Determining a Non-Intrusive Voice Quality Model Using Machine Learning and Signal Analysis in Time

LUIZ CARLOS BRANDAO JUNIOR[1]
DEMOSTENES ZEGARRA RODRIGUEZ[2]

Federal University of Lavras
Department of Computer and Science
Brazil
[1]brandao@gmail.com
[2]demostenes.zegarra@ufla.br

Abstract. The purpose of this paper is to determine a solution to estimate the quality of a signal of using time domain signal information and machine learning algorithms in an environment that simulates wireless networks using Voice over Internet Protocol (VoIP). The methodology employed was divided into three stages, and degradations were initially applied in an environment that simulated wireless networks making changes in two parameters being the signal-to-noise ratio (SNR) and the type of modulation scheme. To perform the degradations on six distinct signals, algorithms implemented in MATLAB were used to simulate the effect of fading in wireless environments. In the second step, time domain graphs were plotted that correspond to the degradations and that were saved, 272 of them were used for training on 12 different learning algorithms. implemented in the Weka tool. In the last step, software-trained algorithms implemented in Java called PredictorFX in order to predict the value of MOS through an audio image in the time domain. The results were satisfactory, the best trained regression algorithms called $r1$ were RandomTree, RandomForest and IBk with correlation coefficients ranging from 0.9798 to 0.9982 in the validation phase. In relation to $r2$ the best were RandomTree, RandomForest, IBk and AditiveRegression with correlation coefficient ranging from 0.9375 to 0.9923 in the validation phase. And finally, for the training algorithms for the named c1 classification the best trained algorithms were IBk, RandomTree, RandomForest and J48 with a range of 48.53% to 98.53% of correctly classified instances.

Keywords: Quality metrics, Voice over IP (VoIP), Voice Quality, Degradation, Fade, Wireless, ITU-T P.862 Recommendation, Weka, Machine Learning .

(Received May 1st, 2019 / Accepted June 1st, 2019)

1 Introduction

According to [9] global mobile data traffic grew 71 % in 2017, reaching 11.5 *exabytes* [1] per month at the end of 2017, up from 6.7 *exabytes* per month at the end of 2016, this is mainly due to the increased number of mobile devices. However, a communication channel can be degraded causing loss of voice quality [23]. Therefore, research regarding the evaluation of the quality of a voice signal is relevant to the areas of networks and telecommunications.

Traditional or fixed telephony to function demands a network infrastructure that involves circuit-switched equipment. According to [16] and [19] although having good call quality the service [17, 32, 12, 4, 8, 45] charged is relatively high and there is a complexity of structure for this system to operate [11, 37, 36, 29]. The

[1]One *exabyte* is equivalent to one billion gigabytes *petabytes*

Fixed Switched Telephone Service (STFC) is restricted to an access point where the handset or terminal will be motionless, in addition to allowing communication it also allows data transfer via Asymmetric Digital Subscriber Line (ADSL). It makes the difference between local and long distance calling, with more expensive international dialing and mobile rates. In VoIP telephony [35], the scenario is different, if the user has a mobile phone, thus allowing mobility. However, if the person is talking about an IP phone, or a softphone / dialer, or Analog Telephone Adapter (ATA) with access to the Internet, transmission can be done, part of a wired network and part in *Wireless*, with no differences between local, long distance or international calls.

Several physical phenomena occur in wireless networks [21, 10, 39, 24, 22, 31, 41, 2, 25, 33, 44] such as signal reflection[1], interference, noise, power and others. The signal-to-noise ratio (SNR) refers to the signal that is transmitted may have noise inserted [40, 6, 30, 7, 34, 15, 3]. Radio waves are subject to reflections on the ground that cause changes in their amplitude and path, causing variations in the received signal strength called fading. It is also caused by obstacles in the direct line of sight or attenuation due to rain.

The ITU-T P.862 [28] recommendation, popularly known as Perceptual Evaluation of Speech Quality (PESQ), is an intrusive objective method that estimates an Opinion Score for end-to-end voice quality assessment. end in narrowband telephone networks. This method needs a reference signal to compare with the signal at the receiver and ensure the quality of the score. For this reason intrusive methods are more reliable and are used as a reference for objective assessment. For the non-intrusive method, the ITU-T P.563 [26] recommendation uses a standard algorithm for assessing voice quality that is applicable for voice quality predictions without a reference signal. These non-intrusive methods are those that only need the signal at the receiver, or at a given point where the signal should be evaluated and thus are faster, which enables its use in real time services.

The main contribution of this paper is to determine a non-intrusive voice quality model based on time domain signal analysis using machine learning, such model achieved satisfactory results. In a transmission channel *wireless* several levels of degradation were entered to obtain 272 degraded audios. These audios were plotted in the time domain and saved in image files that were analyzed. The remainder of this paper is structured as follows: section 2 presents a literature review addressing communication systems, radio frequency (RF), phenomena that happen in an RF channel, performance evaluation parameters, modulations, quality of service and experience, voice quality assessment methods and machine learning (*machine learning*); Section 3 deals with the methodology used, which was divided into two stages. In the first, called the general scenario, degradations were applied in environments that simulate Wireless networks. In the second, the degraded audios were plotted and saved in image files that were analyzed and used to train algorithms in order to build models in the WEKA software and using a software developed in Java that can load other images, performing the MOS prediction in files not used for modeling; Section 4 shows the results and the construction of models for use in PredictorFX software developed in Java for the purpose of predicting the value of MOS by means of an audio image. Section 5 concerns the conclusion.

2 Literature review

This section will cover communication systems, modulation concepts and their types, radio frequency (RF), radio frequency properties, artificial neural networks and machine learning.

2.1 Voice Quality Assessment Method

Voice quality assessment methods assigns a quality score to a given communication and can be classified as subjective and objective [13]. The subjective ones are based on the evaluation of people through hearing [38], in other words, there is only dependence on the users opinions to rate the quality of the voice perceived by those same users [42, 20]. The objectives, on the other hand, are based on mathematical models that may be intrusive, which are those methods that require a speech sample at the point of origin, where the communication happened to be able to compare with the destination point sample, providing a quality evaluation result. And for non-intrusives where a sample of the original communication signal is not required, the evaluation is determined only by the signal at the point being analyzed [42]. It is noteworthy that intrusive methods are more accurate than non-intrusive methods, since they use the original and degraded signals. It is also important to note that non-intrusive methods are most appropriate for assessing the quality of real-time services, such as VoIP for example, where the source signal is not available. ITU-T Recommendation P.862 [28], is an objective intrusive method used to predict the subjective quality of narrowband (0.3 - 3.4 kHz) audio encoders being better known as PESQ (*Perceptual Evaluation of Speech Quality*). The result of PESQ is a prediction of

the quality perceived by an individual on a subjective test, generating a quality score when listening to audio, called the MOS (*Mean Opinion Score*).

2.2 Weka (*Waikato Environment for Knowledge Analysis*)

According to [14] Weka is open source Java software developed under the GNU General Public License, implemented by The University of New Zealand (*The University of Waikato*). It is a collection of machine learning algorithms for data mining tasks, contains tools for data preparation, classification, regression, grouping, association rules mining and visualization. There are numerous ways to use Weka, one of which is to apply a learning method to a dataset and analyze its output to learn more about the data [43]. Another is to use learned models to generate predictions about new instances. A third is to apply several different classifiers and compare their performance to choose one for prediction. Also in [14] mentions that the classification and regression algorithms are called *Classifier* and any learning algorithm is derived from the *abstract* class weka.classifiers.AbstractClassifier. This implements weka.classifiers.Classifier. Even a basic classifier needs a routine that generates a model for a training dataset and another routine that evaluates the model generated in that dataset or generates a probability distribution for all classes. A classifier model is an arbitrary complex mapping of a dataset called an all-but-one dataset that is nothing but attributes for the class attribute. The specific form and creation of this mapping, or model, differs from classifier to classifier. Twelve learning algorithms were used: RandomForest (RdnF), RandomTree (RdnT), IBk and MLPClassifier (MLP) where these four were used for both regression and classification; M5P, SMOreg, AdditiveRegression (AddR) and SimpleLinearRegression (SLR) these four were used for regression only; And only for classification were the J48, OneR, JRip and NaiveBayes (NvBy).

2.2.1 Performance Measures

According to cite witten2005data performance measures are all evaluation measures that belong to classification situations rather than numerical prediction situations. The basic principles do not use the training set, but independent testing for the validation method and cross-validation assessment apply equally well to numerical prediction. The predicted values ãâin the test instances are $p_1, p_2, ..., p_n$ and the real are $a_1, a_2, ..., a_n$.

The mean square error (NDE) is determined by adding the squared prediction errors and dividing by the total number of errors used in the calculation and its mathematical formula can be seen in the equation 1.

$$EQM = \frac{\sum_{i=1}^{n}(p_i - a_i)^2}{n} \tag{1}$$

Mean absolute error (EAM), equation 2, is the error where all error sizes are treated uniformly according to their magnitude. Sometimes, relative rather than absolute error values are important.

$$EAM = \frac{\sum_{i=1}^{n}|p_i - a_i|}{n} \tag{2}$$

The relative quadratic error (ERQ), equation 3 is made relative to what it would have been if a simple *classifier* had been used. The equation is just the average of the actual values of the training data, so the relative squared error takes the total squared error into account and normalizes dividing by the total predictor error.

$$ERQ = \frac{\sum_{i=1}^{n}(p_i - a_i)^2}{\sum_{i=1}^{n}(a_i - \bar{a})^2}, \quad onde \quad \bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i \tag{3}$$

Relative absolute error (EAR), equation 4 is the total absolute error with the same type of normalization. In these three relative error measures, the errors are normalized by the simple predictor error that predicts mean values.

$$EAR = \frac{\sum_{i=1}^{n}|p_i - a_i|}{\sum_{i=1}^{n}|a_i - \bar{a}|} \quad onde \quad \bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i \tag{4}$$

Correlation Coefficient (CC) measures the statistical correlation between the a and p of equation 2.2.1. The correlation coefficient ranges from -1 when the results are perfectly negatively correlated, that is, if one increases, the other always decreases, 0 when there is no relationship to 1 for perfectly correlated results. The correlation is slightly different from other measures because it is independent of the scale at which a particular set of predictions is taken, the error is unchanged if all predictions are multiplied by a constant factor and the actual values ãâare left unchanged. This factor appears in all terms of the $cov_{(P,A)}$ in the numerator and each $var_{(P)}$ term in the denominator, thus canceling. However, this is not true for error numbers, although normalization multiplies all predictions by a large constant, so the difference between the predicted and the actual values ãâwill change dramatically, as will the percentage errors. Unlike good performance leading to a large

correlation coefficient value, whereas, as other methods measure error, good performance is indicated by small values.

$$CC=\frac{cov(P,A)}{\sqrt[2]{var(P)*var(A)}}=\frac{\frac{\sum_{i=1}^{n}(p_i-\bar{p})(a_i-\bar{a})}{n-1}}{\sqrt[2]{\frac{\sum_{i=1}^{n}(p_i-\bar{p})^2}{n-1}*\frac{\sum_{i=1}^{n}(a_i-\bar{a})^2}{n-1}}}=\frac{\sum_{i=1}^{n}(p_i-\bar{p})(a_i-\bar{a})}{\sqrt[2]{\sum_{i=1}^{n}(p_i-\bar{p})^2*\sum_{i=1}^{n}(a_i-\bar{a})^2}} \quad (5)$$

where: $\bar{a}=\frac{1}{n}\sum_{i=1}^{n}a_i \qquad \bar{p}=\frac{1}{n}\sum_{i=1}^{n}p_i$

Where the values of CC are:

* 0.9 to 1.0 positive or negative indicates a very strong correlation.
* 0.7 to 0.9 positive or negative indicates a strong correlation.
* 0.5 to 0.7 positive or negative indicates an average correlation.
* 0.3 to 0.5 positive or negative indicates a weak correlation.
* 0.0 to 0.3 positive or negative indicates a very weak correlation.

According to [14] Kappa is a measure of agreement used on nominal scales that gives an idea of how far the observations deviate from those expected, by chance, thus indicating how legitimate the interpretations are where the numerical value 1.0 means complete agreement. The magnitude of Kappa statistics is a more significant measure of agreement than its own statistical significance. The guidelines for interpreting Kappa are given below.

* 0 indicates a very bad agreement.
* 0 to 0.2 indicates a bad agreement.
* 0.21 to 0.4 indicates considerable agreement.
* 0.41 to 0.6 indicates moderate agreement.
* 0.61 to 0.8 indicates substantial agreement.
* 0.81 to 1.0 indicates excellent agreement.

3 Methodology

This section will address the methodology used to conduct this research and is divided into three steps. It first randomly chose six original audio files (or105, or109, or114, or129, or134 and or137) all belonging to the P.862 [5] recommendation database. These audios were used to make degradation applications in environments that simulate wireless networks called the general scenario, Figure 1, in which two parameters were changed, namely: Signal to Noise Ratio (SNR)) and QPSK and QAM modulations; where the quantization level had a fixed value of 2^{16}. Changes in these parameters modified the speech signal, so as to evaluate the voice quality over IP networks, the ITU-T P.862 [28] recommendation was used. To perform these degradations, algorithms implemented in *MATLAB* were used to simulate the effect of *fading* on Wireless *environments*. Through this implementation it was possible to generate different degraded voice signals, thus obtaining a coherent output with real world phenomena in *Wireless* [18].

After saving the images that represent the graphs of time-domain degraded audios, the next step was to perform image feature extraction for only 272 degradations of the original or114 file, data processing, and model construction in Weka. And finally, Step 3 refers to the Java software implementation called PredictorFX, whose function was to use the models built in the previous step to predict the graphics representing their respective degraded audio files coming from the others. five original files being the or105, or109, or129, or134 and or137.

The general scenario shown in Figure 1 shows the steps followed by the code to obtain the MOS index which is a real number and has two parameters, the modulation that alternated between QPSK, 2-QAM, 4-QAM, 16-QAM, 32-QAM, 64-QAM, 128-QAM, 256-QAM and signal-to-noise ratio that changed from 0dB to 33dB, varying from one to one. These two parameters simulated different types of fading as occurs in a Wireless transmission. The general scenario algorithm loaded each original audio file without any degradation, doing linear quantization, then converting to binary and sequencing these bits. Depending on the configuration, the sequenced bits were QPSK or QAM modulated, so different fading types were generated. Then there was the sequence for grouping and later demodulation with the conversion of binary to sound file with degradation, thereby generating different degraded audio signals. Figure 2 shows the original (no degradation) audio files in the time domain and Table 1 contains their respective information. The total amount of degraded audios came from 8 modulations x 34 SNR x 6 audios

totaling 1632 degraded audios, of which 272 were used to build the models.

We chose to do the ranges as shown in Table 2 since the implementation of the general scenario codes provided unnatural values as in the ITU-T P.800 [27] recommendation, but real with the MOS ranging from 0.5 to 4.5.

Table 3 refers to the MOS obtained from the fading files, we chose to select only the original audio set of images or114 with 272 files, could have been any of the six. Each MOS value from Step 1 that was contained within the ranges shown in Table 3 received an objective quality scale. For example, in $Intervalo_{(1)}$, the $\text{MOS}_{Numérico}$ equal to 1 is contained within the range of values ãâgreater than 0.5 and less than or equal to 1.5. All MOS values ãâthat are in this range are rounded to the value 1, thus receiving the appalling denomination. In the $Interval_{(2)}$ values ãâabove 1.5 and less than or equal to 2.5 are rounded to 2, thus receiving the bad rating. The $Interval_{(3)}$ whose values ãâare over 2.5 and less than or equal to 3.5 are rounded to 3 and rated reasonable. And lastly, in the $Interval_{(4)}$ values ãâover 3.5 and less than or equal to 4.5 are rounded to 4 and given the denomination of good.

3.1 Extracting Features from an Image

The different colors contained in the figure 5 $[a]$ contain a different color for each pixel, as shown by its RGB, where R is the initial letter of the English word textit Red (Red), G for *Green* and B for *Blue*. The combination of the three generates the color that is shown in each marked pixel. The *plot* that appears in the figure is the image `or114 _deg _qpsk _SNR _p12` whose modulation is QPSK, 2^{16} quantization level and 12dB SNR producing an MOS of 4 , 5 with a literal representation of good audio.

The color scale ranges from 0 to 255, for each component (R, G, B), totaling 256 that swapped within each RGB variable produces a large amount of colors. To work with images it is advisable to work on a grayscale containing 256 shades of gray, to make this transformation the equation 6.

$$C = 0,2989 * R + 0,5870 * G + 0,1140 * B \quad (6)$$

3.2 Scanning Images

An image saved in a jpg file is a matrix with dimensions (n X p), where n represents the number of rows and p the number of columns. The images used in this search contain 485 pixels of width by 390 height. From the position (49,5) corresponding to the Pixel L belonging to the $p_{(49)}$ column, the scanning process begins, as shown in Figure ref fig: ZoomPixelAzulPos, since it is from from which begins the area where all time domain signals were plotted, the so-called field area is delimited by pixels L, C, B, and J. When the scanning process begins, each pixel in the $p_{(49))}$ is checked until it reaches the position of Pixel C whose coordinate is (49,345). In this process the column is fixed and the scan starts at line 5 and goes line by line until it reaches line number 345, when this happens there is a column change now going to the column $p_{(50)}$, repeating thus the scanning process only ends when the last pixel is reached at position (480,345).

To scan the 272 degraded images from the original file or114, an algorithm that loads one image at a time was implemented in Java and using the 6 equation, the *threshold* is applied. After this process the code counts how many black pixels there are in each column, averaging and storing this value in a dynamic vector, if there are any columns that have no black pixels, there is no storage, so moving to the next column. , composing the $_{(Field)}$ where it contains only the averages of the positions of each column.

Since Java does not have many of the tools that mathematical software has, it was necessary to create a specific mathematical formula for this type of problem that is verified in Equation 3.2, being similar to the diff function of *Software* R, which is necessary to visualize the behavior of the function for each loaded image and to perform the regression and classification processes in Weka software. These values were placed in a 50-position vector called Vector50.

$$(diff)_{a=0}^{49} = \left(\frac{\sum_{m=1}^{Size\_of[Vector_{(Field)}]-50} (PosVector(m+a) - PosVector(m-1))^2}{Size\_of[Vector_{(Field)}]} \right) \quad (7)$$

The values of $(diff)_{a=0}, (diff)_{a=1}, ..., (diff)_{a=49}$ from Equation 7 represent points that together form a Discrete Function (FD), which in turn represents the loaded and scanned image. These are the points that were used in the training process in the algorithms along with the MOS value. In Figure 2 you can see that each function has its own inclination because it depends on the image, it is also noted that the graph 2 $[d]$ seems to have only 4 discrete functions, but there are 127 of them with almost the same slope, in 2 $[c]$ clearly notices that there is only 4 FD in 2 $[b]$ there are

Table 1: Audio data

	or105	or109	or114	or129	or134	or137
Uncompressed Audio	Yes	Yes	Yes	Yes	Yes	Yes
Channel number	1	1	1	1	1	1
Sample rate	8 kHz	8 kHz	8 kHz	8 kHz	8 kHz	8 kHz
Total of samples	67220	64314	68734	57849	64605	56474
Audio duration	8s 402ms	8s 390ms	8s 591ms	7s 231ms	8s 750ms	7s 593ms
Bit/sample	16 bits	16 bits	16 bits	16 bits	16 bits	16 bits
Bits rate	128 Kbps	128 Kbps	128 Kbps	128 Kbps	128 Kbps	128 Kbps
Format	Wave	Wave	Wave	Wave	Wave	Wave
Size of file	131 KB	126 KB	134 KB	113 KB	126 KB	110 KB
Signal of the sound	3,82 %	3,30 %	3,43 %	2,76 %	2,81 %	2,81 %
No Signal Sound	96,18 %	96,70 %	96,57 %	97,24 %	97,19 %	97,19 %

8 of them and in 2 $[a]$ there are 133 FD with a larger distribution.

It is noted in figure 3 that the discrete functions representing the $Interval_{(4)}$ Good MOS have the lowest slopes compared to the others, followed by the $Interval_{(3)}$ FDs which represents the Fair MOS, $Interval_{(2)}$ which represents the Bad MOS and having the highest slopes of the $Interval_{(1)}$ of the Poor MOS. Note also one or some of the $Interval_{(1)}$ FDs passing near the abscissa axis and four functions being two of MOS Fair and two of MOS Good overlapping, just as between two of Bad MOS with MOS Terrible.

3.3 Weka Algorithm Training

Altogether 3 files were built for analysis in Weka, two involving only numerical values used for regression, called r1 and r2; and one with numerical and literal values ãâused for classification, called c1. All files each contain 272 instances, r1 has only 4 attributes, ie 3 parameters which are SNR ranging from 0 to 33, modulation and quantization level that was set to $2^{16} \rightarrow 65536$, the three parameters generated the MOS $_{(Num)}$. R2 is the junction of Vetor50 and its MOS $_{(Num))}$ of each image. C1, in turn, is the junction of the same data as r2, but with its MOS $_{(Literal)}$ as seen in Table 2. Both r2 and c1 each contain 272 instances with 51 attributes and the three Weka files were analyzed as shown in the schemas in Figure 6.

4 Results and Discussions

This section discusses the results found for both regression and classification. The trained algorithms were built using degraded images from the original or114 file which totaled 272 images out of 1632, the rest were used to validate the model, totaling 1360 images.

4.1 Regression Result for r1

The data used in r1 are the same used in the general scenario for degradation of the original files and refer to the quantization level set at 2^{16}, the eight modulations being: QPSK, 2-QAM, 4-QAM, 16 -QAM, 32-QAM, 64-QAM, 128-QAM and 256-QAM, plus SNR ranging from 0 to 33, totaling 272 instances that match the extracted characteristics of each image with their respective MOS values for the original audio or114 . Each classifier was repeated 50 times over a Cross-Validation K-fold of 10, totaling 500 validations or 5000 iterations, with 4500 used for training and 500 for testing in the 272 instances, Table 4 shows the results of the correlation coefficient averages, where the algorithm with the highest average was RdnF with 0.9858 followed by RdnT with 0.9626.

Table 5 shows the summary of all trained algorithms using the entire or114 database for training and testing. The acronym CC is the Correlation Coefficient, EAM is the Mean Absolute Error which is the difference between Table 3 values and predicted values. NDE is the Mean Square Error which is very useful in comparing the algorithms since it shows that the most effective algorithm is simply the one with the smallest variance. EAR is the Relative Absolute Error which is the ratio between the absolute error and the predicted value of a number. EQR is the Relative Quadratic Error and NTI is the Total Number of Instances.

From the table 5 it can be seen that the IBk algorithm hits all instances with a correlation coefficient of 1 and all errors equal to zero in comparison with the other algorithms, this shows that it is the most effective. , because it has the smallest variance. RdnF has an EAM about 2.36 times higher and NDE about 2.59 times higher than RdnT. Note that from M5P everyone has an increasing NDE. In addition the correlation co-

Table 2: Intervals of MOS$_{Numérico}$ and MOS$_{Literal}$

	Beggining				End		MOS$_{Numérico}$	MOS$_{Literal}$
$Intervalo_{(1)}$	0.5	$<$	x	$\leq$	1.5	$\approx$	1	very bad
$Intervalo_{(2)}$	1.5	$<$	x	$\leq$	2.5	$\approx$	2	bad
$Intervalo_{(3)}$	2.5	$<$	x	$\leq$	3.5	$\approx$	3	moderate
$Intervalo_{(4)}$	3.5	$<$	x	$\leq$	4.5	$\approx$	4	good

Table 3: MOS values for faded audios from original file or114

Repetition	QPSK	2-QAM	4-QAM	16-QAM	32-QAM	64-QAM	128-QAM	256-QAM
p_{00}	0,6	0,8	0,6	0,7	0,7	0,7	0,6	0,7
p_{01}	0,6	1,0	0,6	0,6	0,6	0,7	0,7	0,7
p_{02}	0,6	1,0	0,6	0,6	0,6	0,7	0,6	0,7
p_{03}	0,7	1,0	0,7	0,7	0,6	0,6	0,6	0,7
p_{04}	0,7	2,1	0,7	0,7	0,6	0,7	0,7	0,7
p_{05}	0,8	3,6	0,8	0,6	0,6	0,6	0,6	0,6
p_{06}	1,0	4,4	0,9	0,7	0,6	0,7	0,7	0,7
p_{07}	0,9	4,5	0,9	0,6	0,6	0,7	0,6	0,7
p_{08}	1,0	4,5	1,2	0,7	0,6	0,6	0,6	0,7
p_{09}	2,3	4,5	2,5	0,7	0,6	0,6	0,6	0,7
p_{10}	3,9	4,5	4,0	0,8	0,6	0,6	0,6	0,7
p_{11}	4,4	4,5	4,5	0,8	0,6	0,6	0,7	0,7
p_{12}	4,5	4,5	4,5	0,9	0,7	0,6	0,7	0,6
p_{13}	4,5	4,5	4,5	0,8	0,7	0,7	0,7	0,6
p_{14}	4,5	4,5	4,5	0,7	0,7	0,7	0,7	0,6
p_{15}	4,5	4,5	4,5	1,7	0,8	0,8	0,7	0,7
p_{16}	4,5	4,5	4,5	3,3	0,9	0,8	0,7	0,6
p_{17}	4,5	4,5	4,5	4,3	1,0	0,9	0,7	0,7
p_{18}	4,5	4,5	4,5	4,5	1,0	1,0	0,8	0,7
p_{19}	4,5	4,5	4,5	4,5	1,0	0,7	0,8	0,7
p_{20}	4,5	4,5	4,5	4,5	1,1	0,9	0,9	0,8
p_{21}	4,5	4,5	4,5	4,5	2,1	1,9	0,9	0,8
p_{22}	4,5	4,5	4,5	4,5	3,3	3,3	1,0	0,9
p_{23}	4,5	4,5	4,5	4,5	4,1	4,3	1,1	1,0
p_{24}	4,5	4,5	4,5	4,5	4,5	4,5	1,1	0,8
p_{25}	4,5	4,5	4,5	4,5	4,5	4,5	0,9	0,7
p_{26}	4,5	4,5	4,5	4,5	4,5	4,5	1,2	1,2
p_{27}	4,5	4,5	4,5	4,5	4,5	4,5	2,1	2,3
p_{28}	4,5	4,5	4,5	4,5	4,5	4,5	3,5	3,6
p_{29}	4,5	4,5	4,5	4,5	4,5	4,5	4,3	4,3
p_{30}	4,5	4,5	4,5	4,5	4,5	4,5	4,5	4,5
p_{31}	4,5	4,5	4,5	4,5	4,5	4,5	4,5	4,5
p_{32}	4,5	4,5	4,5	4,5	4,5	4,5	4,5	4,5
p_{33}	4,5	4,5	4,5	4,5	4,5	4,5	4,5	4,5

Table 4: Correlation coefficients means for r1.

	RdnF	RdnT	M5P	IBk	MLP	SMOreg	AddR	SLR
MÃ©dia	0,9858	0,9626	0,9331	0,9122	0,8262	0,7907	0,7777	0,6892

efficients of the first three are very close.

The graph in Figure 3 shows the behavior of the algorithms trained in the different databases, where five of them had their CC above 0.879; but only three of them were close to 1 which were IBk, RdnT and RdnF, almost with an overlap of values, thus indicating wide acceptance of these algorithms for this type of problem.

Figure 4 shows the behavior of the mean absolute error for each algorithm across the different databases and it can be seen that the trained algorithms that had the smallest errors were IBk, RdnT and RdnF where there is almost overlap. of values.

Figure 5 shows the mean square error behavior for each algorithm across the different databases, and it can be seen that the algorithms that had the smallest errors were IBk, RdnT and RdnF, with almost an overlap. values.

4.2 Regression Result for r2

The data contained in r2 are the values of 272 discrete functions that together with their respective MOS from the or114 database were used to train the 8 regression algorithms. Table 6 contains the means of the correlation coefficients corresponding to the 50 validations with the IBk algorithm with the highest correlation coefficient, followed by RdnT, RdnF and AddR.

Table 7 shows the summary of all trained algorithms using the entire or114 database for training and testing.

When analyzing the table it is clear that both the EAM and the NDE of the IBk algorithm are smaller compared to the others, indicating a good acceptance of this classifier for this type of problem. Note also that the correlation coefficients of IBk RandomTree, RandomForest and AdditiveRegression are very close.

The graph in Figure 6 shows the behavior of the correlation coefficients for each database. It can be seen that in the or114 database that contains the degraded file data that was used to train the algorithms, only the SMOreg and the SLR that were below 0.7172, all the others were above 0.9251. From the database of degraded files or105 begins the process of model validation by trained algorithms, it is clear that the trained algorithms AddR, RdnT, IBk, RndF, M5P and MLP can predict, in different databases, the MOS with correlation coefficient values ââranging from 0.8283 to 1. This indicates that algorithms trained using the degraded or114 file database can predict MOS in other databases with other data, values ââthat have never been used for training. the algorithms.

4.3 Classification Result for c1

Each result of the *classifier* corresponds to an average of 50 repetitions of instances correctly classified with the *K-fold* option of *Cross-Validation* equal to 10. Table 8 shows the result of these averages, where the algorithm that had a higher average was the IBk with 84.15 %.

5 Conclusion

The present work has shown that it is possible to measure signal quality non-intrusively in time domain audio representations using a model based on the intrusive method. Different levels of degradation were inserted in a transmission channel *wireless* to obtain different degraded audios, in a total of 1632, and only 272 were used to train the algorithms. This proposed model aimed to determine the quality of the voice signal in the 1360 images that were never used for any kind of training and model proved satisfactory, since using only time domain images and machine learning techniques are enough to do the quality test. Regarding the training time, depending on the amount of data to train the algorithms can be a little time consuming, however, once trained you can use the model quite quickly.

For r1 the best-fitting algorithms were undoubtedly IBk, *RandomForest* and *RandomTree* with very high correlation coefficients, small mean absolute errors in relation to each other and also having the smallest mean square errors indicating that Such trained algorithms were well accepted for this kind of problem. For r2 the IBk, RdnT, RdnF and AddR algorithms were the best, obtaining high correlation coefficients ranging from 0.9375 to 0.9923 from the or105 to or137 database. RandomTree mean absolute errors were the only ones that remained lower than the others and from the or109 database there was almost an overlap of values for the AdditiveRegression, RandomTree and IBk algorithms. RdnT also had the smallest mean square error. For c1 the best algorithms were IBk and RdnT which although they hit fewer instances in the database or134 were the algorithms that were better than the others.

References

[1] Abreu, D. H. S., Rodríguez, D. Z., and Lacerda, W. S. Impact of the linear phase variations on the voice quality index. In *2016 IEEE International Symposium on Consumer Electronics (ISCE)*, pages 33–34, Sep. 2016.

[2] Affonso, E. T., Nunes, R. D., Rosa, R. L., Pivaro, G. F., and Rodríguez, D. Z. Speech quality assess-

Table 5: Summary of Trained Algorithms for or114_r1.

	IBk	RdnT	RdnF	M5P	SMOreg	MLP	AddR	SLR
CC	1	0,9998	0,9985	0,9533	0,7951	0,9080	0,8333	0,6934
EAM	0	0,0232	0,0548	0,3873	0,8934	0,4298	0,8782	1,0280
EQM	0	0,0408	0,1052	0,5612	1,1165	0,7672	1,0119	1,3188
EAR	0%	1,2883%	3,0369%	21,4738%	49,5334%	23,8271%	48,6904%	56,9926%
EQR	0%	2,2281%	5,7459%	30,6613%	61,0002%	41,9164%	55,2853%	72,0557%
NTI	272	272	272	272	272	272	272	272

Table 6: Mean correlation coefficients for r2.

	IBk	RdnF	RdnT	AddR	M5P	MLP	SMOreg	SLR
MÃ©dia	0,9908	0,9913	0,9865	0,9822	0,9822	0,6618	0,6176	0,6030

Table 7: Summary of Trained Algorithms for or114_r2.

	IBk	RdnT	RdnF	M5P	SMOreg	MLP	AddR	SLR
CC	0,9999	0,9998	0,9980	0,9638	0,7157	0,9251	0,9914	0,7172
EAM	0,0077	0,0223	0,0484	0,2974	0,9438	0,4263	0,1383	1,1130
EQM	0,0309	0,0408	0,1166	0,4929	1,5152	0,7069	0,2392	1,2735
EAR	0,4295%	1,2391%	2,6864%	16,5112%	52,4024%	23,6694%	7,6789%	61,7968%
EQR	1,6908%	2,2323%	6,3807%	29,9705%	82,9061%	38,6812%	13,0892%	69,6840%
NTI	272	272	272	272	272	272	272	272

Table 8: Averages of Correctly Rated Instances for c1.

	IBk	RdnF	OneR	J48	RdnT	NaiveBayes	JRip	MLP
MÃ©dia	84,15%	82,54%	81,18%	81,11%	79,90%	78,63%	78,20%	81,11%

ment in wireless voip communication using deep belief network. *IEEE Access*, 6:77022–77032, 2018.

[3] Affonso, E. T., Rodríguez, D. Z., Rosa, R. L., Andrade, T., and Bressan, G. Voice quality assessment in mobile devices considering different fading models. In *2016 IEEE International Symposium on Consumer Electronics (ISCE)*, pages 21–22, Sep. 2016.

[4] Affonso, E. T., Rosa, R. L., and Rodríguez, D. Z. Speech quality assessment over lossy transmission channels using deep belief networks. *IEEE Signal Processing Letters*, 25(1):70–74, Jan 2018.

[5] Annex, T. D. S. I.-T. P. R. Series p: Telephone transmission quality,telephone installations, local line networks. methods for objective and subjective assessment of quality. 2005.

[6] Bai, H. and Atiquzzaman, M. Error modeling schemes for fading channels in wireless communications: A survey. *IEEE Communications Surveys Tutorials*, 5(2):2–9, Fourth 2003.

[7] Baki, A. K. M., Absar, M. W., Rahman, T., and Ahamed, K. M. A. Investigation of rayleigh and rician fading channels for state of the art (soa) lte-ofdm communication system. In *2017 4th International Conference on Advances in Electrical Engineering (ICAEE)*, pages 325–329, Sept 2017.

[8] Begazo, D. C., Rodríguez, D. Z., and Ramírez, M. A. No-reference video quality metric based on the packet delay variation parameter. In *2016 IEEE International Symposium on Consumer Electronics (ISCE)*, pages 83–84, Sep. 2016.

[9] Cisco, W. P. Cisco visual networking index: Global mobile data traffic forecast update, 2017â2022. February 2019.

[10] da Silva, M. J., Begazo, D. C., and Rodríguez, D. Z. Evaluation of speech quality degradation due to atmospheric phenomena. In *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–6, Sep 2019.

[11] Dantas Nunes, R., Pereira, C. H., Rosa, R. L., and Rodríguez, D. Z. Real-time evaluation of speech quality in mobile communication services. In *2016 IEEE International Conference on Consumer Electronics (ICCE)*, pages 389–390, Jan 2016.

[12] de Almeida, F. L., Rosa, R. L., and Rodríguez, D. Z. Voice quality assessment in communication services using deep learning. In *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, pages 1–6, Aug 2018.

[13] El-Hennawey, S. C23. self-healing autonomic networking for voice quality in voip and wireless networks. In *2015 32nd National Radio Science Conference (NRSC)*, pages 297–304, March 2015.

[14] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[15] Kaur, T., Singh, J., and Sharma, A. Simulative analysis of rayleigh and rician fading channel model and its mitigation. In *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–6, July 2017.

[16] Kyeong Deok Moon, Sang Jun Lee, and Jong Kyu Lee. Performance comparison between dar and fr in grid topology circuit switched networks on common channel signaling. In *Proceedings of TENCON '93. IEEE Region 10 International Conference on Computers, Communications and Automation*, volume 3, pages 622–625 vol.3, Oct 1993.

[17] Lasmar, E. L., de Paula, F. O., Rosa, R. L., Abrahão, J. I., and Rodríguez, D. Z. Rsrs: Ridesharing recommendation system based on social networks to improve the user's qoe. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–13, 2019.

[18] Matlab. Fading channels. DisponÃvel em: `http://mathworks.com/help/comm/ug/fadingchannels.html`. Acesso em: 13 novembro 2018, 2018.

[19] Maw, T. and Fontaine, Y. The public switched telephone network and the internet meet. In *CCECE '97. Canadian Conference on Electrical and Computer Engineering. Engineering Innovation: Voyage of Discovery. Conference Proceedings*, volume 2, pages 892–895 vol.2, May 1997.

[20] Militani, D., Begazo, D. C., Rosa, R., and Rodríguez, D. Z. A speech quality classifier based on

signal information that considers wired and wireless degradations. In *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–6, Sep. 2019.

[21] Militani, D., Vieira, S., Valadao, E., Neles, K., Rosa, R., and Rodríguez, D. Z. A machine learning model to resource allocation service for access point on wireless network. In *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–6, Sep. 2019.

[22] Mohammed, A. A., Yu, L., Al-Kali, M., and Adam, E. E. B. Ber analysis and evaluated for different channel models in wireless cooperation networks based ofdm system. In *2014 Fourth International Conference on Communication Systems and Network Technologies*, pages 326–330, April 2014.

[23] Neves, F., Soares, S., and Assuncao, P. A. A. Optimal voice packet classification for enhanced voip over priority-enabled networks. *Journal of Communications and Networks*, 20(6):554–564, Dec 2018.

[24] Nunes, R. D., Rosa, R. L., and Rodríguez, D. Z. Performance improvement of a non-intrusive voice quality metric in lossy networks. *IET Communications*, August 2019.

[25] Orozco, G. N. and de Almeida, C. Performance evaluation of opportunistic wireless transmission in rayleigh fading channels with co-channel interference. In *2013 IEEE Latin-America Conference on Communications*, pages 1–6, Nov 2013.

[26] P.563, I.-T. Single-ended method for objective speech quality assessment in narrow-band telephony applications. Technical report, 2004.

[27] P.800, T. D. S. I.-T. Methods for subjective determination of transmission quality. Technical report, 1996.

[28] P.862, T. D. S. I.-T. Series p: Telephone transmission quality,telephone installations, local line networks. methods for objective and subjective assessment of quality. Technical report, 2007.

[29] Rodríguez, D. Z., Abrahão, J., Begazo, D. C., Rosa, R. L., and Bressan, G. Quality metric to assess video streaming service over tcp considering temporal location of pauses. *IEEE Transactions on Consumer Electronics*, 58(3):985–992, August 2012.

[30] Rodríguez, D. Z., Arjona Ramírez, M., Bernardes, L. F., Mittag, G., and Möller, S. Impact of fec codes on speech communication quality using wb e-model algorithm. In *2019 Wireless Days (WD)*, pages 1–4, April 2019.

[31] Rodríguez, D. Z. and Möller, S. Speech quality parametric model that considers wireless network characteristics. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6, June 2019.

[32] Rodríguez, D. Z., P'ivaro, G. F., Rosa, R. L., Mittag, G., and Möller, S.

[33] Rodríguez, D. Z., Pívaro, G. F., Rosa, R. L., Mittag, G., and Möller, S. Quantifying the quality improvement of mimo transmission systems in voip communication. In *2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–5, Sep. 2018.

[34] Rodríguez, D. Z., Rosa, R. L., Almeida, F. L., Mittag, G., and Möller, S. Speech quality assessment in wireless communications with mimo systems using a parametric model. *IEEE Access*, 7:35719–35730, 2019.

[35] Rodríguez, D. Z., Rosa, R. L., and Bressan, G. No-reference video quality metric for streaming service using dash standard. In *2015 IEEE International Conference on Consumer Electronics (ICCE)*, pages 106–107, Jan 2015.

[36] Rodríguez, D. Z., Rosa, R. L., Costa, E. A., Abrahão, J., and Bressan, G. Video quality assessment in video streaming services considering user preference for video content. *IEEE Transactions on Consumer Electronics*, 60(3):436–444, Aug 2014.

[37] Rodríguez, D. Z., Wang, Z., Rosa, R. L., and Bressan, G. The impact of video-quality-level switching on user quality of experience in dynamic adaptive streaming over http. *EURASIP Journal on Wireless Communications and Networking*, 2014(1):216, Dec 2014.

[38] Sun, L. Speech quality prediction for voice over internet protocol networks. 2004.

[39] Vieira, S. T., Valadoo, E., Rodríguez, D. Z., and Rosa, R. L. Wireless access point positioning optimization. In *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–6, Sep. 2019.

[40] Wang, A. and Xu, H. Comparison of several snr estimators for qpsk modulations. In *2012 International Conference on Computer Science and Service System*, pages 77–80, Aug 2012.

[41] Wang, L.-C. and Cheng, Y.-H. A statistical mobile-to-mobile rician fading channel model. In *2005 IEEE 61st Vehicular Technology Conference*, volume 1, pages 63–67 Vol. 1, May 2005.

[42] Wei Zha and Wai-Yip Chan. Voice quality assessment using classification trees. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 1, pages 537–541 Vol.1, Nov 2003.

[43] Witten, I., Frank, E., Hall, M., and Pal, C. *Data Mining: Practical Machine Learning Tools and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2016.

[44] Zegarra Rodríguez, D., Lopes Rosa, R., and Bressan, G. Improving a video quality metric with the video content type parameter. *IEEE Latin America Transactions*, 12(4):740–745, June 2014.

[45] Zegarra Rodríguez, D., Lopes Rosa, R., Costa Alfaia, E., Issy Abrahão, J., and Bressan, G. Video quality metric for streaming service using dash standard. *IEEE Transactions on Broadcasting*, 62(3):628–639, Sep. 2016.

Figure 1: General scenario that simulates communication similar to a real environment.

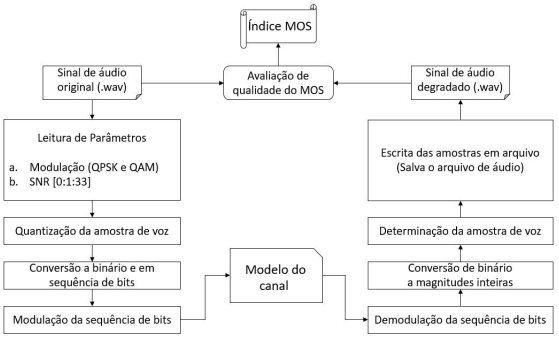
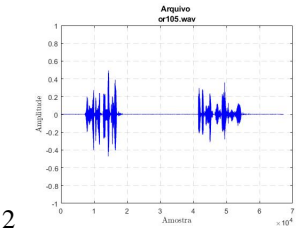
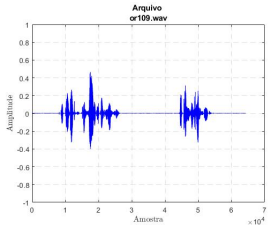


Figure 2: Images of the original audios or105, or109, or114, or129, or134 and or137.

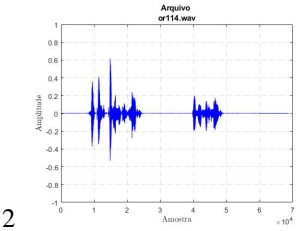


2

[a] file or105

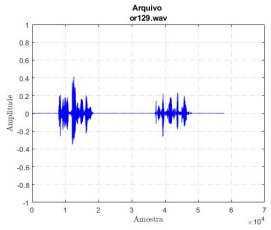


[b] file or109

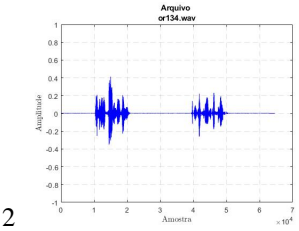


2

[c] file or114

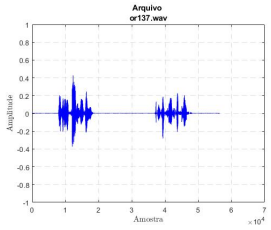


[d] file or129



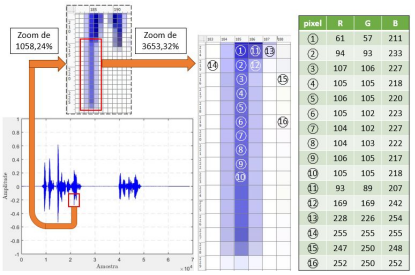
2

[e] file or134

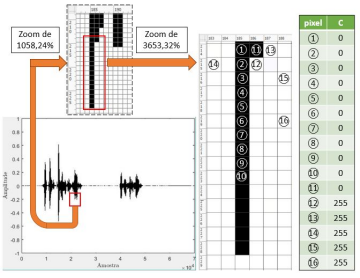


[f] file or137

Figure 3: Pixel classification of images representing the audio signal.



[a] Blue Tinted Pixels



[b] Black and White Pixels

Figure 4: Image Scanning Process.

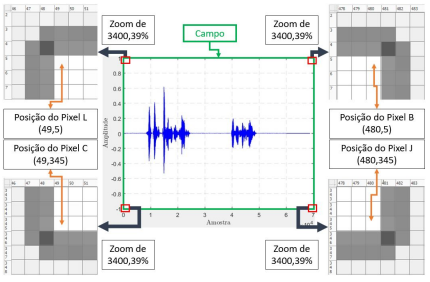


Figure 5: Discrete Functions in 4 Objective Evaluations.

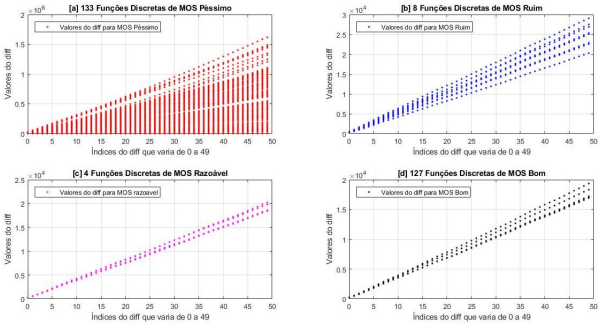


Figure 6: Discrete Function Slopes

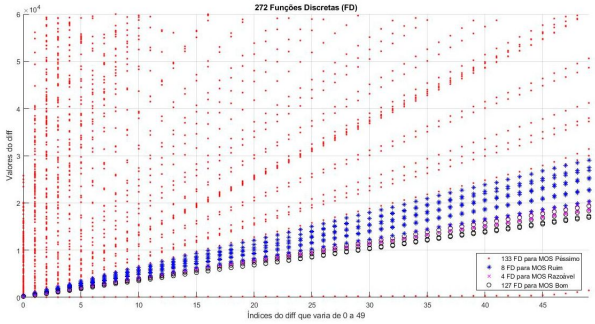
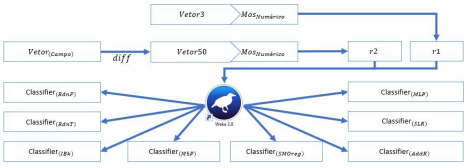
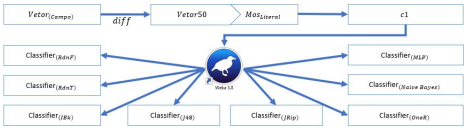


Figure 7: Schema used in regression and classification learning algorithms.



[a] Regression for r1 and r2



[b] Classification for c1

Figure 8: CC behavior for r1

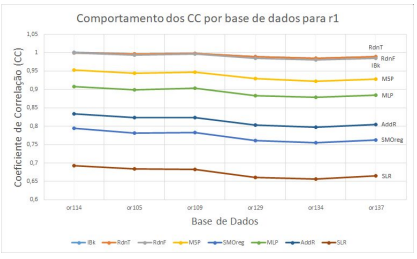


Figure 9: EAM Behavio for r1

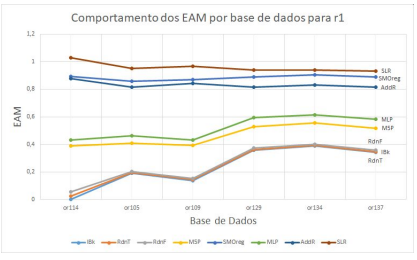


Figure 10: EQM Behavior for r1

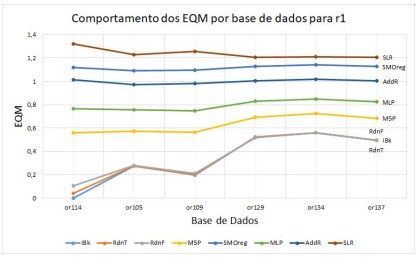


Figure 11: *Software* PredictorFX for r1