UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE GRADUAÇÃO
CURSO DE CIÊNCIA DA COMPUTAÇÃO

ORLANDO VILMAR RODRIGUES HIDALGO JUNIOR

A MACHINE LEARNING MODEL FOR EARLY DIAGNOSIS OF ARTERIOVENOUS FISTULA STENOSIS

São Leopoldo
2020

ORLANDO VILMAR RODRIGUES HIDALGO JUNIOR

A MACHINE LEARNING MODE FOR EARLY DIAGNOSIS OF ARTERIOVENOUS FISTULA STENOSIS

Trabalho de Conclusão de Curso apresentado como requisito parcial para obtenção do título de Bacharel em Ciência da Computação, pelo Curso de Ciência da Computação da Universidade do Vale do Rio dos Sinos - UNISINOS

Orientador(a): Prof. Dr. Cristiano André da Costa

São Leopoldo

2020

A MACHINE LEARNING MODEL FOR EARLY DIAGNOSIS OF ARTERIOVENOUS FISTULA STENOSIS

Orlando Vilmar Rodrigues Hidalgo Junior*
Cristiano André da Costa**

Abstract: The quality of the vascular access of patients with Chronic Kidney Disease is extremely important and proves to be a decisive factor in the patient's longevity and well-being. Currently, arteriovenous fistula is one of the most recommended vascular access and some concerns about this access are evident, such as arteriovenous fistula stenosis. The aim of this work is to develop a machine learning model for analysis and prediction based on monitoring of the data generated by the hemodialysis equipment and hemodialysis session. This study is a partnership between Clinical Research Center located in Porto Alegre, Brazil and the Graduate Programs in Applied Computing and Nursing at UNISINOS. The project was previously approved by the Research Ethics Committee of UNISINOS and HCPA and uses 1483 samples from 27 patients. Logistic Regression, K-Nearest Neighbors, Support Vector Machine and Random Forest have been trained and tested using 10-fold cross validation. Random Forest achieved the best performance with an F1-score of 98.40%, sensitivity of 98.80% and specificity of 98.50%. We also found that patient's age, fistula age and gender had higher importance for Random Forest in Predicting stenosis. This model used a new set of features and had higher results compared to the related works, making it a promising predictor of arteriovenous fistula stenosis.

Keywords: *Arteriovenous fistula (AVF), stenosis, hemodialysis (HD), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR).*

1 INTRODUCTION

Chronic kidney disease represents a gradual loss of kidney functions. In these cases, hemodialysis is presented as the immediate treatment commonly used.

* Computer Science Undergraduate. Email: orlandohj@unisinos.br
** Tutor, Unisinos Professor, Ph. D. in Computer Science from Universidade Federal do Rio Grande do Sul (2008), M. Sc. in Computer Science from Universidade Federal do Rio Grande do Sul (1997). Email: cac@unisinos.br

Dialysis performs critical functions that the kidneys are unable to perform, such as removing toxic substances from the metabolism and balancing the body's fluid levels, purifying the patient's blood using the hemodialysis device (BELMONT et al, 2019).

As the volume of blood flow increases, the team of surgeons creates an arteriovenous fistula (AVF), a vascular access between an artery and a vein. Thus, the vein dilates with the pressure between the artery and the vein, allowing faster blood flow, which makes the AVF the best access for hemodialysis (YEIH et al., 2014). As time goes by, it is possible for the patient to present a reduction in the functioning of the vein with the narrowing of the blood passage, also known as arteriovenous stenosis, a complication that can cause fistula failure, making it necessary to develop a new fistula, exposing the patient to invasive catheter treatment (BHATIA et al, 2018).

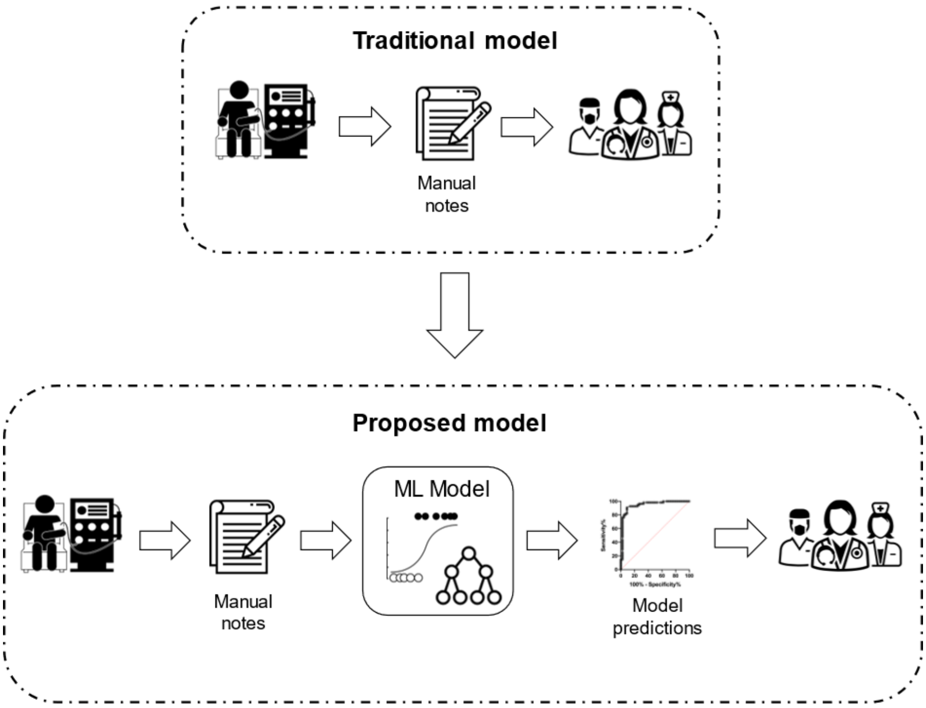
The constant monitoring and the use of a computer algorithm capable of producing reliable and accurate results will impact on the reduction of vascular complications, increasing the patient's well-being and reducing treatment costs since early diagnosis avoids the need for new fistulas and surgical procedures (BELMONT et al., 2019). Techniques such as ultrasound, angiography and collection of sound waves from the fistula are some of the most used forms of information in search of effective characteristics in the prediction of complications of fistula and stenosis. Some of these processes are not applied with great frequency because they are invasive, have side effects and high cost. These obstacles often result in a small sample of information, making the classification of the algorithms less accurate, mainly due to the variation of information of each patient (CHIANG et al., 2019).

The dialysis process tends to be continuous and with the application of several sessions within a short period. Each session generates a varied amount of information, most of which are not used as these are only available during the procedure. For reasons of access to equipment data, the collection of this data is done manually by the nephrologist present in the procedure (BHATIA et al, 2018).

The motivation of this work is to enable the early diagnosis of arteriovenous fistula stenosis and its aggravations, making use of the largest amount of information available in dialysis sessions, observed and collected by specialists, taking advantage of the large amount of data that can be lost, seeking new characteristics, as well as the best way to take advantage of them through machine learning models.

Figure 1 illustrates the traditional model for detecting arteriovenous fistula stenosis in patients versus the model proposed in this work. In the conventional model, the specialist’s manual records are analyzed by the team based on known information. On the other hand, the proposed model preprocesses these records and uses them as input to the machine learning algorithms.

Figure 1 – Traditional model vs. Proposed model.



Source: Elaborated by the author.

Possible discoveries in the approach of this model can increase the quality of life of the patient during the course of his treatment, helping specialists in the field to make decisions in advance and not reactive to complications, avoiding surgical procedures and even more severe problems in the patient.

The main scientific contribution of this work is to provide an effective model to predict arteriovenous fistula stenosis, as well as to identify the most significant features using data available in the hemodialysis machine and dialysis session. For that, we will use artificial intelligence-based solutions that are trained and tested with stenosis dataset to them compare the results obtained by the models, according to the accuracy and other metrics applicable to the problem, providing valuable information for early prediction of stenosis.

This article is structured in six sections. The second section presents the theoretical foundations necessary to understand the concepts presented in this work.

The third section describes the related works and opportunities and research open points. The fourth section describes the proposed solution and development methods, detailing the preprocessing and modeling steps. The fifth section discusses the results achieved. Finally, the sixth section presents the conclusion, final considerations and possible future research.

2 THEORETICAL FOUNDATION

This section presents the relevant concepts about chronic kidney disease, hemodialysis, vascular access, arteriovenous fistula, possible access's complications, as well as forms of monitoring used in the diagnosis of stenosis.

2. 1 Chronic Kidney Disease

Chronic kidney disease (CKD) is a serious condition in which there is a progressive and irreversible decrease in kidney function, perceived by the reduction in the glomerular filtration rate, when kidney function is affected and is no longer able to sustain life in the long term. There are two main modalities of treatment: renal replacement therapy, which includes peritoneal dialysis, hemodialysis and kidney transplantation, or conservative care, also called palliative or non-dialysis care that consists of non-invasive treatments whose objectives are to mitigate the symptoms of disease. Currently the main form of treatment in more advanced cases of the disease is hemodialysis (WEBSTER et al., 2017).

Chronic kidney disease has a great impact on the patient's quality of life, who must undergo the hemodialysis process to reverse the hydro electrolytic and acid-base imbalance caused by it. Data suggest that more than 750 million people worldwide have CKD. The economic impact on health systems is worrying in view of a high population of elderly with comorbidities. (CREWS; BELLO; SAADI, 2019)

2.2 Hemodialysis

Hemodialysis is a form of treatment that is part of the modality of renal replacement therapy for chronic kidney disease. It is a process in which it aims to eliminate harmful residues from the body, assist in the control of blood pressure and maintain the balance of substances such as sodium, potassium, urea and creatinine. Hemodialysis performs functions that the kidneys are unable to perform, remove toxic substances from metabolism and balance the body's fluid levels (SBN, 2016).

The process consists of cleaning the blood by filtering it through selective diffusion from an extracorporeal circulation system and membrane extracorporeal. The time of hemodialysis sessions varies according to the patient's clinical status, but on average they last four hours and occur three times a week (SBN, 2016).

The treatment has a major impact on the patient's life, as it affects diet, physical and social condition. Despite changing many aspects and causing several changes in the patient's life, it should not be a negative factor, since it brings an improvement and well-being to the condition of the chronic kidney patient, contributing positively (GRASSELLI et al., 2012).

2.3 Permanent Vascular Access

Establishing and maintaining adequate vascular access is essential for chronic renal patients on hemodialysis, both for an appropriate dialysis dose and for satisfactory therapy. The ideal access for hemodialysis should have a long lifespan, provide an adequate blood flow rate and present no complications. Currently, the arteriovenous fistula (AVF) is the one that best meets these requirements and the one that requires the least number of interventions (HAYASHI; HUANG; NISSENSON, 2006).

The vascular accesses are subject to the development of stenosis. Avoiding this complication is a challenge, as this problem contributes to the reduction of the access life and to an inadequate blood flow. Physical examinations and instruments to evaluate the blood flow are some of the measures taken to monitor vascular access, which must be constant to guarantee its perfect performance. Early diagnosis of the dysfunction is important so that it can be repaired before it can result in inefficient hemodialysis therapy or major complications for the chronic kidney patient (ABREO, K.; AMIN, ABREO, A., 2019).

2.4 Arteriovenous Fistula

The arteriovenous fistula (AVF) is the access for hemodialysis that allows a better life span and a lower number of complications, it is usually done on the upper limbs, preferably on the non-dominant forearm. They are built from an anastomosis between an artery and a superficial vein through a small surgical intervention, the most common vessels are the cephalic or basilic vein and radial or brachial artery. There is a process of arterialization of the veins where they become stronger and more resistant, this process can take some time, resulting in the complete maturation of the fistula. For this reason, the early use of the fistula can lead to a failure in access and its loss, because the veins are not yet prepared to be punctured. Therefore, an AVF is considered mature when it has an adequate vein diameter, which allows for successful cannulation and can provide the appropriate blood flow for dialysis. Although the maturation time is not yet known, it is estimated to be more than two or three months to obtain an AVF with less risk of failure (HAYASHI; HUANG; NISSENSON, 2006).

Even though AVF is the most recommended type of vascular access, strict supervision must always be maintained both for monitoring, in which it refers to the examination and evaluation of vascular access through physical examination, and for the surveillance that refers to the periodic evaluation of vascular access using tests that may involve special instrumentation, being the two forms of supervision of the AVF complementary. This control is very important for the hemodialysis therapy of the chronic renal patient to occur without major complications (COENTRÃO; RODRIGUES, 2013).

2.5 Complications of Permanent Vascular Access

Stenosis is a narrowing of the vein or artery that can interfere with the blood flow of the AVF, usually causing a decrease or obstruction of blood flow. Normally, patients with vascular access should not complain of pain or weakness and this information can be useful for the diagnosis of stenosis, which can be done through physical examination, which includes inspection (appearance), palpation (sensation) and auscultation (sound). The diagnosis can become difficult in cases where the lesion is very small or the stenosis is very prominent, which can lead to thrombosis, with venous stenosis being responsible for 80-85% of the thrombosis of the access and arterial stenosis is responsible for 1-2% of access thrombosis. Therefore, the findings on physical exams always need to be validated through Doppler ultrasound (HAYASHI; HUANG; NISSENSON, 2006).

Early detection of stenosis and low AVF blood flow are essential in preventing thrombosis and access failures, leading to loss of access. When the diagnosis is late and thrombosis is already present, it is characterized by an undetectable flow by physical examination. Aneurysm formation may be indicative of skin thinning by repeated needles in the same location or a high intra-access blood flow (COENTRÃO; RODRIGUES, 2013).

2.6 Vascular Access Monitoring and Surveillance

Monitoring and surveillance in vascular access are related to control and precautionary measures to guarantee the adequate function of the access, allowing a longer useful life and preventing its loss from occurring. Monitoring refers to physical examination while surveillance is through tests or special instrumentation, the two forms of which are based on the identification of patients at risk of developing a future dysfunction in access and allied with an early intervention, reducing the incidence of failure in vascular access, ensuring satisfactory hemodialysis therapy for chronic renal patients (HADDAD et al, 2012).

These methods are applied by the dialysis team and a combination of several techniques is usually used to detect patients at risk of access dysfunction or the dysfunction itself. As there are still no studies to prove which is the best method of controlling vascular access, monitoring and surveillance are still very important for the prevention of failures, as they are easy to learn, quick to perform and relatively economical (GRASSELLI et al., 2012).

3 RELATED WORKS

In order to obtain more details regarding the solutions proposed in relevant studies in the area, an exploratory research on prediction of stenosis and machine learning was carried out. The chosen databases were Scopus and Google Scholar, some of the largest bibliographic bases available, in addition to having an advanced search engine. The search criteria for the articles used was the combination of the words “machine learning", "arteriovenous fistula", "stenosis”, With a date equal to or greater than 2015, and articles ordered by relevance. From the search result, five articles were selected and described below.

Grochowina (2015) used the sound emitted by the radio cephalic fistula that is located on the wrist, with the collection point being five centimeters from the anastomosis. The purpose of the study is to compare the Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) algorithms, with the objective of identifying the stenosis as early as possible, evaluating the model that best fits the problem. The data sample was distributed in such a way that the rarest cases (stenosis), have a balanced occurrence with the cases in which there is no stenosis, as these are usually in greater quantity. The algorithms were trained with 60% of the data and tested with the remaining 40%, resulting in an accuracy of 81% by the SVM, while the KNN showed an accuracy of 85%.

Bhatia et al (2018) proposes the use of the characteristics presented in the hemodialysis session to enable the preventive creation of a new fistula. It points out the steps that a system needs to present such results, namely: importing data, graphical analysis of each characteristic, evaluation with different classification algorithms, and performance reports comparing the results obtained. Since only the idea of the model was presented, characteristics such as blood flow, venous pressure, dialysis flow and blood pressure, were analyzed graphically and considered important within the study sample for use in the future diagnostic process.

In the study carried out by Kordzadeh et al (2019), Artificial Neural Network (ANN) was used to predict fistula maturation in Radiocephalic arteriovenous fistula. The data was collected from 266 patients for four years. Some of the features used in the study were comorbidities, blood flow, bruit, thrill and pulse. Patients who did not achieve fistula maturation after 6 months were classified in the group of functional maturation failure. The dataset was split into 70% training set, 15% for F-measure evaluation and 15% for model testing. Age, vein and artery size were selected as the best subset of predictive characteristics, resulting in an accuracy of 89% for fistula maturation diagnosis.

Chiang et al (2019) developed a photoplethysmography sensor as a noninvasive, portable and low-cost alternative to ecodoppler ultrasound. The data consisted of 153 samples of soundwaves, having 74 samples for degree of stenosis evaluation (DOS) and 79 samples for blood flow volume (BFV). High-risk patients were labeled based on a degree of stenosis larger than 30% and blood flow volume greater than 600 ml/min. Naive Bayes, KNN and SVM were applied for assessing the DOS and BFV. The SVM algorithm proved to be the most effective classifier with a DOS accuracy of 87.84% and 88.61% accuracy for BFV.

Using data of approximately fourteen thousand patients from United States Renal Data System, Qian et al (2020) studied important predictor characteristics for AVF maturation focused on older patients, having all the patients aged 67 years or older. Arteriovenous fistula maturation has shown to be reduced in older patients compared to younger ones, resulting in the prolonged need of a central venous catheter, which is classified as an invasive monitoring technique. The study considered different types of features as predictors, such as: demographic (age, gender and race), geographical and laboratorial. The data was randomly split into 66% training data and 34% for testing and was classified with Random Survival Forest (RSF), which resulted in 45.9% Out-of-bag score with 34 features.

Lastly, Grochowina (2020) research aimed at developing a low-cost phono-angiography based device for non-invasive monitoring of AVF condition. The data was collected from 38 patients and 23 features were extracted from the resulting phono-angiogram, which were used in the pre-processing step. The samples were divided in 6 groups and labeled from A to F, having A as the best scenario for AVF condition and F as the worst condition present in the dataset. This classification was done through mathematical analysis and ecodoppler results evaluated by the team of specialists. Random Forest (RF), SVM and KNN were used for classification of the data, obtaining the best accuracy of 81% with KNN.

Table 1 shows a comparison between the articles. The items taken into consideration were the type of collected data labeled as sound or hemodialysis (HD), the number of patients that were part of the study, the algorithm that showed the best result in arteriovenous fistula classification problem as well as its predictive accuracy.

It is possible to notice that the approach of collecting sound waves emitted by the fistula has shown to be one of the most common options for non-invasive monitoring of the arteriovenous fistula. However, it presents some problems such as the variation of sounds for each patient, in addition to the difficulty of obtaining diversified samples and the complex classification. Since studies focus on this type of data, there are other viable forms of monitoring and collection that are generally not used. Through this exploratory research we identified the opportunity to make use of the data presented in the hemodialysis machine that are often discarded after the session. This way we can discover new characteristics through analysis and modeling using such data as well as the most diversified hemodialysis session information possible, increasingly optimizing the work performed by specialists in the field, serving as an auxiliary method in the early prediction of complications of AVF. In addition, the present work has the opportunity to explore different metrics for analyzing the results obtained in the process.

Table 1. Related works comparison. (ANN, Artificial Neural Network; HD, hemodialysis; KNN, K-Nearest Neighbor; RSF, Random Survival Forest; SVM, Support Vector Machine)

Article	Grochowina (2015)	Bhatia (2018)	Kordzadeh (2019)	Chiang (2019)	Qian (2020)	Grochowina (2020)
Data	Sound	HD	HD	Sound	HD	Sound
No. of patients	9	200	266	-	14892	38
Algorithm	KNN	-	ANN	SVM	RSF	KNN
Accuracy	85%	-	89%	87%	-	81%

Source: Elaborated by the author.

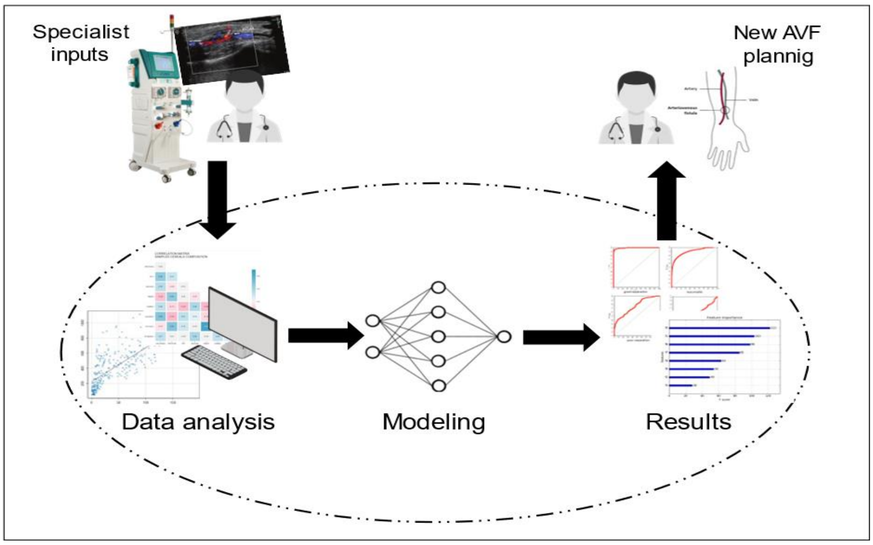
4 MATERIALS AND METHODS

This section aims to describe the model general steps, the dataset used in the study, develop the preprocessing stages from the exploratory analysis to the feature selection that will compose the final dataset. Based on these data, several algorithms will be tested, evaluated with relevant metrics to choose the model that best fits the stenosis prediction problem.

4.1 Process Overview

The general model (Figure 2) consists of an exploratory analysis of the data collected by observing the hemodialysis machine, physical examination and doppler ultrasound results, which are preprocessed, modeled and outputted back to the specialist to help on planning a new arteriovenous fistula based on the model results.

Figure 2. General process for planning a new AVF based on model predictions.



Source: Elaborated by the author.

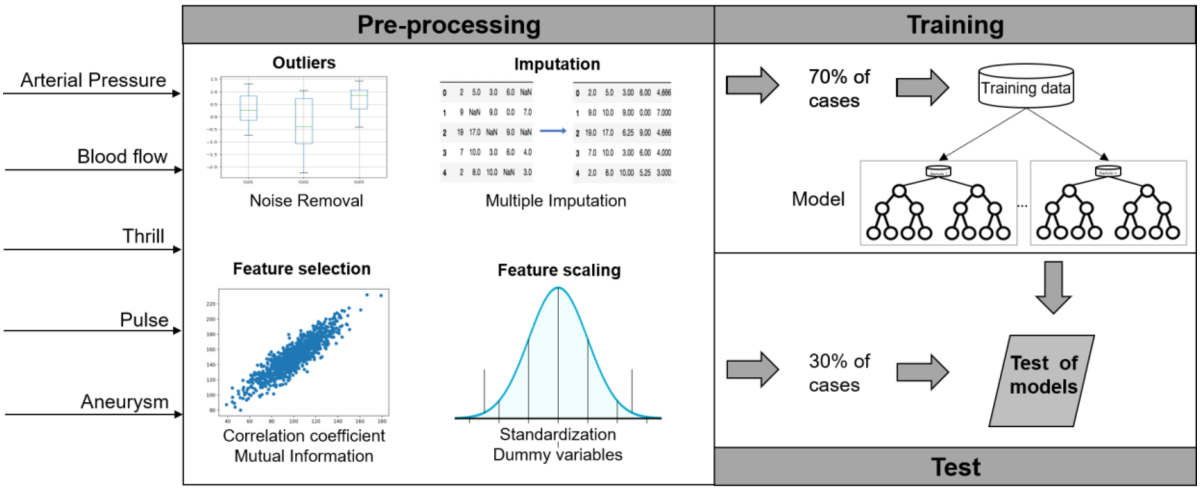
The detailed model (Figure 3) contains three main steps, starting with the pre-processing of the input variables. Such data compose information from the real world, that is, they may have several errors and formatting problems since their collection is done manually by the specialist.

One of the pre-processing steps is dealing with the outliers, which consist of non-standard information presented by the sample. The identification of these values leads to a better understanding of the data (DENESHKUMAR; MANIKANDAN; KALIYAPERUMAL, 2014). Another step is identifying null or invalid values. The validation of these data eliminates the possibility of erroneous trends, since minority data can lead the algorithms to mistaken conclusions. Several imputation techniques can be applied, but it is necessary to pay attention to the nature of the data and understand the best way to apply them (VERGARA; ESTÉVEZ, 2015).

The feature selection uses statistical tests to check the features correlations. Redundant data can be found and eliminated, making the modeling and training faster. In the feature scaling step, the previously selected features, undergo a scale adjustment for their respective types. For categorical data, techniques such as label encoding and conversion to dummy variables may be applied. These conversions are important because some algorithms only work with continuous data. For the numerical features the standardization technique is applied, leaving all values on the same scale.

The pre-processed inputs are then split into training and test set. The objective is to train several algorithms in order to identify the model that best fits the data. Training set is then used to validate the models with unseen records. The result is defined by the model that has the best overall performance.

Figure 3. Detailed model for pre-processing, training and testing stenosis inputs.



Source: Elaborated by the author.

4.2 Dataset

This study uses stenosis dataset obtained through a partnership between Clinical Research Center located in Porto Alegre, Brazil and the Graduate Programs in Applied Computing and Nursing at UNISINOS. The project was previously approved by the Research Ethics Committee of UNISINOS and HCPA (protocol number – CAEE: 19551019.8.3001.5327).

The data sample consists of 1483 records of hemodialysis sessions of 27 patients. The collection cycle was approximately of three months, based on a 50% decrease in the difference in diameter and an increase of 100% in the venous systolic peak between two points of the fistula, obtained at the end of the treatment through ecodoppler tests, having these results as the gold standard in the identification of stenosis by specialists (HAYASHI; HUANG; NISSENSON, 2006).

The data used in this work characterize a binary classification problem, with the output of stenosis as YES or NO. Table 2 describes the data as well as their type and values. Altogether there are 21 characteristics, 6 of the numerical type and 15 of the categorical type, including the class stenosis. As shown in Figure 4, the dataset has 1096 samples marked as no stenosis, while 387 indicate stenosis. This class distribution represents an imbalanced distribution of 1 positive case (stenosis) for every 3 negative ones (healthy patients).

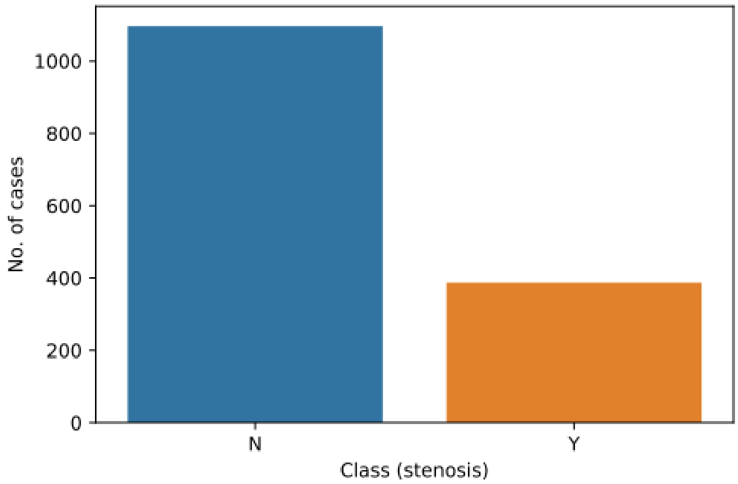
Table 2. Stenosis dataset described by feature name, description, type in unit or category, missing values percentage of each feature.

Name	Description	Type: unit/category	Missing percentage
Age	Patient’s age	Numerical: years	0%
Comorbidities	Does the patient has comorbidities or not	Categorical: yes, no	0%
Fistula age	Patient’s fistula age	Numerical: years	0%
Race	Patient’s race	Categorical: white, black, brown	0%
Sex	Patient’s gender	Categorical: yes, no	0%
Arterial pressure	Arterial pressure	Numerical: mm/Hg	2%
Flow	AVF’s blood flow	Numerical: ml/min	2%
Kt/v online	Dialysis dose per session	Numerical: mmHg	38%
Venous pressure	Venous pressure	Numerical: mm/Hg	2%
Aneurysm	Does the patient’s AVF has aneurysm or not	Categorical: yes, no	0%
Bruise	Presence of bruise in the AVF path or not	Categorical: yes, no	0.50%
Clots	Presence of clots on the puncture needle or not	Categorical: yes, no	1%
Collateral vein	Presence of thrill / pulse in collateral venous network or not	Categorical: yes, no	3%
Edema	Does the patient’s AVF limb has edema or not	Categorical: yes, no	4%
Hemostasis	Hemostasis time changed (>5 min) when removing the needles or not	Categorical: yes, no	0.50%
Hypoperfusion	Does the patient has hypoperfusion or not	Categorical: yes, no	37%
Pain	Does the patient’s AVF has pain or not	Categorical: yes, no	0%
Pulse	Change of pulse in the AVF or not	Categorical: yes, no	40%
Punctures	Are the difficulties in puncture or not	Categorical: yes, no	0%
Thrill	Change of thrill in the AVF or not	Categorical: yes, no	40%
Stenosis (class)	Does the patient has stenosis or not	Categorical: yes, no	0%

Source: Elaborated by author.

In medical datasets, high risk patients (abnormal cases) tend to represent the minority class as normal (healthy) cases compose most of the records. This may lead to a learning bias towards the majority class resulting in poor performance, specifically for the minority class that is the target for an accurate classification. In order not to lose important information, this study uses balancing parameters whenever possible as seen in the following sections.

Figure 4. Distribution of class in stenosis dataset.



Source: Elaborated by the author.

4.3 Data Preprocessing

Real world datasets are very likely to contain missing, extreme, redundant and inconsistent data, causing misinterpretation of the data as well as inaccurate model results. This study's dataset is no exception. Therefore, the first step to build a useful machine learning model is to explore and treat the data, making use of data processing techniques.

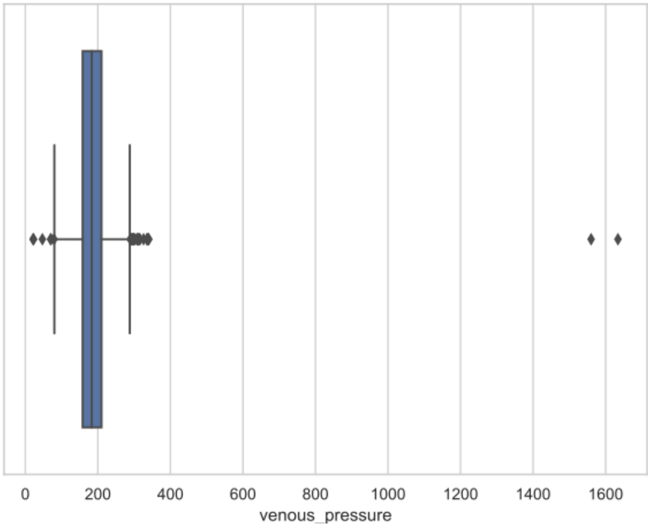
4.3.1. Outliers

Extreme values or outliers are extreme data points that deviate from the central tendency. These outliers are generally referred to as a noise in the data or exceptions, meaning that they may be discarded. However, medical data demands attention since outliers can be legitimate and have influence on the results of stenosis detection (DENESHKUMAR; MANIKANDAN; KALIYAPERUMAL, 2014).

In stenosis dataset, these extreme values may be originated from observation errors since the values are manually noted by the specialist. For this reason, each outlier detected in the dataset is checked to know whether they can be treated as noise or realistic data that can present valuable information.

In this study, the extreme data points that go beyond the acceptable range clinically have been handled as missing data and then modified as described in later sections.

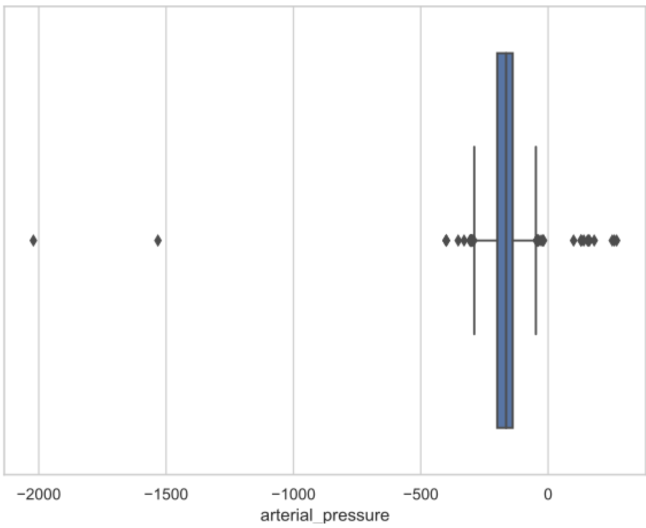
Figure 5. Boxplot for venous pressure.



Source: Elaborated by the author.

Box plots have been used to detect stenosis, as seen in Figure 5, some records presented a venous pressure of 1500 and 1600. High venous pressure is commonly used as AVF surveillance parameter. However, there are no registries of such high values meaning that these outliers are not legitimate (WHITTIER, 2009).

Figure 6. Boxplot for arterial pressure.



Source: Elaborated by the author

Figure 6 shows extreme values of -2000 and -1500mm/Hg for arterial pressure. Recent studies evaluate the importance of this parameter in the AVF, but there are no mentions of values this low, so these values cannot be accepted as normal information (SCHOLZ et al, 2019). A common explanation for these values is a simple typo, as the data is observed and noted manually.

4.3.2 Null Values

The presence of null or invalid values is a very common problem in datasets, usually resolved by eliminating them, as well as filling the data with the mean, mode or median of the attribute values. Table 2 shows that approximately 57% of the stenosis dataset has null variables. The missing percentage range from 0.5% to 40% according to each variable. Special care is needed to replace these values, because incorrectly imputed clinical data can generate erroneous trends and patterns.

In this study, Multiple Imputation (MI) was applied as imputation technique. In MI, the null data is replaced m times, where m is usually a small number from 3 to 10. Each of the simulated datasets are analyzed and adjusted according to trends observed by the algorithm used as estimator (JONATHAN et al, 2009). In the case of numerical features Linear Regression was chosen. As for categorical variables, Logistic Regression was used.

4.3.3 Feature Selection

This stage employs statistical tests as well as mutual information concepts to discover relationships between the input parameters to gain knowledge of the data, examine their relevance to the early diagnosis of stenosis, as well as association between themselves. The stronger the correlation between variables the closer they are to be linearly correlated, meaning that one of them can be removed as a redundant information. However, strong association between a parameter and the class variable means that this parameter may be relevant to the optimal prediction of stenosis. Conversely, weak correlations generally mean that data can be discarded as irrelevant parameter. This is an important step to understand the level of overlap between healthy patients and stenosis, building an effective model that can predict valid results.

From the result of Pearson's correlation (Figure 5), it is observed that blood flow has a moderate positive correlation of 0.41 with venous pressure (VP). Venous pressure consists of the positive pressure exerted to return the blood to the AVF after being filtered. Elevations in VP may suggest complications of the fistula, however there is no evidence that this measure is effective in monitoring the arteriovenous fistula (WHITTIER, 2009). With a 0.04 correlation between VP and the stenosis class, this parameter was considered not relevant.

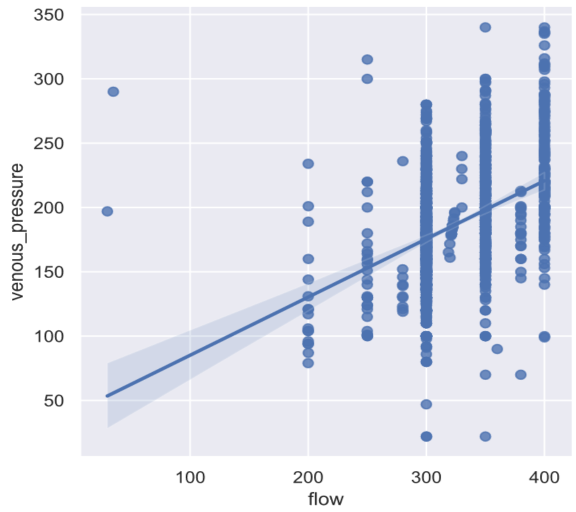
Although there is no consensus among the studies, the blood flow achieved in the hemodialysis equipment with values lower than 300ml / min proved to be an indicator of dysfunctions in the AVF (POLKINGHORNE, 2006). However, only 2% of patients diagnosed with stenosis had values in this range (Figure 6), in addition to having a 0.05 correlation with the class of stenosis, proving to be a low impact parameter.

The Cramer V correlation test showed a strong positive correlation of 0.85 between thrill and pulse. These characteristics are commonly used together in the physical evaluation of the fistula, indicating the vibration caused by pulsatile arterial blood inside the vein. The thrill is usually continuous and can be felt, while the pulse must be smooth. Reduction or absence of these parameters indicates difficulty in maturing the fistula or stenosis. Studies report the importance of observing the thrill for the early detection of stenosis and the importance of observing this parameter at each hemodialysis session (POLIMANTI, 2018). In the dataset of this study, 60% of the patients with stenosis had a change in thrill, pointing it as a possible parameter for identifying stenosis.

Pain at the site where the AVF is present and aneurysm have a significant correlation of 0.71 according to the Cramer V test. Aneurysm represents areas of enlarged AVF with a larger diameter and swirling blood flow. Pain in AVF is commonly related to aneurysm, being reported by 48% of patients and recognized as a parameter for surgical intervention (PASKLISNKY et al, 2011).

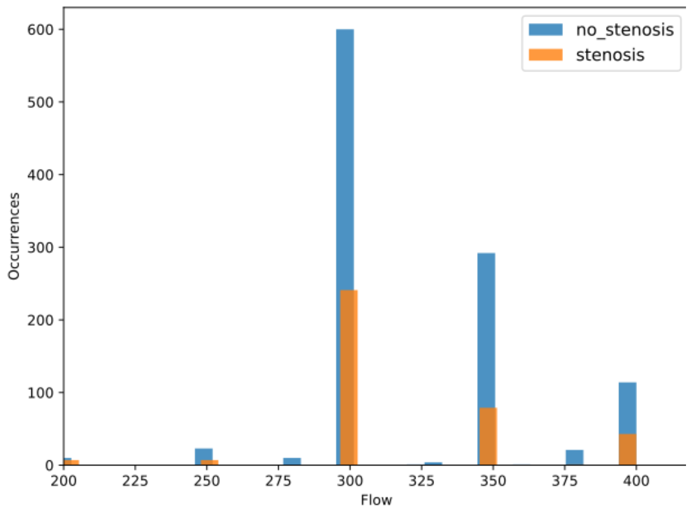
Reports of pain in the AVF also showed a correlation with difficulty in puncture, this being 0.58 according to Cramer V. The performance of the puncture is essential to obtain a good venous access, guaranteeing a quality dialysis (DIAS; NETO; COSTA, 2008). The puncture process has a repetitive nature and can be somewhat painful, due to the size of the needle, chosen technique and difficulties in the process, generating reports of pain in the AVF in up to 60% of patients (KORTOBI et al, 2019). The parameter of pain in the AVF was identified as redundant, showing an association with difficulty in puncture and aneurysm, confirmed by the chi-square test with p-value <0.001.

Figure 5. Scatterplot between flow and venous pressure.



Source: Elaborated by the author.

Figure 6. Blood flow distribution grouped by class category.



Source: Elaborated by the author.

Puncture and aneurysm difficulties are also correlated with a coefficient of 0.61. Puncture problems may be related to stenosis and should be checked at each hemodialysis session. In addition, repeated puncture can cause weakening of the vein walls, causing an aneurysm that limits the possibility of punctures depending on its size (KORTOBI et al, 2019). Approximately 3% of the records with a positive diagnosis in the dataset had difficulty in puncture. The chi-square test showed a p-value <0.008 between difficulties in puncture and aneurysm.

Aneurysm is present in 70% of the records with a positive diagnosis for stenosis, although this high number may originate from false aneurysms or pseudoaneurysms, studies show equivalent numbers among patients with stenosis and aneurysm (PATEL et al, 2014). In this study, chi-square showed a correlation between aneurysm and the stenosis class with p-value < 0.001. Thus, difficulties in puncture were eliminated as a redundant feature.

Lastly, Mutual Information was used as a filter method. Mutual Information measures reduction in uncertainty between random variables, including non-linear relationships, and is also independent to the learning algorithm (VERGARA; ESTÉVEZ, 2015).

At the end of the tests and analysis of the results, the following attributes were discarded from the experiment: Blood flow, venous pressure, edema, pulse, bruise, pain, hypoperfusion, clots, punctures, comorbidities and hemostasis.

4.3.4 Feature Scaling

It is common for the dataset to have values on different scales, such as centimeter and milligrams. The data sample used in this work is no exception, however, some techniques allow the standardization of these values. For the numerical attributes, the standardization technique was used, which consists of distributing the values so that the mean is zero and the standard deviation is 1, centralizing the data on the same scale (BOLLEGALA, 2017). The categorical attributes were adjusted using the dummy variables and label encoding techniques, which transform the values to 0 and 1. The race column was broken into three, representing the race variations of the dataset (white, black, brown) in the described range. These conversions help the application of some algorithms that cannot handle categorical variables, especially regression problems.

4.4 Modeling

Five algorithms were used in this step, namely: Random Forest, K Nearest Neighbor (KNN), Support Vector Machine (SVM) and Logistic Regression. These classifiers were chosen based on the studies carried out in related works, which proved that these models can be very accurate tools for predicting complication in arteriovenous fistula, especially stenosis. In addition, these algorithms can benefit the classification of unbalanced data through the implementation of class weighting.

4.4.1 Logistic Regression

Logistic Regression (LR) predicts whether something is True or False (binary classification). Instead of fitting a line to the data, logistic regression fits it to a sigmoid (logistic function). The curve goes from 0 to 1, meaning the probability that a patient has the target variable based on the inputs. Although it tells the probability of the class, it’s used for classification (HYEOUN-AE, 2013). For example, if the patient has a high probability of having stenosis, then we will classify him as “stenosis”, otherwise “no stenosis”. Logistic regression can work with continuous and discrete data and its ability to provide probabilities and classify new samples using continuous and discrete measurements makes it a popular machine learning method and a base algorithm for classification problems.

4.4.2 K-Nearest Neighbor

In K-Nearest Neighbor (KNN) algorithm, learning is based on how similar one data (vector) is to another and the classification of this vector occurs according to the calculation of its distance in relation to other data. This means that a new input will be classified according to the shortest distance from its neighbor. The calculation of this distance is commonly performed by the Euclidean distance, but there are other metrics that can be used accordingly to the parameters. Despite its high computational resources demand, KNN’s flexibility on working with categorical and numerical data makes it popular for feature selection and classification of stenosis (GROCHOWINA; LENIOWSKA, 2015).

4.4.3 Support Vector Machine

Support Vector Machine (SVM) is a learning model that is commonly used in classification problems. It is reported to reveal promising results in the diagnosis of arteriovenous fistula condition (GROCHOWINA; LENIOSWKA; BŁĄDZIŃSKA, 2020). The idea of SVM is to create an optimal hyperplane and divide the data in two dimensions, separating the points with their respective classes. This hyperplane is established by points near the edge of the plane, called support vectors, which are all the same distance from the edge. New data will be categorized according to the dimensions already classified and from the data already available (training) that are mapped by mathematical functions known as Kernels, varying from radial basis function (RBF), linear, sigmoid, to polynomial (CHIANG et al., 2019). In this study, RBF has been chosen based on parameter tuning results.

4.4.4 Random Forest

Random Forest (RF) is a type of decision tree, widely used and that provides good results. The purpose of decision trees is to divide data into subsets following a series of rules and with a goal. This process is usually achieved by considering the correlation between the data through the entropy calculation. However, the decision tree can become very complex and have a series of edges that leads to overfitting. In this case, there is a very specific learning and there may be a high computational cost and the generalization for future data will be very weak. The RF algorithm imposes the diversity of each tree separately by selecting a random feature. This is a way to control the complexity of the decision trees because it creates random trees from some labels that limit the overgrowth of the tree. Random Forest can deal with unbalanced data that is present in healthcare datasets and it is robust against overfitting (KHALILIA, CHAKRABORTY, POPESCU, 2011).

5 RESULTS AND DISCUSSION

Several machine learning methods and classifiers were used for early diagnosis of arteriovenous fistula stenosis. The experiments are conducted in Jupyter notebook, using Python v. 3.8 programming language and Sklearn v. 21 algorithms.

The result of each classification algorithm was evaluated using different metrics and validated with k-fold cross validation, which consists of separating the algorithm into training and test samples in k subsets, repeating k times with a different subset. This way, it is possible to eliminate problems of bias and overfitting, using data samples never seen by the algorithm. The stenosis records were split into training set (70%) and test set (30%) with stratify parameter respecting the class distribution on the original dataset. The evaluation metrics used to compare the overall algorithms performances are classification accuracy, F1-score, sensitivity, sensibility. Moreover, the evaluation includes ROC curve and measures the area under the curve (AUC).

Each model lists different results according to the selected parameters. For parameter tuning Grid Search and cross validation were used to estimate the best parameter for each algorithm. The best performance for RF was obtained with number of trees = 10, maximum depth of three = 6, minimum of 2 samples per leaf, and balanced class weight to handle imbalance classification. These parameters resulted in a sensitivity 98.8% and F1-score of 98.4%. KNN performed best with number of neighbors = 15 and distance metric = “manhattan”, obtaining a F1-score of 96%. The SVM best parameters were C = 10, radial basis function (RBF) as kernel and balanced class weight, resulting in a sensitivity of 96%. Finally, the best performance for LR was with C = 0.1 and penalty = “L2” and balanced class weight with an accuracy of 82.7% and F1-score of 83%.

Table 3 summarizes the average results of each model in terms of accuracy, F1-measure, sensitivity, specificity and AUC based on 10-fold cross-validation.

Table 3. The predictive performance of machine learning models.

Algorithm	Accuracy	F1-score	Sensitivity	Specificity	AUC
Random Forest	98.20%	98.40%	98.80%	98.50%	99.90%
K-Nearest Neighbor	97.10%	96.60%	92.00%	99.10%	99.40%
Support Vector Machine	99.00%	98.30%	96.50%	99.10%	99.80%
Logistic Regression	82.70%	83.60%	90.70%	80.60%	91.00%

Source: Elaborated by the author.

From the results showed in Figure 7, the predictive models have achieved very good results in predicting AVF stenosis. The highest sensitivity was 98.8% achieved by Random Forest, followed by SVM’s rate of 96.5%. These are significant results for imbalance classification because sensitivity represents the rate of true positives. K-Nearest Neighbor and SVM reached the best specificity performance of 99.1%. Although KNN has obtained the best specificity it was also one of the less sensitive to predict stenosis, second only to 90% from LR. The overall accuracy of predictions was excellent, ranging from the lowest value of 82.7% from LR to the very high 99% and 98.2% from SVM and RF respectively.

As Figures 8 illustrates, the highest F1-score of 98.4% gives RF a slight advantage over SVM score of 98.3%. The same is also true for AUC values, with a result of 99.9% for RF and 99.8% for SVM. The ROC plot (Figure 9) summarizes the performance of the classification models on the positive class, demonstrating the similarity in the models results except for LR.

The precision/recall curve (Figure 10) shows that LR was the least performed model, while all the other classifiers show a very similar curve as well, proving the high skill of these models to predict the positive cases of stenosis. Perhaps, the number of samples was too small for LR when compared to the number of input features.

Table 4. Test set confusion matrices of the four models. (TN, True Negative; FN, False Negative; FP, False Positive; TP, True Positive)

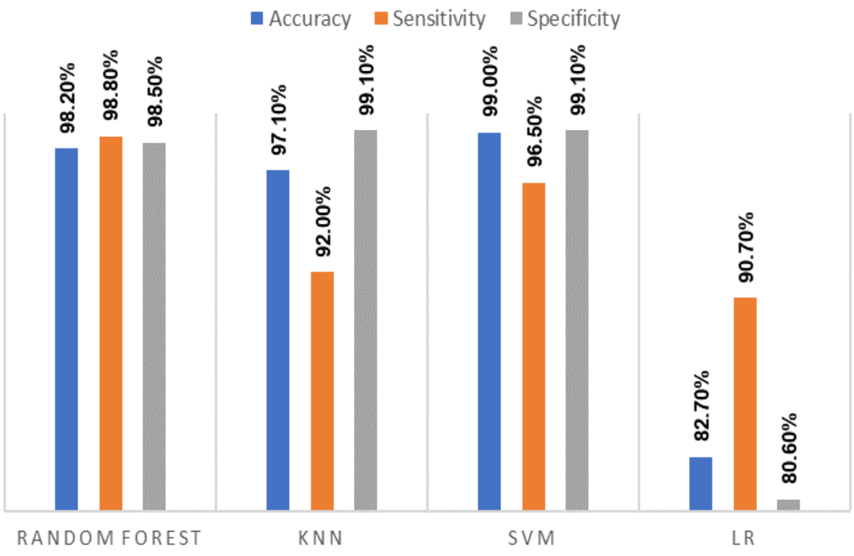
Confusion matrix	Actual	Prediction	
		Not Stenosis	Stenosis
Random Forest	Not Stenosis	TN = 329	FP = 0
	Stenosis	FN = 2	TP = 114
K-Nearest Neighbor	Not Stenosis	TN = 328	FP = 1
	Stenosis	FN = 2	TP = 114
Support Vector Machine	Not Stenosis	TN = 327	FP = 2
	Stenosis	FN = 1	TP = 115
Logistic Regression	Not Stenosis	TN = 262	FP = 67
	Stenosis	FN = 7	TP = 109

Source: Elaborated by the author.

The confusion matrix was also calculated for the four models (Table 4). As shown in Table 4, LR generated a number of False Negatives (FN) = 7 and a large amount of False Positives (FP) = 67 during the test process, while the other models had very close number of FN (ranging from 1 to 2), where the SVM model produced the least amount of FN =1 . The RF model produced the minimum number of FP = 1.

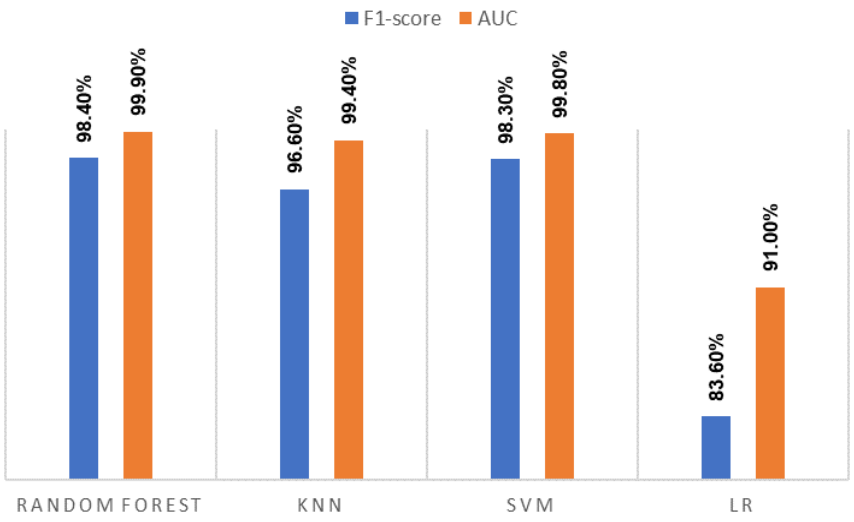
Considering the metrics of F1-score, sensitivity and specificity, the overall performances showed very accurate models, even the least significant results achieved by Logistic Regression are up to 90%.

Figure 7. Accuracy, sensitivity and specific from the models.



Source: Elaborated by the author.

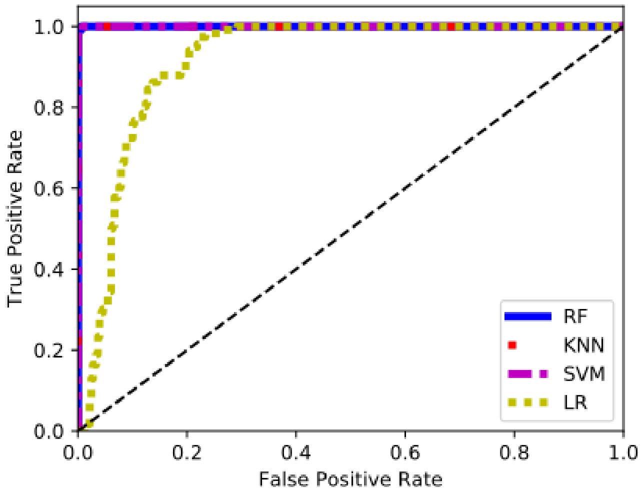
Figure 8. F1-score and area under the curve from the models.



Source: Elaborated by the author.

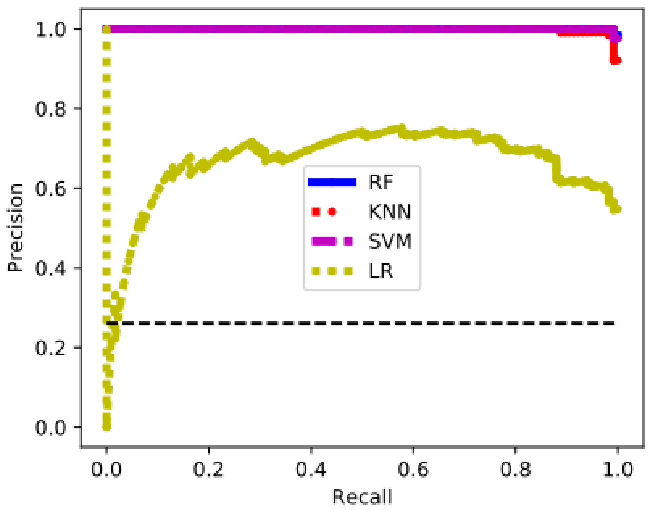
These results are higher than the performance reached by Grochowina (2020), who's highest F1-score of 92% was achieved with KNN, compared to the F1-score of 96% achieved in this study by the same classifier. Also, higher performance compared to the RF accuracy of 76% and F1-score of 88%. The SVM classifier accuracy of 99% was higher than the accuracy of 87% registered by Chiang (2020) in the prediction of fistula's degree of stenosis. Hence, with an almost perfect ROC and precision/recall curves, and a very balanced scores of 98% on average, RF had the best performance among the classifiers. Comparing these results with Kordzadeh (2019) accuracy of 89% achieved by ANN and the overall performance of Random Survival Forest presented in the study of Qian (2020), we can prove the high predictive value of this model for early predictions of AVF stenosis.

Figure 9. The ROC curves from the four evaluated models.



Source: Elaborated by the author.

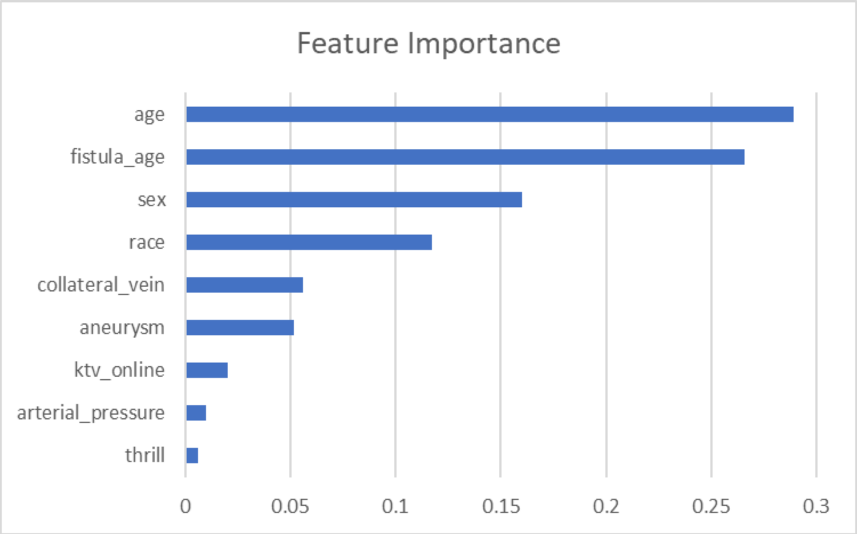
Figure 10. The Precision/Recall curves from the four evaluated models.



Source: Elaborated by the author.

Since RF achieved the most significant results, we investigate the importance of the features. As shown in Figure 11, the patient's age has the highest score while thrill has the lowest score. With a degree of importance of 0.29 the age parameter shows to be a relevant factor in the AVF condition monitoring. According to studies, the older the patient is, the lower the chance of successful fistula maturation (QIAN et al, 2020). In this study, 66% of the patients diagnosed with stenosis are older than 65 years, proving the importance of this parameter.

Figure 11. Importance of features in Random Forest model.



Source: Elaborated by the author.

The fistula age also shows an importance of 0.26 for stenosis prediction. According to the research carried out by BORZUMATI et al (2013), the survival rate of the AVF in older patients was 76% in the first year and 71% in two years. In this study, 34% of the patients diagnosed with stenosis and older than 65 years had complications within the first year of the fistula creation. For the fistula age greater than or equal four, 33% of the patients diagnosed with stenosis are older than 65.

In the stenosis dataset, 54% of all patients are male, whereas 38% of the male patients had stenosis. Only 10% of the female patients had stenosis. Some studies stated that the gender has interference on the fistula outcome and generally the majority of AVF failures occur in female patients, contrary to this dataset results (QIAN et al, 2020). However, these results were usually achieved in bigger datasets with more diversified samples. In addition, the patient's race, development of collateral vein, aneurysm, dose of Kt/v per session, arterial pressure and thrill were also identified as important predictors of stenosis.

Machine learning is proving to be one of the most relevant tools to help in health care. The large sets of data originated from the medical area in combination with machine learning has led to great advances in the early diagnosis of innumerous diseases. Patients undergo hemodialysis therapy for years and the arteriovenous fistula deteriorates over time, which can lead to serious complications such as stenosis. The use of machine learning enhances prediction accuracy and can find new features and relationship not previously explored, thus AVF monitoring is essential for early identification of complications, improving the patient’s quality of life and reduced treatment and surgical intervention costs (CHIANG, 2019). In this study, we applied four machine learning classifiers that have already proven to be effective in predicting stenosis and complications of arteriovenous fistula based on related work researches. All the models showed good results, mainly RF who’s was able to predict stenosis with an accuracy rate of 98.2% and a F1-score of 98.4%, which indicates the balance of precision and recall considering the false positives and false negative, proving the model excellent performance for early diagnosis of stenosis in a imbalanced classification.

This is one of the few studies to use data from the hemodialysis machine. This data is usually discarded as the machine doesn’t provide an easy way to access them, making its collection only possible by manual notes which is more susceptible to errors. Besides that, this stenosis dataset had diversified features obtained through physical examination and clinical data, proving to be a valuable source of data. However, this dataset is considered to be small (1483), which affects the results and the finding of new features. Another difficult is to find datasets with the same features for comparison.

Hence, it is necessary to build a larger dataset, thus containing more diversified samples in order to obtain more predictors for stenosis and to consolidate these results. Though, it would take years and thousands of patients to build a large dataset with all these features, which is aggravated by the necessity of manual notes that are very error prone. Therefore, it is important to come up with optimal solutions for collecting data and AVF monitoring. Many researchers are developing non-invasive, low costs devices to assess the quality of the arteriovenous fistula. These devices are generally based on photoplethysmography and can provide predictions at any time and place due to its portability (GROCHOWINA; LENIOSWKA; BŁĄDZIŃSKA, 2020).

6 CONCLUSION

This work presented a proposal for a machine learning model to assist in the early diagnosis of arteriovenous fistula stenosis in such a way that it is possible to start appropriate treatment before the need for a surgical procedure and possible decrease in the patient's quality of life. We approach this aim by applying four machine learning classifiers: Random Forest, KNN, SVM and Logistic Regression on a dataset of 1483 records. In order to reduce the number of features and redundant data, the association between variables have been studied.

The classifiers have been trained and tested using 10-fold cross-validation. Random Forest achieved the best performance with an F1-score of 98.40%, sensitivity of 98.80% and specificity of 98.50%. This result is the highest among related works. We also found that patient’s age, fistula age and gender had higher importance for Random Forest in predicting stenosis.

The main scientific contribution of this work is to provide an effective model to predict arteriovenous fistula stenosis, using data available in the hemodialysis machine and dialysis session. This data is usually discarded as the machine is not accessible. This study is one of the few to use this information in order to find possible relevant parameters to assist in the early diagnosis of stenosis.

There are some limitations to this work. This dataset is small, affecting the results and the finding of new features. Another difficult is to find datasets with the same features for comparison. These limitations open possibilities for further development of future works, for instance, testing this model in a larger and more diverse dataset, to compare and validate the obtained results in order to create an optimal data pattern for early diagnosis of stenosis. Another opportunity for future work is to develop more automate and easier ways of monitoring and collecting arteriovenous fistula information, investing in low-cost solutions through portable hardware or other solutions that can be replicated and validated in a less time-consuming manner.

REFERENCES

ABREO, K.; AMIN, B. M.; ABREO, A. P. **Physical examination of the hemodialysis arteriovenous fistula to detect early dysfunction.** J Vasc Access. Vol. 20(1), p. 7–11, 2019.

BELMONT, B. et al. **A pilot study to measure vascular compliance changes during fistula maturation using open-source software.** J Vasc Access. Vol. 20(1), p.41-45. 2019.

BHATIA, G. et al. **Machine Learning for Prediction of Life of Arteriovenous Fistula.** In: 3rd International Conference for Convergence in Technology (I2CT), 2018, Pune. Proceeding. IEEE, 2018. p. 1-6.

BOLLEGALA, D. **Dynamic feature scaling for online learning of binary classifiers.** Knowledge-Based Systems. Vol. 129, p. 97-105, 2017.

BORZUMATI, M. et al. **Survival and complications of arteriovenous fistula dialysis access in an elderly population**. J Vasc Access. Vol. 14(4), p. 330-334, 2013.

CHIANG, P.Y. et al. **Machine Learning Classification for Assessing the Degree of Stenosis and Blood Flow Volume at Arteriovenous Fistulas of Hemodialysis Patients Using a New Photoplethysmography Sensor Device**. Sensors (Basel). Vol. 19(15), p. 3422, 2019.

COENTRÃO, L.; RODRIGUES, L. T. **Monitoring dialysis arteriovenous fistulae: it's in our hands.** J Vasc Access. Vol.14(3), p. 209-215, 2013.

CREWS, D. C.; BELLO, A.K.; SAADI, G. **Burden, Access, and Disparities in Kidney disease.** Kidney International. Vol. 95, p. 242-248, 2019.

DENESHKUMAR, V.; MANIKANDAN, M.; KALIYAPERUMAL, S. K. **Identification of Outliers in Medical Diagostic System Using Data Mining Techniques**. International Journal of Statistics and Applications. Vol. 4(6), p. 241-248, 2014.

DIAS, T. S.; NETO, M. M.; COSTA, J. A. C. **Arteriovenous Fistula Puncture: An Essential Factor for Hemodialysis Efficiency**. Renal Failure. Vol. 30(9), p. 870-876, 2008.

GRASSELLI, C. S. M. et al. **Assessment of quality of life in patients undergoing hemodialysis**. Rev Bras Clin Med. Vol. 10(6), p. 503-507, 2012.

GROCHOWINA, M.; LENIOSWKA, L.; BŁĄDZIŃSKA, B-C. **The prototype device for non-invasive diagnosis of arteriovenous fistula condition using machine learning methods**. Sci Rep. Vol. 10(1), no. 16387, 2020.

GROCHOWINA, M.; LENIOWSKA, L. **Comparison of SVM and k-NN classifiers in the estimation of the state of the arteriovenous fistula problem.** In: 2015 Federated Conference on Computer Science and Information Systems (FedCSIS). 2015, Lodz. Proceeding. IEEE, p. 249-254.

HADDAD. N. J. et al. **Hemodialysis access monitoring and surveillance, how and why?** Semin Intervent Radiol. Vol. 33(1), p. 25–30, 2014.

HAYASHI, R.; HUANG, E.; NISSENSON A. R. **Vascular access for hemodialysis.** Nature Clinical Practice Nephrology. Vol. 2, p. 504–513, 2006.

HYEOUN-AE, P. **An Introduction to Logistic Regression: From Basic Concepts**

to Interpretation with Particular Attention to Nursing Domain. J Korean Acad Nurs. Vol. 43(2), p. 154-164, 2013.

JONATHAN, A. C. et al. **Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls**. BMJ. Vol. 338, p. 157-160, 2009.

KHALILIA, M.; CHAKRABORTY, S.; POPESCU, M. **Predicting disease risks from highly imbalanced data using random forest**. BMC Med Inform Decis Mak. Vol. 11(1), p. 51-64, 2011.

KORDZADEH, A.; ESFAHLANI, S. S. **The Role of Artificial Intelligence in the Prediction of Functional Maturation of Arteriovenous Fistula**. Ann Vasc Dis. Vol. 12(1), p. 44-49, 2019.

KORTOBI, L. et al. **Management of pain at arteriovenous fistula puncture: Cryotherapy versus lidocaine/prilocaine**. Saudi J. Kidney Dis. Transpl. Vol. 31(3), p. 597-603, 2020.

SCHOLZ, S. S. et al. **Effects of Arteriovenous Fistula on Blood Pressure in Patients with End-Stage Renal Disease: A Systematic Meta-Analysis**. J Am Heart Assoc. Vol. 8(4), 2019.

PASKLISNKY, G. et al. **Management of true aneurysms of hemodialysis access fistulas**. J Vasc Surg. Vol. 53(5), p. 1291-1297, 2011.

PATEL, M. S. et al. **Evaluating and treating venous outflow stenoses is necessary for the successful open surgical treatment of arteriovenous fistula aneurysms**. J Vasc Surg. Vol. 61(2), p. 444-448, 2014.

POLIMANTI, A. C. et al. **Influence of intraoperative findings on immediate flow through radial-cephalic arteriovenous wrist fistulas for hemodialysis access**. J Vasc Bras. Vol. 17(3), p. 208-214, 2018.

POLKINGHORNE, K. R. et al. **Does monthly native arteriovenous fistula blood-flow surveillance detect significant stenosis—a randomized controlled trial**. Nephrol Dial Transplant. Vol. 21(9), p. 2498-506, 2006.

QIAN, J. et al. **Selecting important predictors for arteriovenous fistula maturation in older hemodialysis patients by using random survival forests**. Seminars in Dialysis. Vol. 33(2), p. 148-155, 2020.

Sociedade Brasileira de Nefrologia (SBN). **Hemodiálise**. <https://www.sbn.org.br/orientacoes-e-tratamentos/tratamentos/hemodialise/>. Accessed on: Jun. 27, 2020.

VERGARA, J.R.; ESTÉVEZ, P. A. **A Review of Feature Selection Methods Based on Mutual Information.** Neural Comput & Applic. Vol. 24, p. 175-186, 2014.

WHITTIER, W. L. **Surveillance of Hemodialysis Vascular Access**. Semin Intervent Radiol. Vol. 26(2), p. 130-138, 2009.