

# Annals of Clinical and Medical Case Reports

Technical Paper

ISSN: 2639-8109 | Volume 8

## Structure and development of a clinical decision support system: application in high Digestive Gastroenterology

Lagareiro A and Domingues N\*

Polytechnic Institute of Lisbon, Higher Institute of Engineering of Lisbon, Portugal

### \*Corresponding author:

Nuno Domingues,  
Rua Conselheiro Emidio Navarro,  
1, 1959-007 Lisbon, Portugal,  
E-mail: nuno.domingues@isel.pt

Received: 11 Feb 2022

Accepted: 21 Feb 2022

Published: 28 Feb 2022

J Short Name: ACMCR

### Copyright:

©2022 Domingues N. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and build upon your work non-commercially.

### Citation:

Domingues N, Structure and development of a clinical decision support system: application in high Digestive Gastroenterology. *Ann Clin Med Case Rep.* 2022; V8(11): 1-13

### Keywords:

Esophagus' Cancer; stomach's cancer; Clinical decision support system; Classification; Therapeutic decision

### 1. Abstract

Clinical decision requires to infer great, diverse and not suitably organized quantity of information and having low time to decide. The therapeutic choice is fundamental to formulate a strategy to avoid complications and to achieve favorable results, being more important in some specialties. In addition, medical decision-makers are overloaded with clinical tasks, have an intense work rate and are subject to a great demand, and are prone to greater tiredness. In this sense, computer tools can be extremely useful, as can deal with a lot of information in much less time than decision-makers. Thus, the existence of a tool that assists them in decision-making is of crucial importance.

This paper reflects on a system of assistance in clinical decision making developed and that had its application in Gastroenterology of the Upper Digestive. Its development was divided into three main stages: obtaining the database, subject to treatment and through which subsets were created for use in the following phases; the definition of the model and construction of the system, for testing and combinations of the various subsets with the Algorithms Multiclass Neural Network and Multiclass Decision Forest and using holdout and cross validation methods; testing the application of the template through a web-service, with support of Excel and a web app.

The model created presents or good performance, however, in its application for new cases, there was a decrease in performance. And there is room for development and improvement. To this end, the first step is to fill in all variables that have weight in the therapeutic decision. This will make it possible to obtain a more robust database and more grounded therapeutic suggestions.

<http://acmcaseports.com>

### 2. Introduction

Gastric and esophageal cancers represent the 5th and 7th most common neoplasms worldwide, with over 1,000,000 and 500,000 new cases in 2018, respectively [1]. Late presentation of symptoms is common in both, with approximately half of patients with esophageal tumors and up to 65% of patients with gastric tumors exhibiting locally advanced or metastatic disease at diagnosis [2]. Consequently, survival is low, with 12.0% and 19.9% of patients living 5 years after diagnosis, in the United Kingdom and the United States of America, respectively, for cases of esophagus, and 17.0% and 32.0% for stomach cases [3-6]. Thus, the therapeutic choice is fundamental to achieve favorable results in this patient population. However, the decision is based on a broad set of information and with little time to decide. In this sense, it tools can be extremely useful because they can process large amounts of information in much less time than decision makers. Several decades ago, with the growth of computational power, large data sets were developed and studied, as well as methods for their classification. As a result, a large number of applications have been created, such as Decision Support Systems (DSS) [7]. A DSS is a software developed to help professionals solve problems or evaluate opportunities, that is, the purpose of its use is to support and improve the making of one or more decisions associated with a given area of knowledge [8].

In the clinical context, these software provides specific and intelligently filtered information, in order to directly impact the decision-making of a clinician, with the objective of improving health care [8, 9], correlating the characteristics of an individual patient and a computerized knowledge base, and then presented specific

assessments or recommendations regarding the patient whose decision to therapy, or other action, is at stake [8-10].

Medical errors are costly, causing thousands of deaths worldwide every year [11], and with the increased focus on its prevention, which has occurred since the publication of the Institute of Medicine report, *To Err Is Human* [12], cdss (Clinical Decision Support Systems) have been proposed as key elements to improve patient safety [7], its potential to reduce errors and increase the quality and efficiency of health care [10]. Thus, health organizations face increasing pressures to improve the quality of care and reduce costs, taking an interest in improving medical practices, disease management and the use of resources through artificial intelligence(AI) [7], which has been shown to be very effective in the development of clinical support systems [11].

In general, all health services have internal protocols that guide and guide professionals in decision-making in various situations. This decision-making system, based on experience and knowledge, is defined in protocols that aim for a good procedure of action, protege the clinical decision-making and facilitates timely decision-making. In the case of a therapeutic decision, this reality of following a defined path is even more pronounced, especially in the oncological aspect. Although most protocols are presented in a simplified way, each step implies complicated decision-making, as there are several factors and criteria to be taken into account that influence the treatment plan of an individual patient and can considerably change the course to follow. Thus, a decision support system that incorporates the information needed to make an adjusted decision, and according to the protocol of the institution concerned, can bring enormous advantages, for the decision-makers, for the patient and for the institution itself.

In this sense, with the present work, it was intended to study a system of decision aid and develop its application in Gastroenterology of the Upper Digestive, being used coded data from the institute Portuguese of Oncology of Lisbon.

### 3. Methodologies

The development of this work was divided into three stages: the first was to obtain a structured database to be used as input of the

model of the decision support system; the second was the definition of the model and the construction of a system that receives the database as input and has as output recommendations to support the therapeutic decision; and the third consisted of testing and validating the operation of the system.

#### 3.1. Database

A database on IPO patients was used, d the multidisciplinary group of cancer of the esophagus and stomach (GMCEE), which resulted from the combination of an older database, with data since 1937, and a more recent one by the GMCEE, with the inclusion of more variables and criteria, for as complete information as possible. The database resulting from this junction consists of 1982 records with 184 variables.

#### 3.2. Data Protection

Since the database used contains actual data, there was a concern to ensure the protection of them. Thus, a coding of patients' personal data was created and the coding key is known only by the IPO medical team. From the original database, only the information needed to perform this FMT was used, and the data provided are not specific and do not allow the identification of patients.

#### 3.3. Data Processing

Data processing tasks, such as screening, deletion and formatting, were performed in order to obtain a usable data set from the initial raw data. The selection of variables was performed based on the protocol provided by the IPO, in which the process is sum-able from diagnosis and staging to the selection of the most appropriate treatment and follow-up. Variables that do not add value, such as institutional and management data and information, such as dates and order states, were eliminated. In addition to these, the variables that allowed the identification of the patient were eliminated, that is, the process number and the name, maintaining the ID, which is a fictitious number and is useful only for the distinction of the records informatically.

As heterogeneity was observed in filling variables, simple changes were made, shown in Table 1, in order to reduce inconsistencies, reduce variability and eliminate errors.

**Table 1:** Changes in the completion of records in the database.

Variable	Original fill	Alteration
Initial biopsy	1st "GIST??" 20th "GIST"	21st "GIST"
	1 "not available" 1 "not effected" 2nd "ND"	4th "ND"
	94 "cell clower carcinoma" 113 "CPC"	207 "CPC"
	79 "adenocarcinoma" 117th "ADC"	196th "ADC"
Base Table Degree Differentiation	10th "NR" 6th "ND"	16th "ND"
Base Table Location	1 "body and den" 1 "den/gastric body" 1 "body/den transition"	3 "body/den"

<b>Weight loss</b>	4 "0" (should be filled with Yes, No or ND)	4th "No"
<b>Anemia/HDA</b>	3 "0" (should be filled with Yes, No or ND)	3rd "No"
<b>Obstruction</b>	4 "0" (should be filled with Yes, No or ND)	4th "No"
<b>Staging - EcoEDA - T</b>	1st "uT1N0"	1 "1"
<b>Staging - EcoEDA - N</b>	1st "N0"	1 "0"
<b>PET staging - result</b>	80 "tumor and N+"	83 "tumor and N+"
	3 "tumor-only and N+" marking	
	37 "tumor" 2 "tumor-only marking"	39 "tumor"
<b>Staging TC - T</b>	2 "4?"	2 "4"
<b>Staging TC - N</b>	2 "1?"	2 "1"
	1 "n0???"	1 "0"
<b>Staging TC - M</b>	1 "0 see description"	1 "0"
	13 "1?"	13 "1"
<b>Cn</b>	5 "(+)"	5 "+"
<b>Table Base Surgery Location</b>	1 "cardia siewert II??"	1 "cardia siewert II"
<b>ASA score</b>	20th "I"	24th "I"
	4 "1"	
	106 "2"	517 "II"
	411 "II"	
	35 "3"	
155 "III"	190th "III"	

### 3.4. Creating Subsets

Since a large number of unfilled records were found in several variables, it was necessary to build new subsets of the database, including only variables of interest and without blank records. Different subsets were created and compared to each other in order to understand which is most feasible to use. For the creation of the first subset, only records for the years 2017, 2018 and 2019 were considered, i.e., the three most recent years contained in the database, since they were pointed out by the GMCEE as having a higher probability of more complete completion of the various variables. Thus, all blank records for each of the variables of interest were eliminated, or filled out where possible, to the extent that only fully completed records remained. For the second subset, no ipo registration year was considered as a basis, thus having a starting point with more content. Moreover, it has advanced in a similar way to the first version. In the construction of the third subset, all records with the variables of interest filled were considered. All others were deleted, not being filled records, as happened in previous versions.

### 4. Therapeutic Decision Analysis

The initial therapy was compared in the subsets with the initial therapy expected by the protocol, in order to verify the existence of discrepancies. As the organ variable only has the options esophagus, cardia and stomach, and cardia treatment can follow two approaches, depending on the specific location, for this analysis this distinction is necessary. In addition, the stage is very important in the choice of the following therapy. Thus, the variables for and this phase were "table base\_cirurgia location", "cT", "cN", "cM" and "Initial Tx". The variable "base\_cirurgia table" presents as an option "Cárdia Siewert I/II/III", which by itself does not allow to know which protocol to follow. Thus, cardia I/II/III records in the various subsets were not analyzed. For the analysis of esophagus cases, the options "lower thoracic esophagus", "middle thoracic

esophagus", "upper thoracic esophagus", "siewert I" and "cardia siewert I/II" options were selected in the variable "table base\_cirurgia location". In the case of stomach, "den", "stump", "body", "bottom", "notch", "cardia II", "cardia II/III" and "cardia III" were selected.

### 5. System Development

Having the three subsets of the database ready for use, the classification phase was carried out with them, in order to understand which one provides the best results, and then the created model was application. For all phases of classification was used the Microsoft Azure Machine Learning Studio platform, which allows the creation and management of machine learning projects using modules interconnected with each other, and for the application phase was used this platform, but also Microsoft Visual Studio.

#### 5.1. Classification

To perform the classification tests for each of the three subsets created, as represented in Table 2, several combinations of factors were made: classifier used, number of variables considered in each subset and method of separation of the data, in addition to the modification of some of the parameters that concerns each one. The classifiers used were the Multiclass Neural Network and the Multiclass Decision Forest. the number of variables used was 17 (configuration 1) or 32 (configuration 2) and the methods of data division were holdout or cross validation. The choice of the method of data division influences the distribution of modules in the work schema, as can be seen in Figure 1. The classifiers have several parameters that can be adjusted, however, most were left intact, with the default options.

In the case of the Multiclass Neural Network classifier, the following parameters were adjusted: "hidden layer specification", which allows you to choose the architecture for the network, having selected the option "fully-connected case", which is the standard

neuronal network architecture for multiclass models;

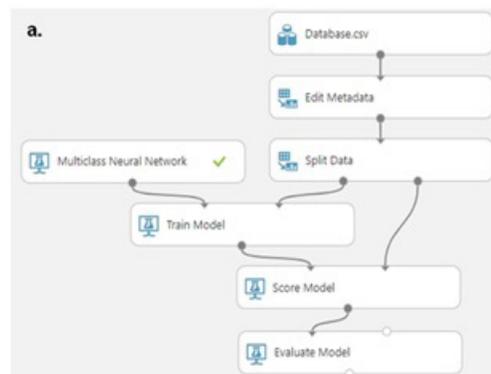
"number of hidden nodes", i.e. neurons in the hidden layer, and this parameter was maintained with the value 100 in all tests; and "number of learning iterations", which is the maximum number of times that the algorithm should process training cases, and this value was varied between 1 and 100.

For tests with the Multiclass Decision Forest classifier, the following parameters were adjusted: "number of decision trees", which allows you to choose the maximum number of decision trees that can be created in the set; "maximum depth of decision

trees", which allows to limit the maximum depth of any decision tree; "number of random splits per node", which is the number of divisions to use when constructing each node in the tree, and a split means that the characteristics at each tree level (node) are randomly divided; and "minimum number of samples per terminal node", which is the minimum number of cases required to create any terminal node (leaf) in a tree. Once the parameters of the algorithms were defined, and the remaining modules, and after their execution, the classifier performance for each test was evaluated, using the accuracy metric.

**Table 2:** Tests performed for each subset.

Classifier	Configuration	Method	Parameters
Multiclass Neural Network OR Multiclass Decision Forest	1	Holdout	Split: 0.7 - 0.3 ; random split Classifier: vary parameters
			Split: vary proportion; random split Classifier: fixed parameters
	2	Cross Validation	Classifier: standard parameters
			Classifier: parameters with better results in the holdout <i>method</i>
	1	Holdout	Split: 0.7 - 0.3 ; random split ; stratified split Classifier: vary parameters
			Split: vary proportion; random split; stratified division Classifier: fixed parameters
	2	Cross Validation	Classifier: standard parameters
			Classifier: parameters with better results in the holdout <i>method</i>



**Figure 1:**

**5.2. Application**

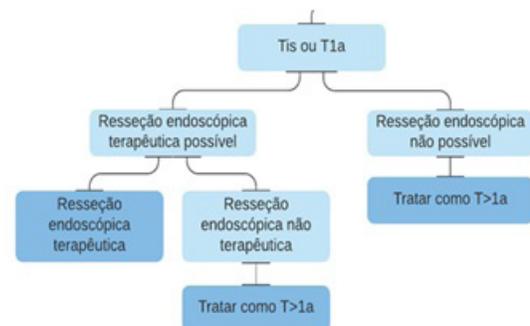
After analyzing the performance results obtained for the various subsets in the various configurations, one was selected for incorporation into the model and explored the application of this through the creation of a web service, which allows the model to be used in various applications regardless of the hardware or software platform in which it was implemented. The web service was then applied through excel and a web app.

**6. Theoretical Concepts**

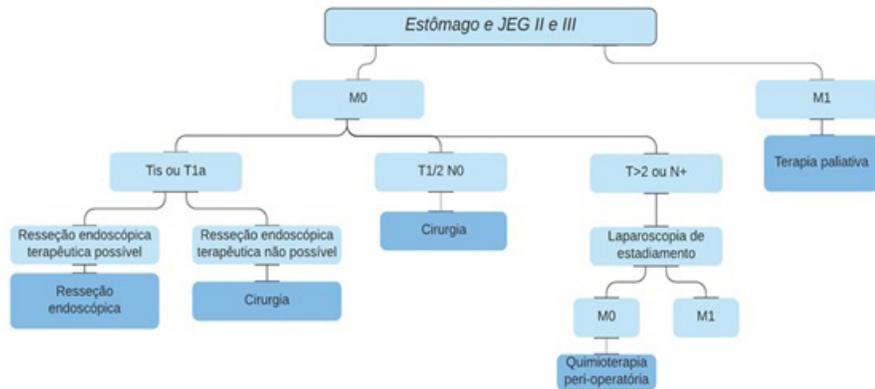
**6.1. Therapeutic Approaches**

Contemporary therapies are stage-specific and highly complex, including organ preservation techniques, minimally invasive options, and multimodality therapy, which includes combinations of surgery, chemotherapy, and radiation. According to the IPO protocol, which describes the treatment for Adenocarcinoma (ADC)

and Cell Carcinoma (CPC), the treatment modalities for cancer of the oesophagus and stomach are those presented in Figuras 2 and 3. In cases of gastroesophageal junction neoplasia (JEG), the following approach will be that of oesophagus if it is JEG I or stomach cancer if it is JEG II or III [13]. Cases of JEG I/II are usually treated as the oesophagus.



**Figure 2:** Therapeutic decision-making scheme for oesophageal cancer.



**Figure 3:** Therapeutic decision scheme for stomach cancer.

**6.2. Classification**

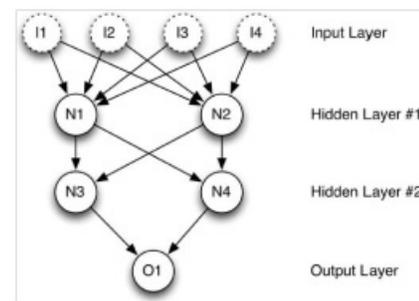
Classification is the process of finding a model to designate the information presented in classes and categories of the same type [14]. The model is derived based on the analysis of training data, for which the class is known, and is later used to predict the class of objects for which the class is unknown [15]. A variety of classification algorithms have been proposed in the literature for intelligent medical applications, including neural networks and decision trees. A sort task with more than two classes is called multiclass classification. For example, determining for several patients which therapy is most appropriate between surgery, chemotherapy, or radiotherapy is a multiclass classification task. The multiclass classification assumes that each object is assigned only one label: therapy can be chemotherapy or surgery, but not both simultaneously.

**6.3. Multiclass Neural Network**

Artificial Neural Networks (ANNs) were created to mathematically model human intellectual abilities through biologically plausible engineering creations [16, 17]. Being biologically inspired, ANNs are made up of artificial neurons (nodes) and synapses (edges) [16, 18]. ANNs have at least one input layer and an output layer. A pattern is displayed to the input layer, and then an output pattern is obtained on the output layer. Among these layers, in the hidden layers, it is not known for sure what happens, what is known is that neurons add up to the input of all input synapses, apply a nonlinear function and emit the computation result for all output synapses [18]. Figure 4 shows that artificial neurons N1 and N2 receive input from one or more sources, which may be other neurons or data fed into the network, and multiply each input by a weight. They then add these multiplications and pass the sum to an activation function. For the construction of an ANN, several neurons are chained, as visible in the figure, which represents an ANN composed of four neurons. This ANN has four inputs and one output. The outputs of neurons N1 and N2 feed N3 and N4 to produce output O.

The classifier used in Microsoft Azure Machine Learning Studio, Multiclass Neural Network, is a learning method that allows you to create a neuronal network model that can be used to predict an object with various possible values. The relationship between

inputs and outputs is learned by training the neuronal network in the input data. The direction of the chart proceeds, from the inputs, through the hidden layer(s) to the output layer. To calculate the output value for a given input, a value is calculated on each node in the hidden layers and in the output layer, through the weighted sum of the node values of the previous layer, and an activation function is applied to the weighted sum [19].



**Figure 4:** Structure of a completely interconnected neuronal network.

**6.4. Multiclass Decision Forest**

A decision tree is a flowchart with a tree-like structure consisting of the root (root node), branches, inner nodes, and leaves (terminal nodes), organized hierarchically, as represented in Figure 5, in which the root and inner nodes are represented with circles and the terminal nodes (leaves) with squares. All nodes (except root) have exactly one input border [15, 20]. Root and internal nodes branch the decision process, while leaves assign classes [17]. That is, each inner node corresponds to a question, representing a test on an attribute value; each branch represents the answer to the question, that is, a test result; and the leaves of the tree represent classes or distributions of classes, that is, the final decision, obtained based on all tests performed [15, 18]. Thus, the object to be sorted travels a path from the root to the leaf, where a class is assigned [15, 17]. Decision trees are easy to understand, can be interpreted intuitively by humans and can be easily converted into classification rules [15, 18].

Generally, set models provide better coverage and accuracy than single decision trees. Thus, the classifier used in Microsoft Azure Machine Learning Studio (classic), Multiclass Decision Forest, is

a learning method that consists of a set of decision trees. The algorithm builds multiple decision trees and then votes in the most popular exit class. Voting is a form of aggregation, in which each tree produces an un-normalized frequency histogram. The aggregation process sums these histograms and normalizes the result to get the "probabilities" of each class. Trees that have high forecast confidence have greater weight in the final decision of the set [20, 21].

### 6.5. Data Division

There are several methods for dividing the data for classification, being very common the techniques holdout and cross validation [15].

In the holdout method, the data provided is randomly divided into two independent sets. Typically, two-thirds of the data is allocated to the training set, used to derive the model, and the remaining third is retained in the test set for further estimation of the accuracy of the model, hence the name holdout [15]. It is important that the test data is not used in any way to create the classifier, because to predict the performance of a classifier it is necessary to evaluate its error rate in a set of data that has not played any role in the formation of the classifier [22].

A more general way to mitigate any bias caused by the specific sample chosen for holdout is to repeat the entire process, training and testing, several times with different random samples. In each iteration, a certain proportion, for example two-thirds, of the data is randomly selected for training and the remainder used for testing. In this method, called cross validation, also known as k-fold cross validation, the initial data is randomly divided into k subsets, or folds, D1, D2,..., Dk, each of approximately equal size. The training and the test are performed k times. In iteration i, the Di subset is reserved as the test set, and the rest are used collectively to train the model. That is, in the first iteration, the Subsets D2,..., Dk collectively serve as the training set to obtain a first model, which is tested in D1; the second iteration is trained in subsets D1, D3,..., Dk and tested on D2; and so on. Unlike the holdout method,

here each sample is used the same number of times for training and once for testing. For classification, the accuracy estimate is the total number of correct classifications of the k iterations, divided by the total number of classes in the initial data [15].

## 7. Results and Discussion

### 7.1. Therapeutic Decision

Regarding the comparison of the therapy performed with the therapy expected by the decision protocol used in the IPO, the results obtained for the three subsets are represented in Table 3. It was found that both for oesophagus and stomach, the results observed did not vary much between the three subsets.

In an oesophagus, it is possible to verify that, for some cases, the therapy was performed does not coincide with what was expected by the protocol, however, for each of these cases a justification was found, given by the GMCEE. For example, in the esophagus regimen, there is an option that is neo-adjuvant chemo-radiotherapy, however, in the database there are some cases with neo-adjuvant chemotherapy, this is because these patients underwent chemotherapy in other institutions and were only referred for surgery on IPO after that. As for the other cases, considered as coincident, it was found that one of two options could be performed and, in fact, one of these options was performed, but the information contained in the database does not allow us to understand which factors led to the decision between the two options in question.

In the stomach, there was a higher proportion of cases that did not coincide, also having justifications similar to that previously presented for oesophagus. In addition to these and those that are within the expected options, it was also possible to observe completely coincident cases.

Based on the results obtained, and after a new screening, in which cases of cardia I/II/III and others with contradiction between variables were eliminated, only the coincident cases were maintained, thus the total number of records 115, 396 and 281, for subset 1, 2 and 3, respectively.

**Table 3:** Analysis of therapeutic decision in the various subsets.

	Oesophagus			Stomach		
	Subset 1	Subset 2	Subset 3	Subset 1	Subset 2	Subset 3
Matches	-	-	-	-	361 (66,4%)	259 (67,6%)
Within the options	56 (81,2%)	69 -81%	15 -75%	70 -66%	10 (1,8%)	8 (2,1%)
Does not match	13 (18,8%)	16 -19%	5 -25%	36 -34%	173 (31,8%)	116 (30,3%)

### 7.2. Classification

After importing the complete file from each subset into Azure Machine Learning Studio, it was noticed that the new variables had missing values, that is, blank values. Consequently, in the execution of the training of the model a message of error appeared. The acceptable solution to the problem, to keep all records, would be to fill in the missing values manually, a time-consuming approach <http://acmcaseports.com>

that may not be feasible due to large data sets with many missing values, performed by doctors, if they had access to information in another format. Not being feasible to follow this path and since no other option would be adequate, because it was not intended to tamper with data, but rather have real data, two alternatives remained: eliminate cases with blank variables or eliminate the variables keeping the cases. As the number of cases had already been sharply reduced, the second option was chosen.

### 7.3. Multiclass Neural Network

Initially, the holdout method was used and, in relation to the parameters of the Multiclass Neural Network classifier, the data were divided with a fraction of 0.7, with the random method, the number of neurons of the hidden layer was maintained as 100 and the number of learning iterations between 1 and 100 was varied.

In configuration 1, although not linear, in the subset, there was an increase in accuracy with the increase in the number of learning iterations, with accuracy of 32.35% for a single iteration but with all correct classifications, i.e., 100% accuracy, from 23 iterations. In subset 2, it was verified that initially with the increase in the number of iterations, from 1 to 12, in general, greater accuracy was obtained, although oscillations were observed; from 13 to 30 iterations, there was a decrease compared to the previous figures, and there were also some fluctuations; from 31 to 34, the highest accuracy was achieved (96.12%); from 35 to 38 iterations the accuracy decreased to 92.44% and, although in the following iterations, from 39 to 78, if there was a new increase, this never exceeded the maximum accuracy already obtained and, from 79 to 100 iterations, the value remained at 92.44%. In subset 3, the overall accuracy increased several times throughout the tests with the increase in the number of iterations, although with some oscillations. Initially, from 1 to 14 iterations there was always growth, with 61.90% for 1 iteration, 88.10% for 2 to 7 iterations, 89.29% for 8 iterations and 90.48% for 9 to 14 iterations. From 15 to 23 iterations the accuracy was 92.89%, except for the 22 iterations, with 94.05%. From 24 to 80 iterations, the accuracy was 94.05%, with some exceptions: for 44 to 46, 48, 49, 53 and 58 iterations, it was 95.24%, the highest value observed, already for 79 iterations, decreased to 92.86%. From 81 to 100 iterations, there was a new decrease, returning to 92.86%. For iterations from 95 to 97 decreased further, to 91.67%. Table 4 summarizes the best values obtained for each subset.

For configuration 2, it was proceeded in the same way as in configuration 1, but also activating the stratified division. In subset 1 there was an increase in accuracy with the increase in the number of learning iterations, and for a single iteration the accuracy was 44.12% and from 70 iterations all classifications were correct, i.e.,

the accuracy was 100%. In subset 2, in general, no major variations in accuracy were observed, although for 1 iteration the value is substantially lower. From 1 iteration to 2, accuracy increased from 76.27% to 94.07%, thus remaining until 22 iterations, except for 3 and 4 iterations, whose accuracy was 93.22%, as well as for 23 and 24 iterations. From 25 to 85 iterations the value was 94.92%, and from 86 to 100, it ranged from 94.92% to 95.76%, the highest value obtained. In subset 3, for 1 iteration the accuracy was 51.19% and then there was an increase in iterations 2 to 6, 7, 8 to 10, 11 to 13, 14 to 25 and 26 to 50, to 90.48%, 92.86%, 95.24%, 96.43%, 97.62% and 98.81%, respectively. From 51 to 80 iterations, accuracy decreased to 97.62%, and from 81 iterations the highest value previously observed was resumed. Table 5 summarizes the best values obtained for each subset.

Moving on to the tests performed with the cross validation method, in configuration 1, the initial dataset was divided into 10 subsets, each consisting of 11 or 12 instances, for subset 1. In the case of subsets 2 and 3, it was not possible to perform the tests using the cross validation method, due to an error identified as impossibility of shuffling in the training data. Thus, the shuffle option, selected in the standard parameters of the classifier, was deactivated and the retry was made. Still, it continued to remain an error message, referring to an internal system error. We could not solve this problem or understand why in subset 1 all modules worked correctly and in subsets 2 and 3 did not. For the tests with subset 1, for comparison purposes, the standard values for the classifier parameters were used, but also the values with the best results in the holdout method. The accuracy values are found in Table 6, and it is possible to verify that they are the same in subset 1, since the standard parameters and parameters with the best results in the holdout method are the same.

In configuration 2, as in 1, the initial dataset was divided into 10 subsets, and again, in the case of subsets 2 and 3, the tests could not be performed due to the same errors. The standard values of the classifier were used, but also the values with the best results in the holdout method, and in this case are also equal. The accuracy values are found in Table 7, and it is possible to verify that they are the same in subset 1, because the standard parameters and parameters with the best results in the holdout method are the same.

**Table 4:** Comparison of the highest accuracy obtained in each test set.

	<i>Subset 1</i>	<i>Subset 2</i>	<i>Subset 3</i>
<b>Accuracy</b>	100,00	96,12	95,24

**Table 5:** Comparison of the highest accuracy obtained in each test set.

	<i>Subset 1</i>	<i>Subset 2</i>	<i>Subset 3</i>
<b>Accuracy</b>	100,00	95,76	98,81

**Table 6:** Accuracy obtained for configuration 1 tests with the cross validation method.

Accuracy	Subset 1		Subset 2		Subset 3	
	Standard parameters	Parameters used in the holdout method	Standard parameters	Parameters used in the holdout method	Standard parameters	Parameters used in the holdout method
	96,52	96,52	Error	Error	Error	Error

**Table 7:** Accuracy obtained for configuration 1 tests with the cross validation method.

Accuracy	Subset 1		Subset 2		Subset 3	
	Standard parameters	Parameters used in the holdout method	Standard parameters	Parameters used in the holdout method	Standard parameters	Parameters used in the holdout method
	95,65	95,65	Error	Error	Error	Error

#### 7.4. Multiclass Decision Forest

The tests were performed for each subset, first using the holdout method and then the cross validation method, making the various parameters of the classifier vary at a time, starting with the number of decision trees and moving on to the maximum depth of the decision trees, number of random divisions per node and minimum number of samples per terminal node.

In configuration 1, in subset 1, with the increase in the number of decision trees, an increase in accuracy was observed only once, maintaining this value in all tested numbers. Then, increasing the maximum depth of the decision trees, the scenario was similar. In the variation of the third parameter, number of random divisions per node, from 100 to 159 the accuracy value remained the same as that already obtained, and for divisions from 160 it was possible to obtain maximum accuracy, 100%.

Regarding subset 2, changing the number of trees from 8 to 100, there was a decrease in accuracy; for values from 200 to 800, there was an increase, still lower than the first value obtained; and for 1000 there was a new decrease. Varying the depth of the decision tree, there was no change in accuracy in relation to that obtained for the predefined parameters. Changing the number of random splits per node occurred several changes, however, none of them exceeded the highest value ever obtained. Varying the fourth parameter, minimum number of samples per terminal node, a decrease in the value was observed. Thus, the highest accuracy was 95.7983%.

As for subset 3, increasing the values of the first parameter, there was a decrease in accuracy, compared to the presets, except for the values 500 and 800, for which the performance was equal. Increasing the value of the second parameter and keeping the first as 8, the accuracy was the same; changing to 100 decreased; and changing to 500 resumed the initial value. By varying the third parameter, from 100 to 150 the accuracy decreases and from 200 to 500 remained the same. When the fourth parameter varied, the highest accuracy was obtained, 91.6667%.

Table 8 summarizes the accuracy values of the tests performed. Subset 1 stands out for its 100% accuracy, subset 2 has the second best performance and subset 3 was the one that got the farthest, although with a very good result.

Moving to configuration 2 with the holdout method, in subset 1, varying only the values of the first parameter, it was possible to

obtain accuracy of 100%, for values equal to or greater than 160.

For subset 2, either by varying the first parameter or the second, the accuracy values obtained were always the same, equal to the predefined ones. By changing the number of random divisions per node some changes were observed: for 8 trees, with a maximum depth of 32 and 100 random divisions, the accuracy decreased, while for 100 and 200 trees with a maximum depth of 32 and for values of 100 in the three parameters, with 100 random divisions, the accuracy increased, to the highest value so far. In the remaining attempts the results were lower. Varying the fourth parameter, the results were lower. Thus, the maximum accuracy obtained in subset 2 was 96.6102%.

In subset 3, both for variations in the first parameter and in the second, the results remained the same as those of the predefined parameters. Varying the third parameter and maintaining the predefined values in the two previous parameters, a higher value was observed and a value lower than that already observed. When the fourth parameter was varied, with the first two with predefined values and the third with 100, the value decreased to the lowest observed. Thus the maximum accuracy in this subset was 98.8095%.

Table 9 summarizes the accuracy values of the tests performed. Subset 1 has the best performance in terms of accuracy, with 100%, and subset 2 is the one with the lowest accuracy, albeit with a very good value of 96.61%.

In the tests performed with the cross validation method, in configuration 1, it was possible to obtain results for all subsets. The initial data set was again divided into 10 subsets, each consisting of 11 or 12, 39 or 40 and 28 or 29 instances, for subsets 1, 2 and 3, respectively, and the standard values of the classifier were used, but also the values with the best results in the holdout method, for comparison. The accuracy results are found in Table 10, from which it can be concluded that the subset with the best results was the 2 and that with the standard parameters, the performance result obtained is higher than with the parameters used in the holdout method, except for subset 2, in which the results are even. In configuration 2, the entire procedure was as described in configuration 1. The accuracy results are found in Table 11, from which it can be concluded that the subset with the best results was the 2 and that, as verified in configuration 1, with the same parameters used in the holdout method, the performance result obtained is higher than with the standard parameters, with the exception of subset 1, in which the results are equal.

**Table 8:** Comparison of the highest accuracy obtained in each test set.

	<i>Subset 1</i>	<i>Subset 2</i>	<i>Subset 3</i>
<b>Accuracy</b>	100,00	95,80	91,67

**Table 9:** Comparison of the highest accuracy obtained in each test set.

	<i>Subset 1</i>	<i>Subset 2</i>	<i>Subset 3</i>
<b>Accuracy</b>	100,00	96,61	98,81

**Table 10:** Accuracy obtained for configuration 1 tests with the cross validation method.

	<i>Subset 1</i>		<i>Subset 2</i>		<i>Subset 3</i>	
	Standard parameters	Parameters used in the <i>holdout method</i>	Standard parameters	Parameters used in the <i>holdout method</i>	Standard parameters	Parameters used in the <i>holdout method</i>
<b>Accuracy</b>	94,78	93,91	96,21	96,21	95,37	93,95

**Table 11:** Accuracy obtained for configuration 2 tests with the cross validation method.

	<i>Subset 1</i>		<i>Subset 2</i>		<i>Subset 3</i>	
	Standard parameters	Parameters used in the <i>holdout method</i>	Standard parameters	Parameters used in the <i>holdout method</i>	Standard parameters	Parameters used in the <i>holdout method</i>
<b>Accuracy</b>	94,78	94,78	96,98	95,47	95,73	95,37

### 7.5. Fraction of Data Division

Regarding the fraction of data division, it was possible to conclude, based on the results in Table 12, that, in general, with the increase in the training fraction, there is also an increase in the accuracy of the classifier, although some exceptions are visible.

### 7.6. Considerations

Both classifiers tested demonstrated good performance in the various subsets used and tests performed, as visible in Tables 13 and 14, where the best values obtained throughout the various tests performed are summarized. In subset 1, the values of both settings were always the same as each other, with the exception of tests with the cross validation method for the Multiclass Decision Forest classifier. Although it was not possible to obtain results with the cross validation method for subsets 2 and 3, based on the results obtained for subset 1, it can be said that the performance was better with the holdout method, and in both classifiers 100% accuracy was obtained. Comparing the two classifiers, the first was superior. In subset 2, the superiority of one configuration over the other is not evident, with variation depending on the classifier or method used. Still, looking at the two higher values, it is possible to notice that they relate to configuration 2. In the results referring to the Multiclass Decision Forest, there were very close results between the two methods, with all values contained in the range of 95.80% to 96.98%, and the best results of each method differ only 0.37% from each other.

In subset 3, configuration 2 proved to be superior. As for the methods of division, in the Multiclass Decision Forest classifier the holdout method showed superiority in configuration 2 while the cross validation method was superior in configuration 1, not being able to distinguish one of them.

Despite the lack of results in several tests due to internal system

errors, the Multiclass Neural Network classifier showed superior results in six tests, similar in five and lower in only one, so based on these results can be considered as the top classifier.

Regarding the number of variables and the fraction of data division, it was found that, in general, with a greater number of variables, that is, in configuration 2, the classifier performance is better, as well as with a greater number of data in the test group. These results are due to the fact that, in the first case, there is more information for decision-making and, in the second case, the classifier has more data to incorporate and be able to find similarities and patterns that lead to decision-making.

As for the best database, one would expect that with more data, the performance would be better. However, the best results were obtained for subset 1, with 100% accuracy for all configurations tested in any classifier, with the holdout method. These results are in line with the indications of the GMCE, which stated that in this group it would be more likely to have more complete and more reliable records. Although there were some variables to be filled in a posteriori, based on information from others, the fact that they are recent data, with more information, may have helped to make this filling true. In the case of subset 2, which was filled in the same way, as there are more incomplete data, the filling may have led to more errors, leading to lower performance. In subset 3, no filling was performed and, in configuration 2, the data were higher than those of subset 2, which validates the ideas previously presented. Considering the results, although it is desirable to have as many objects as possible, the most appropriate subset for embedding in a CDSS would be subset 1, by the accuracy results presented, or subset 3, which despite having lower accuracy, refers to reliable and unchanged data, in configuration 2 and using the holdout method and the Multiclass Neural Network classifier.

**Table 12:** Influence of the separation fraction of the data on the performance of the classifier.

Fraction test	Multiclass Neural Network						Multiclass Decision Forest					
	Configuration 1			Configuration 2			Configuration 1			Configuration 2		
	Subset 1	Subset 2	Subset 3	Subset 1	Subset 2	Subset 3	Subset 1	Subset 2	Subset 3	Subset 1	Subset 2	Subset 3
<b>0,9</b>	100,00	38,46	44,44	100,00	38,46	44,44	100,00	32,90	44,44	100,00	35,90	44,44
<b>0,8</b>	100,00	92,50	43,64	100,00	92,50	43,64	100,00	95,00	43,64	100,00	95,00	43,64
<b>0,7</b>	100,00	95,76	98,81	100,00	95,76	98,81	100,00	96,61	98,81	100,00	96,61	98,81
<b>0,6</b>	96,43	94,97	96,43	95,65	94,97	96,43	97,83	93,08	96,40	97,83	93,08	93,75
<b>0,5</b>	95,65	93,40	96,41	96,43	93,40	96,41	94,64	93,91	93,75	94,64	93,91	96,40
<b>0,4</b>	94,20	Error	Error	94,20	Error	Error	92,75	92,44	92,90	92,75	92,44	92,90
<b>0,3</b>	92,50	Error	Error	92,50	Error	Error	96,25	92,47	92,39	96,25	92,47	92,39
<b>0,2</b>	90,22	Error	Error	91,30	Error	Error	92,39	91,48	45,67	92,40	91,48	45,67
<b>0,1</b>	79,61	Error	Error	81,55	Error	Error	75,73	37,99	44,25	75,73	37,99	44,25

**Table 13:** Summary of the best accuracy results obtained for the Multiclass Neural Network classifier.

	Multiclass Neural Network					
	Holdout		Cross Validation			
	Config. 1	Config. 2	Config. 1		Config. 2	
			Pattern	Holdout parameters	Pattern	Holdout parameters
<b>Subset 1</b>	100,00	100,00	96,52	96,52	95,65	95,65
<b>Subset 2</b>	96,12	95,76	Error	Error	Error	Error
<b>Subset 3</b>	95,24	98,81	Error	Error	Error	Error

**Table 14:** Summary of the best accuracy results obtained for the Multiclass Decision Forest classifier.

	Multiclass Decision Forest					
	Holdout		Cross Validation			
	Config. 1	Config. 2	Config. 1		Config. 2	
			Pattern	Holdout parameters	Pattern	Holdout parameters
<b>Subset 1</b>	100,00	100,00	94,78	93,91	94,78	94,78
<b>Subset 2</b>	95,80	96,61	96,21	96,21	96,98	95,47
<b>Subset 3</b>	91,67	98,81	95,37	93,95	95,73	95,37

**7.7. Application**

For the creation of the web service, the model obtained with subset 3 was used. Two application options were tested: excel and web app. As for excel, the file was downloaded in Azure Machine Learning Studio and "new patients" were created to obtain suggestions for therapy. When you opened excel, it was found to be blank, but with a section titled "Azure Machine Learning" with the web service available for use. Selecting it and the "use sample data" option, 5 rows and 31 columns of the sheet were automatically filled in, with data from the database, as shown in Figure 6, in which only the right end of the sheet appears. In addition to the 31 columns of the input variables, there is another called "scored labels", in which in the 5 completed records automatically appears the therapy that the model classified. To get suggestions for new cases, it was enough to fill each variable in a new row, and then in the side column, write which input cells and which cell to output, and validate. The corresponding cell was automatically filled in the "scored labels" column, with the suggestion for each case. For the test and, data were filled in for 17 new patients, for esophagus and stomach, and compared the suggestions obtained with the protocol.

As for the second application option, the web app, after its publication, through Visual Studio and the Azure portal, was obtained the link <https://decisaoterapeutica.azurewebsites.net/>, which allows access to the form to fill out and obtain a therapeutic suggestion, as shown in Figure 7. Para test and, the filling was performed with the same data used in excel. <http://acmcaseports.com>

Table 15 shows the results for cases of esophagus, CPC and ADC, and stomach, ADC, and dark green results are represented, light green results corresponding to one of the expected options and red non-coincident results.

It is possible to conclude that there are some inconsistencies, however, it is noticeable that this is the case, since the data set used to create the model only contained 10 cases of esophagus and only 5 were used for the training phase. Of all of them only 1 is ADC, the CPC being, so the results are better than expected for cases of ADC, most likely due to similarity of characteristics in other variables with records used for training.

Regarding cases that could be endoscopic resection or surgery, the result obtained is in line with what was expected, because of the 197 records that were used in the training phase of model only 1 performed endoscopic resection, so, since the classification is made based on these data and through probabilities, it would be unlikely that the suggestion was endoscopic resection, even for a case with all very similar variables. Moreover, even if there are more cases, the suggestion could happen to be surgery, since 85 of the cases used for training performed surgery and some important variables in the decision to perform endoscopic resection were not considered for the training of the model, because a high number of records did not contain information in them. The cases in which surgery was chosen and this is not an option, according to the protocol, may be due to the fact that this therapy was chosen in 85 of the cases used for training and the new cases created may have similarities with these 85 in several variables, leading the system

to suggest this therapy.

As for the last two cases, whose therapy should be palliative QT, because they were filled with 1 in the variable "cM", it was found that the system does not suggest this option for all cases with this characteristic. Once again, the reason may be the small number of training cases for which palliative QT was performed, since there were only 3, and the similarity in other variables with cases that underwent surgery.

Based on these results of the application of the model, it is con-

cluded that it responds well to the request, with the exception of some specific situations that were previously analyzed. To solve these situations and, consequently, obtain results more similar to reality, a more robust database will be required, with all the variables of interest filled, especially those that are crucial to decide between two therapies and that, for the most part, were not considered due to lack of data. The fact that there is a large difference in the proportion of cases that have undergone some therapies compared to others is also a factor that influences the suggestion given by the model.

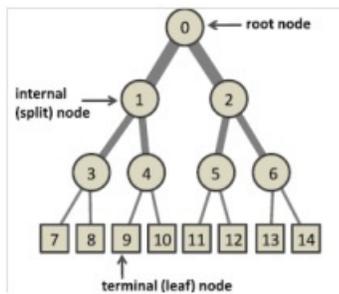


Figure 5: General structure of a decision tree.

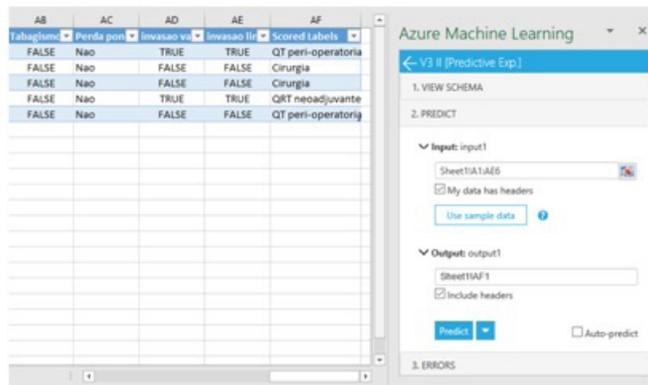


Figure 6: Application of the created template, in excel, using a web service.

### Decisão Terapêutica

Sexo  Masculino  Feminino

QT prévia  True  False

RT prévia  True  False

Cirurgia prévia  True  False

Disfagia

Anemia / HDA

Obstrução

Barrett associado  True  False

Estadiamento BF  True  False

Estadiamento ORL  True  False

Estadiamento Lap est  True  False

cT

cN

cM

Patologia

Órgão

Localização

Laparoscopia  True  False

Toracoscopia  True  False

Totalmente mini-invasiva  True  False

Doença pulmonar  True  False

DHC  True  False

Diabetes  True  False

Insuficiência renal  True  False

Infeção HIV  True  False

Patologia cardíaca  True  False

Etilismo  True  False

Tabagismo  True  False

Perda ponderal  True  False

Invasão vascular  True  False

Invasão linfática  True  False

[Submeter](#)

A sugestão de terapia aparecerá aqui, após clicar em "submeter".

Figure 7: Form obtained by accessing <https://decisaoterapeutica.azurewebsites.net/>.

**Table 15:** Comparison between the suggestion obtained for the "new cases" and the suggestion expected by the protocol.

Case	Oesophagus		Protocol	Stomach	
	Suggestion			Suggestion	Protocol
	CPC	ADC			
1	Surgery	Surgery	Endoscopic resection or surgery	Surgery	Endoscopic resection or surgery
2	Surgery	Surgery	Endoscopic resection or surgery	Surgery	Endoscopic Resection or Surgery
3	Surgery	Surgery	Surgery	Surgery	Surgery
4	Surgery	Surgery	Surgery	Surgery	Surgery
5	Neoadjuvant QRT	Neoadjuvant QRT	Neoadjuvant QRT or definitive QRT	Surgery	QT perioperative
6	Surgery	Surgery	Neoadjuvant QRT or definitive QRT	Surgery	QT perioperative
7	Neoadjuvant QRT	Neoadjuvant QRT	Neoadjuvant QRT or definitive QRT	QT perioperative	QT perioperative
8	Surgery	Surgery	Neoadjuvant QRT or definitive QRT	Surgery	QT perioperative
9	Neoadjuvant QRT	Neoadjuvant QRT	Neoadjuvant QRT or definitive QRT	Surgery	QT perioperative
10	Surgery	Surgery	Neoadjuvant QRT or definitive QRT	Surgery	QT perioperative
11	Neoadjuvant QRT	Neoadjuvant QRT	Neoadjuvant QRT or definitive QRT	QT perioperative	QT perioperative
12	Neoadjuvant QRT	Neoadjuvant QRT	Neoadjuvant QRT or definitive QRT	QT perioperative	QT perioperative
13	Neoadjuvant QRT	Neoadjuvant QRT	Neoadjuvant QRT or definitive QRT	QT perioperative	QT perioperative
14	Neoadjuvant QRT	Neoadjuvant QRT	Neoadjuvant QRT or definitive QRT	Surgery	QT perioperative
15	Palliative QT	Palliative QT	Palliative QT	Palliative QT	Palliative QT
16	Neoadjuvant QRT	Neoadjuvant QRT	Palliative QT	Palliative QT	Palliative QT
17	Surgery	Surgery	Palliative QT	Surgery	Palliative QT

## 8. Conclusions and Limitations of the Study

Among the solutions that information technology offers are clinical decision support systems, which offer several advantages, such as better clinical management, cost containment, diagnostic support and therapeutic support. The latter was the focus of this study, and a system of support for therapy was studied.

For the development of the model, a database provided by the multidisciplinary group of esophagus and stomach cancer (GMCEE) of the Lisbon IPO was used, from which three subsets were created, excluding variables without relevance to the objective and varying in the number of records included and information present, in order to explore which, one is most suitable for integration system. Different classification algorithms were tested using the Microsoft Azure Machine Learning Studio platform, Multiclass Neural Network, and Multiclass Decision Forest.

Comparing the results obtained for the two classifiers, varying their settings and the subset used, it was possible to conclude about the performance of the classifier, but also how the data used influence it. Although both showed excellent results, with accuracy values obtained between 91.67% and 100%, in general, the Multiclass Neural Network was superior.

Regarding the subsets used, it was noticed that, although the greater the number of records available, the better the classifier performance, the content of the same is a crucial factor. When using a set of data in which some cells were filled based on others, the classifier showed worse performance than when a smaller number of records were used, but with more variables filled by GMCEE physicians. Thus, for the creation of a decision support system, it is important to exist a large data set, but also complete and fidedigno the. In addition, during the tests, it was possible to notice that, for the training phase, a higher proportion of data should be used in the training set and lower in the test set, and the ideal proportion for the subsets used was 70%-30%. It is also important to highlight the importance of the existence of, approximately, the same proportion of cases of each type for model training. This is because, as observed in the classification and application tests of the system, if a large amount of data is of one type and there is a small amount of another type, with some characteristics similar to the larger group, there is a high probability that the records of the second group will be classified as those of the first.

It can be concluded that the model created presents a good performance, however, in its final application for new cases, there was a decrease in performance. In this way, there is room for future development and improvement. For this, the first and most important step is to fill in all variables that have weight in the therapeutic decision, since a limitation in the creation of this model was the fact that several important variables are eliminated, namely "EDA

12

- extension", "Base Table

Grade Differentiation", "Nº gg+", "Type Resseção (R)", "Margins", "Performance Status" and "ASA score", because the vast majority of the records did not contain information. Once all the important variables for the decision are filled, it will be possible to obtain a more robust database and, consequently, more reasoned therapeutic suggestions. It may also be an important point to explore the removal of variables that may be considered without effectively presenting great weight in the decision, such as the variables related to pathologies. This way, you can optimize the system, making it simpler but more specific.

## References

1. Who. Who report on cancer: setting priorities, investing wisely and providing care for all. 2020.
2. Hayes T, Smyth E, Riddell A, Allum W. Staging in Esophageal and Gastric Cancers. Vol. 31, Hematology/Oncol Clin of North Am. W.B. Saunders; 2017; 31(3): 427-440.
3. National Cancer Institute. Esophageal Cancer, Cancer Stat Facts. 2020
4. National Cancer Institute. Stomach Cancer. Cancer Stat Facts. 2020.
5. Cancer Research UK. Oesophageal cancer survival statistics. 2020.
6. Cancer Research UK. Stomach cancer survival statistics. 2020.
7. Eta S. Berner, editor. Clinical Decision Support Systems: Theory and Practice. 3rd editio. Health Informatics. Springer; 2015.
8. Middleton B, Sittig DF, Wright A. Clinical Decision Support: a 25 Year Retrospective and a 25 Year Vision. Yearb Med Inform. 2016 Aug 2 Suppl 1(Suppl): S103-16.
9. Yes I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical Decision Support Systems for the Practice of Evidence-based Medicine. J Am Med Informatics Assoc. Dec 2001; 8(6): 527-34.
10. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success NPJ Digit Med. 2020; 3: 17.
11. Amin SU, Agarwal K, Beg R. Data Mining in Clinical Decision Support Systems for Diagnosis, Prediction and Treatment of Heart Disease. Int J Adv Res Comput Eng Technol. 2013; 2(1): 218-23.
12. Kohn LT, Corrigan JM, Donaldson MS, editors. To Err is Human: Building a Safer Health System. Washington (DC); 2000.
13. Castile J, Leitão C, Sousa P, Guerreiro I, Ferro SM de, Serrano M, et al. Comparison of the clinicopathological characteristics and the survival outcomes between the Siewert type II/III adenocarcinomas. Rev Port Cir. 2016;18.
14. Ranawana R, Palade V. Multi-Classifer Systems: Review and a roadmap for developers. Int J Hybrid Intell Syst. 2016 Apr 20; 3(1): 35-61.
15. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. Data Mining: Concepts and Techniques. Elsevier Inc.; 2012.
16. Heaton J. Artificial Intelligence for Humans: Deep learning and neural networks. Heaton Research, Incorporated.; Artificial Intelligence for Humans (2015).
17. Kuncheva L. Combining Pattern Classifiers: Methods and Algorithms. 2nd editio. Combining Pattern Classifiers. 2014.
18. Beyerer J, Richter M, Nagel M. Pattern Recognition: Introduction, Features, Classifiers and Principles. From Gruyter Oldenbourg; 2017. (De Gruyter Textbook).
19. Microsoft, i'm sorry. Multiclass Neural Network. 2020.
20. Criminisi A, Shotton J, editors. Decision Forests for Computer Vision and Medical Image Analysis. Springer; 2013.
21. Microsoft, i'm sorry. Multiclass Decision Forest. 2020.
22. Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical Machine Learning Tools and Techniques. 4th editio. Cambridge: Elsevier; 2011.