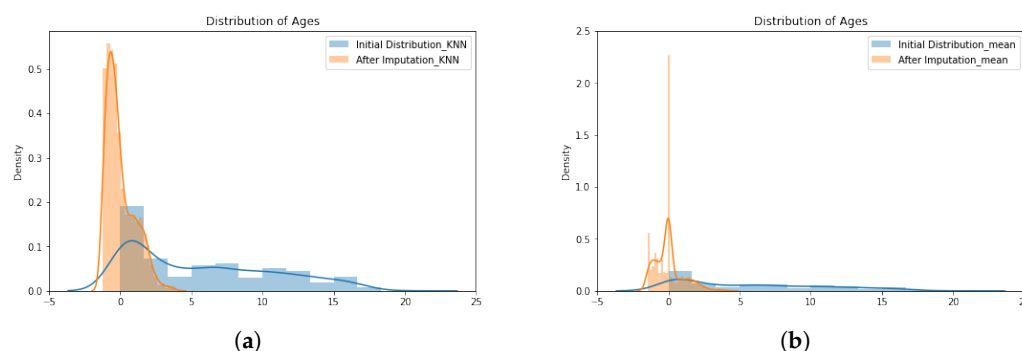
A 10-fold cross-validation strategy was followed to obtain an average performance and standard deviation. Performance measures derived from confusion matrices were employed to represent real and predicted classes. The main metrics used were: *Accuracy*, *Recall*, *Precision*, and *F-measure*; for a detailed description of the metrics used, please refer to the work of Castro et al. [43].

## 3. Experimental Results

The proposed system was implemented using Python v3.6.8, and the libraries *joblib*, *numpy*, *pandas*, *date*, *pytz*, *scikit-learn*, *scipy*, and *xlrd*. The server employed to execute the comparison of the system using distinct algorithms included a virtual machine running on Linux Centos 7.0 with 32 processor chips (Intel Xeon (R) CPU 2.60 GHz), and 20 MB of RAM.

### 3.1. Pre-Processing Results

The result of the encoding phase was the assignment of an integer to each categorical feature value. After the imputation of the lost data with the KNN imputer and simple imputer, the dataset was normalized to validate the most accurate methods. The distributions of age after data completion with both methods are shown in Figure 3.



**Figure 3.** (a) Distribution of the age data using the K-nearest neighbor (KNN) imputer; and (b) the distribution of age data using simple imputer.

By comparing both distributions, it can be seen that the KNN imputer provides a distribution that complies with the principle of the central theorem, where a higher area is close to zero. In other words, errors are normally distributed, and meaningful samples can be parameterized. Similar results were observed with the rest of the features, and the rest of the preprocessing was conducted on KNN imputed data.

### 3.2. Main Processing

With sample records meeting the new distribution of data provided by the KNN imputer, a subset of six features was selected through the SFS wrapper. The features included DI, ICD10, Medicine, TT\_Medic, Complications, and D\_Hospital, producing a *precision* up to 82.4%. For data filtering, the best hyperparameters of the DBSCAN were explored on validation data using a search strategy based on random candidate combinations (provided by the *ParameterSampler* function). The hyperparameters found include  $eps = 4.5285$ ,  $minsamples = 9$ ,  $p = 1$ , and a cohesion of cluster of 87%; as a result, nine records were deleted from the original dataset. For a fair comparison, the *GridSearch* algorithm was applied to all classifiers with partitions of data following 70% for training

and 30% of samples for test. The resulting hyperparameters for each classifier are shown in Table 3.

**Table 3.** List of hyperparameters for each algorithm.

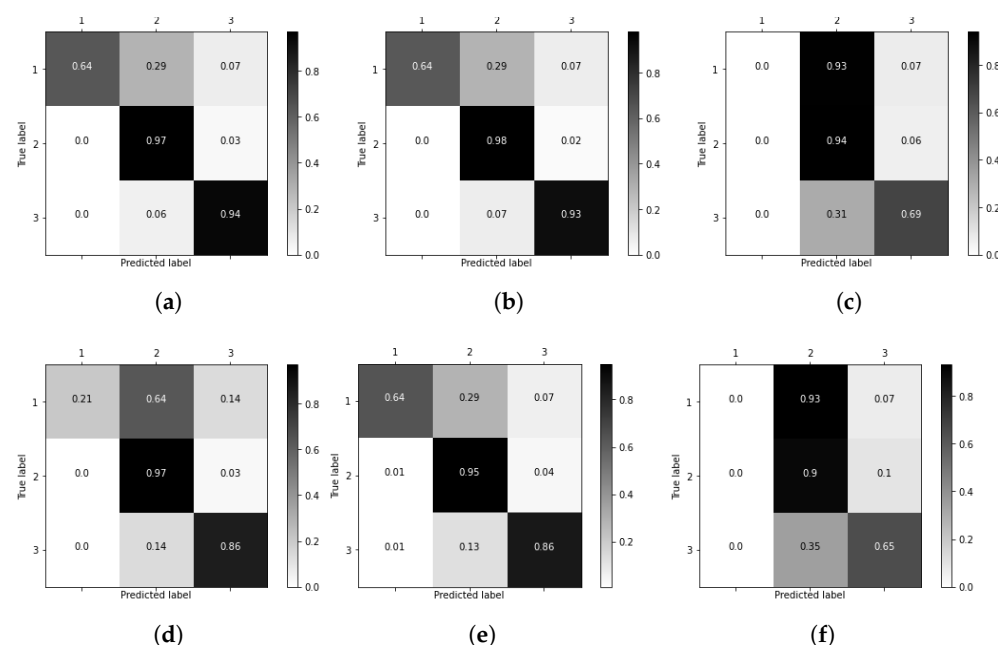
Technique	Hyperparameter	Values
XGBoost	max_depth	{4,5,6,7}
	learning_rate	{0.001, 0.01, 0.1, 0.2, 0.3}
	subsample	{0.5, 0.6, 0.7, 0.8, 0.9, 1.0}
	colsample_bytree	{0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0}
	colsample_bylevel	{0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0}
	min_child_weight	{0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0}
	gamma	{0, 0.25, 0.5, 1.0}
	reg_lambda	{0.1, 1.0, 5.0, 10.0, 50.0, 100.0}
AdaBoost	base_estimator_criterion	{gini, entropy}
	base_estimator_splitter	{best, random}
	algorithm	{SAMME.R, SAMME}
	n_estimators	{10, 100, 200, 250}
	learning_rate	{.05, 0.5, 1.5, 2.5}
Gradient boosting	loss	{deviance}
	learning_rate	{0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2}
	min_samples_split	{0.1,0.5,12}
	min_samples_leaf	{0.1,0.5,12}
	max_depth	{3,5,8}
	max_features	{log2,sqrt}
	criterion	{friedman_mse,mae}
	subsample	{0.5, 0.618, 0.8, 0.85, 0.9, 0.95, 1.0}
	n_estimators	{10}
Random bagging	criterion	{gini, entropy}
	learning_rate	{0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2}
	min_samples_split	{3,4,5,6,7}
	min_samples_leaf	{1,2,3}
	random_state	{123}
	n_jobs	{-1}
	n_estimators	{10,15,20,25,30}
CART	max_features	{auto, sqrt, log2}
	min_samples_split	{2,3,4,5,6,7,8,9,10,11,12,13,14,15}
	min_samples_leaf	{1,2,3,4,5,6,7,8,9,10,11}
	random_state	{123}
Voting ensemble	lr_C	{1.0, 100.0}
	svm_C	{2,3,4}

The average performance of each classifier after a 10-fold cross-validation strategy for classifier evaluation and comparison is presented in Table 4, where the numbers in parenthesis represent the standard deviation for ten replications of the experiment. Bold numbers highlight the highest performance. Results in Table 4 reveal that the proposed XGBoost algorithm achieves the best performance in terms of both *accuracy* and *F-measure*. These results are consistent with those presented in the literature for different applications. For instance, Nguyen et al. [44] mentions the XGBoost model as a robust algorithm to build predictive models when applied to predict the environmental effects around a mine.

**Table 4.** The performance of the proposed system with XGBoost, and compared to other decision tree (DT) ensemble algorithms.

Technique	Accuracy	F-Measure
XGBoost	<b>84.62%</b>	<b>73.99</b>
AdaBoost	82.83%	46.95
Gradient boosting	78.44%	44.96
Random bagging	84.82%	49.77
CART	77.63%	46.79
Voting ensemble	74.05%	46.35

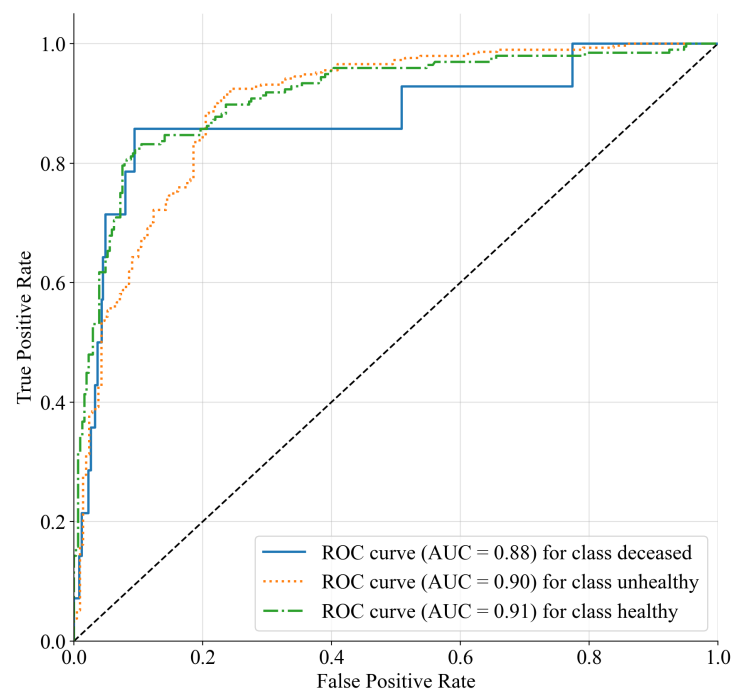
Further performance analysis can be retrieved by analyzing the confusion matrices for each classifier. Confusion matrices in Figure 4 summarize the predictability of each model: the dark gray colors represent high values, whereas light gray colors represent low values. Comparing Figure 4a–f, the highest level of correctly classified samples is provided by XGBoost: Figure 4b. Here, most of the errors occur when class 1 (deceased patients) are confused with class 2 (unhealthy patients).

**Figure 4.** Confusion matrices for classifier comparison and performance analysis. (a) AdaBoost, (b) XGBoost, (c) Gradient boosting, (d) Random bagging, (e) CART, and (f) Voting ensemble.

Although the overlap between classes 1 and 2 in Figure 4b is close to one-third of the decisions, it might be considered that pathologies and complications may appear after the patient is discharged. Similarly, the highest precision presented in all cases corresponds to class 2, and the particular situation for discharging unhealthy patients should require further analysis by practitioners. A higher level of errors is presented between classes 1 (deceased patient) and 2, and close attention must be paid to every particular case. Although approaches in Figure 4a,b,e present a dark diagonal (correct class predictions), the diagonal values for XGBoost are consistently higher than those of other approaches.

Finally, Figure 5 presents the receiver operating characteristics (ROC) curves per class computed for the proposed system designed with XGBoost. In order to construct the ROC curves, the output probabilities from each class were computed, and the area under the ROC curve (AUC) was estimated for each curve using the approximation by rectangles. The lowest AUC was achieved by the blue solid ROC curve, which represents the performance of the system for class 3 (deceased patients), and the operational point closest to the upper-left corner corresponds to  $tpr = 0.85$  and  $fpr = 0.1$ . Furthermore,

selecting the correct operational point is relevant to tune the system and reduce errors for specific classes.



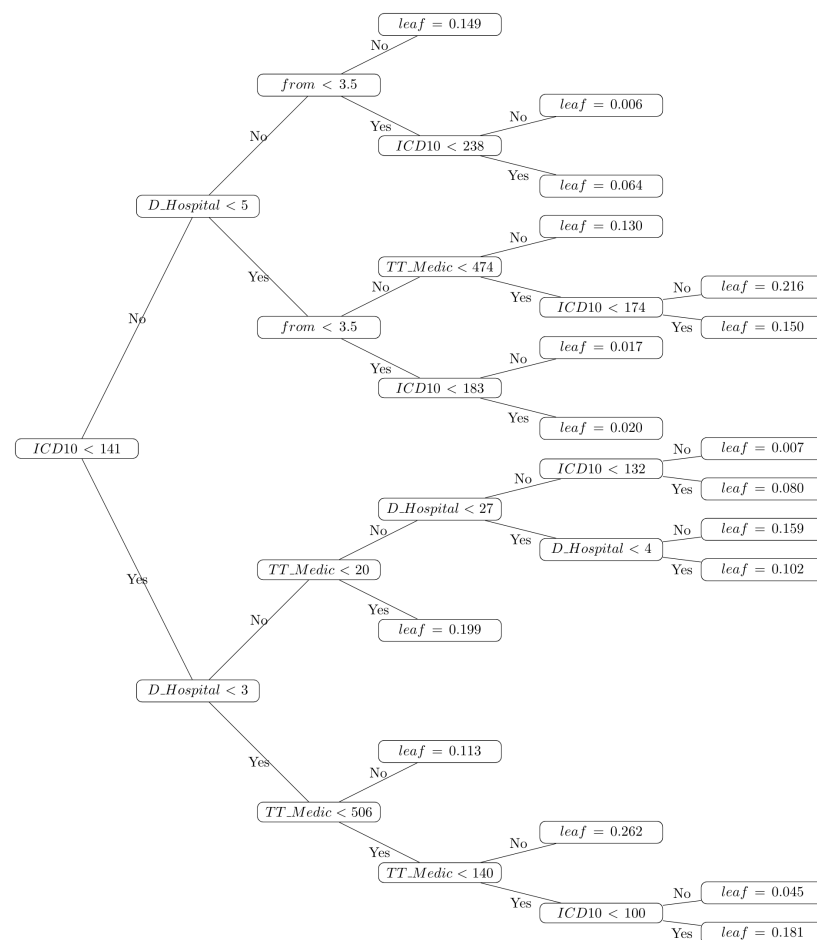
**Figure 5.** ROC curves that show the system performance for each class (e.g., discharge diagnosis: deceased, unhealthy, and healthy). The diagonal dashed line represents a random classifier that does not consider any information to make decisions.

### 3.3. Decision Trees for Computer-Aided Diagnosis

According to Chen and Guestrin [39], the resulting probabilities of decision trees can be computed using the logistic function. As shown in Figure 6, the resulting decision tree allowed identifying malformations of the digestive tract, firstly the atresia of the esophagus. This was identified in the node ( $ICD10 < 141$ ), where it makes the code Q39 (ICD10).

According to Stoll et al. [45], this treatment with reserved diagnosis was that the child cannot feed or pass the saliva. The latter causes the same to pass to the lungs, and the other problem that complicates the treatment is that if the distal segment of the atrial (malformed) esophagus attached to the trachea causes the gastric juice to also pass to the lung, the child will develop pneumonia. Currently, this pathology's complexity is often associated with premature heart problems and other associated malformations such as rectal anus malformations, which makes the reserved prognosis of the child and the urgency of the solution: the subsequent three days are crucial in the treatment.

The nodes ( $TT\_Medic < 506$  and  $TT\_Medic < 140$ ) refer to the correct treatment that has been successful in the historical data. The correct treatment is described: since the patient cannot be fed, it must be hydrated with a good intravenous route with prophylactic antibiotics, hydration, and feeding (total parenteral nutrition). The catheter designed for this does not go to a peripheral or superficial vein because it does not hold the NPT; it must go to a more prominent or central vein. If a peripheral vein is used, the solution's osmolarity will inflame and affect it.



**Figure 6.** Exemplar of the resulting DT that provides treatment and diagnosis information retrieved from training samples.

For this reason, we reference a tunneled CVC (TT\_Medic < 140) and a probe (TT\_Medic < 506) to aspirate the saliva until it is operated. After surgery, the procedure consists of placing a drain in the thorax to confirm if the surgery points (anastomosis) have a saliva leak. In this sense, it is corrected until everything is well. The probability of dying with these recommendations is 54.5% since these cases are extremely critical and even more so if not in newborns. The complexity of the treatment of these malformations lies in the fact that there are other added affectations. Likewise, pathologies associated with the digestive tract's malformations are of the genitourinary category belonging to the N category of ICD10, especially rectal atresia. For this reason, the tree places (ICD10 < 100) [46].

Other malformations that interrupt the communication between the mouth and the anus would be intestinal atresia and rectal atresia. Supposing these atresias are not treated on time, we consider the three days referred to at the tree. In this situation, the case will present an increase in the probability of complications such as sepsis (generalized infection), multi-organ failure, coagulation disorders, shortness of breath, and the acidotic state will be high. The probability of death increases from 54.5% to 59.4%, being that in previous lines, and the proposal of natural treatment of atresias was explained within three days. Therefore, time is vital, and as happens in third-level hospitals, patients arrive from other cities where they had previously had unsuccessful treatments and therefore are very delicate cases.

Another example is shown in the tree (TT\_Medic < 20). This reveals that a treatment such as an ostomy (removing the proximal part of the intestine to the skin in order to let it drain) involves drainage and complicated infection control. Therefore, it is vital to consider

the days of treatment if it is more significant than three triggers in a perforation of the intestine and the chances of death increase.

#### 4. Conclusions

In this paper, a computer-aided diagnosis system was proposed to be employed as a tool for accurate and timely diagnosis in a third-level hospital. Based on decision trees, the system considers the historical records of input pediatric patients to predict an estimated discharge diagnosis and possible treatments. The proposed CAD system predicts three categories of patient discharge: (1) deceased, (2) unhealthy, and (3) healthy. The CAD learns from the previous diagnoses, the treatment applied, and the patient's discharge condition in order to assist the pediatric surgeon's decision-making so that the best treatment can be offered. If the patient is discharged with pathology, this condition is due to the fact that the patient requires management or surgical treatment in stages due to the complexity of the pathology or the need for its correction at a later age. For example, there is the case of high anorectal malformations, in which at the neonatal stage, a procedure is performed that allows the child to have a bowel movement through an artificial orifice, and then, at a later age, the definitive correction is performed. Let us suppose that the patient dies later, before completing the entire treatment. In that case, the factors that led to the patient's poor evolution would have to be recorded, and the system would have to be fed back. Therefore, the system must have continuous feedback to improve medical decision support.

The main advantages that are foreseen with using a prediction system to support the making decision process include the following. First, a graphical representation of the possible paths from admission to output diagnosis can provide the means for better decisions. Second, the resulting decision tree can be employed to inform possible paths and risks to the patients' parents, according to previous system experience. Third, statistics on previous cases may provide experience-based evidence in the case of legal conflicts. Finally, an adaptive system may learn from new cases to make more accurate decisions as the knowledge is improved with experience.

Future work could further improve the results by using XGBoost with ensemble learning (Boosting and Bagging) methods, which consists of selecting the samples that obtained the least error during the learning of sequentially constructed trees. This distributed learning environment can solve problems beyond billions of examples, making it much more versatile when it comes to treatments as delicate as pediatric, and hyperparameters for ensemble learning can be optimized considering a trade-off between *accuracy* and stability [47]. Additionally, given the continuous entry of novel cases, the system might be adapted to incorporate new information on the new sample records. Ensemble learning methodologies have been proven to provide good performance after incremental learning and fusion adaptation. Finally, although the use of the ICD10 standard provides a helpful framework, within the next few years, the system should consider the recent ICD11 codification to include a more accurate diagnosis [48]. Backward compatibility may be resolved by adding a module to translate medical records from ICD10 to ICD11.

**Author Contributions:** Conceptualization, J.S.-G. and D.D.; methodology, J.S.-G., W.C., M.D.-I.-T., and H.A.-G.; software, J.S.-G. and M.D.-I.-T.; validation, W.C., H.A.-G., D.D., and J.E.T.-C.; formal analysis, M.D.-I.-T.; investigation, D.D.; resources, J.E.T.-C.; data curation, J.S.-G.; writing—original draft preparation, J.S.-G., M.D.-I.-T., and H.A.-G.; writing—review and editing, J.S.-G., M.D.-I.-T., W.C., J.E.T.-C., and H.A.-G.; visualization, J.S.-G. and H.A.-G.; supervision, W.C.; project administration, J.S.-G.; funding acquisition, J.E.T.-C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

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

**Informed Consent Statement:** Not applicable.



**Data Availability Statement:** Not applicable.

**Acknowledgments:** Not applicable.

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

## References

1. Yanase, J.; Triantaphyllou, E. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Syst. Appl.* **2019**, *138*, 1–25. [\[CrossRef\]](#)
2. Yanase, J.; Triantaphyllou, E. The seven key challenges for the future of computer-aided diagnosis in medicine. *Int. J. Med. Inform.* **2019**, *129*, 413–422. [\[CrossRef\]](#)
3. Coventry, B.J. *Pediatric Surgery*; Springer: London, UK, 2014.
4. Frongia, G.; Mehrabi, A.; Ziebell, L.; Schenk, J.P.; Gunther, P. Predicting Postoperative Complications After Pediatric Perforated Appendicitis. *J. Investig. Surg.* **2016**, *29*, 185–194. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Ozgediz, D.; Langer, M.; Kisa, P.; Poenaru, D. Pediatric surgery as an essential component of global child health. *Semin. Pediatr. Surg.* **2016**, *25*, 3–9. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Fowler, B.; Rajendiran, M.; Schroeder, T.; Bergh, N.; Flower, A.; Kang, H. Predicting patient revisits at the University of Virginia Health System Emergency Department. In Proceedings of the 2017 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA, 28 April 2017; pp. 253–258. [\[CrossRef\]](#)
7. Greenberg, S.; Ng Kamstra, J.; Ameh, E.; Ozgediz, D.; Poenaru, D.; Bickler, S. An investment in knowledge: Research in global pediatric surgery for the 21st century. *Semin. Pediatr. Surg.* **2016**, *25*, 51–60. [\[CrossRef\]](#)
8. Escobar, M.; Caty, M. Complications in neonatal surgery. *Semin. Pediatr. Surg.* **2016**, *25*, 347–370. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Peker, M. A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM. *J. Med. Syst.* **2016**, *40*, 116. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Bennett, T.D.; Callahan, T.J.; Feinstein, J.A.; Ghosh, D.; Lakhani, S.A.; Spaeder, M.C.; Szefer, S.J.; Kahn, M.G. Data science for child health. *J. Pediatr.* **2019**, *208*, 12–22. [\[CrossRef\]](#)
11. Wong, H.R. Intensive care medicine in 2050: Precision medicine. *Intensive Care Med.* **2017**, *43*, 1507–1509. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Suri, J.S.; Singh, S.; Reden, L. Computer Vision and Pattern Recognition Techniques for 2-D and 3-D MR Cerebral Cortical Segmentation (Part I): A State-of-the-Art Review. *Pattern Anal. Appl.* **2002**, *5*, 46–76. [\[CrossRef\]](#)
13. Khuwaja, G.A.; Abu-Rezq, A.N. Bi-modal breast cancer classification system. *Pattern Anal. Appl.* **2004**, *7*, 235–242. [\[CrossRef\]](#)
14. Moore, M.M.; Slonimsky, E.; Long, A.D.; Sze, R.W.; Iyer, R.S. Machine learning concepts, concerns and opportunities for a pediatric radiologist. *Pediatr. Radiol.* **2019**, *49*, 509–516. [\[CrossRef\]](#)
15. Triantaphyllou, E. *Data Mining and Knowledge Discovery via Logic-Based Methods: Theory, Algorithms, and Applications*; Springer Science+Business Media: New York, NY, USA, 2010.
16. Kuncheva, L. *Combining Pattern Classifiers*; Wiley: Hoboken, NJ, USA, 2014. [\[CrossRef\]](#)
17. Hameed, K.; Chai, D.; Rassau, A. A Progressive Weighted Average Weight Optimisation Ensemble Technique for Fruit and Vegetable Classification. In Proceedings of the 2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV), Shenzhen, China, 13–15 December 2020; pp. 303–308.
18. Hameed, K.; Chai, D.; Rassau, A. A Sample Weight and AdaBoost CNN-Based Coarse to Fine Classification of Fruit and Vegetables at a Supermarket Self-Checkout. *Appl. Sci.* **2020**, *10*, 8667. [\[CrossRef\]](#)
19. Lu, W.; Li, Z.; Chu, J. A novel computer-aided diagnosis system for breast MRI based on feature selection and ensemble learning. *Comput. Biol. Med.* **2017**, *83*, 157–165. [\[CrossRef\]](#)
20. Savasci, D.; Ornek, A.H.; Ervural, S.; Ceylan, M.; Konak, M.; Soyulu, H. Classification of unhealthy and healthy neonates in neonatal intensive care units using medical thermography processing and artificial neural network. In *Classification Techniques for Medical Image Analysis and Computer Aided Diagnosis*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 1–29. [\[CrossRef\]](#)
21. Simon, A.; Vinayakumar, R.; Sowmya, V.; Soman, K.P.; Gopalakrishnan, E.A.A. A deep learning approach for patch-based disease diagnosis from microscopic images. In *Classification Techniques for Medical Image Analysis and Computer Aided Diagnosis*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 109–127. [\[CrossRef\]](#)
22. Zheng, Q.; Furth, S.L.; Tasian, G.E.; Fan, Y. Computer-aided diagnosis of congenital abnormalities of the kidney and urinary tract in children based on ultrasound imaging data by integrating texture image features and deep transfer learning image features. *J. Pediatr. Urol.* **2019**, *15*, 75.e1–75.e7. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Hernandez-Suarez, D.F.; Ranka, S.; Kim, Y.; Latib, A.; Wiley, J.; Lopez-Candales, A.; Pinto, D.S.; Gonzalez, M.C.; Ramakrishna, H.; Sanina, C.; et al. Machine-learning-based in-hospital mortality prediction for transcatheter mitral valve repair in the United States. *Cardiovasc. Revascularization Med.* **2020**. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Kandil, H.; Soliman, A.; Taher, F.; Ghazal, M.; Khalil, A.; Giridharan, G.; Keynton, R.; Jennings, J.R.; El-Baz, A. A novel computer-aided diagnosis system for the early detection of hypertension based on cerebrovascular alterations. *Neuroimage Clin.* **2020**, *25*, 102107. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Rajagopal, R. Automated arrhythmia classification for monitoring cardiac patients using machine learning techniques. In *Classification Techniques for Medical Image Analysis and Computer Aided Diagnosis*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 153–177.



26. Amami, R.; Smitib, A. An incremental method combining density clustering and support vector machines for voice pathology detection. *Comput. Electr. Eng.* **2017**, *57*, 257–265. [\[CrossRef\]](#)
27. Samikannu, R.; Ravi, R.; Murugan, S.; Diarra, B. An Efficient Image Analysis Framework for the Classification of Glioma Brain Images Using CNN Approach. *Comput. Mater. Contin.* **2020**, *63*, 1133–1142. [\[CrossRef\]](#)
28. Rahman, M.M.; Ghasemi, Y.; Suley, E.; Zhou, Y.; Wang, S.; Rogers, J. Machine learning based computer aided diagnosis of breast cancer utilizing anthropometric and clinical features. *IRBM* **2020**. [\[CrossRef\]](#)
29. Moon, W.K.; Lee, Y.W.; Ke, H.H.; Lee, S.H.; Huang, C.S.; Chang, R.F. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Comput. Methods Programs Biomed.* **2020**, *190*, 105361. [\[CrossRef\]](#)
30. Renz, D.M.; Böttcher, J.; Diekmann, F.; Poellinger, A.; Maurer, M.H.; Pfeil, A.; Streitparth, F.; Collettini, F.; Bick, U.; Hamm, B.; et al. Detection and classification of contrast-enhancing masses by a fully automatic computer-assisted diagnosis system for breast MRI. *J. Magn. Reson. Imaging* **2012**, *35*, 1077–1088. [\[CrossRef\]](#)
31. Hospital San Bartolomé. Official Website Consulted on February 4th, 2020. 2020. Available online: <https://www.sanbartolome.gob.pe/> (accessed on 5 April 2021).
32. World Health Organization. International Classification of Diseases for Mortality and Morbidity Statistics (10th Revision). 1992. Available online: <https://icd.who.int/browse10/2016/en> (accessed on 5 April 2021).
33. Meesad, P.; Hengpraprom, K. Combination of KNN-Based Feature Selection and KNNBased Missing-Value Imputation of Microarray Data. In Proceedings of the 2008 3rd International Conference on Innovative Computing Information and Control, Dalian, China, 18–20 June 2008; p. 341.
34. Troyanskaya, O.; Cantor, M.; Sherlock, G.; Brown, P.; Hastie, T.; Tibshirani, R.; Botstein, D.; Altman, R.B. Missing value estimation methods for DNA microarrays. *Bioinformatics* **2001**, *17*, 520–525. [\[CrossRef\]](#)
35. Buuren, S.v.; Groothuis-Oudshoorn, K. MICE: Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **2011**, *45*, 1–67. [\[CrossRef\]](#)
36. Torres, R.; Judson-Torres, R.L. Research Techniques Made Simple: Feature Selection for Biomarker Discovery. *J. Investig. Dermatol.* **2019**, *139*, 2068–2074. [\[CrossRef\]](#)
37. Sanchez, N.; Alonso, A.; Calvo, R.M. A Wrapper Method for Feature Selection in Multiple Classes Datasets. In *Bio-Inspired Systems: Computational and Ambient Intelligence*; Cabestany, J., Sandoval, F., Prieto, A., Corchado, J.M., Eds.; Springer: Berlin/Heidelberg, Germany; Salamanca, Spain, 2009; pp. 456–463.
38. Tiwari, S.; Singh, B.; Kaur, M. An approach for feature selection using local searching and global optimization techniques. *Neural Comput. Appl.* **2017**, *28*, 2915–2930. [\[CrossRef\]](#)
39. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
40. Ponomareva, N.; Colthurst, T.; Hendry, G.; Haykal, S.; Radpour, S. Compact multi-class boosted trees. In Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; pp. 47–56.
41. Braga, I.; do Carmo, L.P.; Benatti, C.C.; Monard, M.C. A Note on Parameter Selection for Support Vector Machines. In *Advances in Soft Computing and Its Applications*; Castro, F., Gelbukh, A., González, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 233–244.
42. Feurer, M.; Hutter, F. *Hyperparameter Optimization*; Springer: Cham, Switzerland, 2019.
43. Castro, W.; Oblitas, J.; De-La-Torre, M.; Cotrina, C.; Bazan, K.; Avila-George, H. Classification of Cape Gooseberry Fruit According to its Level of Ripeness Using Machine Learning Techniques and Different Color Spaces. *IEEE Access* **2019**, *7*, 27389–27400. [\[CrossRef\]](#)
44. Nguyen, H.; Bui, X.; Bui, H.; Cuong, D. Developing an XGBoost model to predict blast-induced peak particle velocity in an open-pit mine: A case study. *Acta Geophys.* **2019**, *67*, 477–490. [\[CrossRef\]](#)
45. Stoll, C.; Alembik, Y.; Dott, B.; Roth, M.P. Associated malformations in patients with esophageal atresia. *Eur. J. Med. Genet.* **2009**, *52*, 287–290. [\[CrossRef\]](#) [\[PubMed\]](#)
46. German, J.C.; Mahour, G.H.; Woolley, M.M. Esophageal atresia and associated anomalies. *J. Pediatr. Surg.* **1976**, *11*, 299–306. [\[CrossRef\]](#)
47. Xiao, B.; Luo, P.C.; Cheng, Z.J.; Zhang, X.N.; Hu, X.W. Systematic Combat Effectiveness Evaluation Model Based on Xgboost. In Proceedings of the 2018 12th International Conference on Reliability, Maintainability, and Safety (ICRMS), Shanghai, China, 17–19 October 2018; pp. 130–134.
48. World Health Organization. International Classification of Diseases for Mortality and Morbidity Statistics (11th Revision). 2018. Available online: <https://icd.who.int/browse11/l-m/en> (accessed on 5 April 2021).