26th International Congress of Mechanical Engineering
COBEM 2021
November 22-26, 2021, Virtual Congress, Brazil

COB-2021-1082
QUANTIFICATION OF CLASSROOM DESIGN OVER SPEECH INTELLIGIBILITY INDEX AND REVERBERATION TIME THROUGH DEEP LEARNING

Eriberto Oliveira do Nascimento
Paulo Henrique Trombetta Zannin
Institution Laboratory of Environmental and Industrial Acoustics and Acoustic Comfort, Federal University of Paraná, Curitiba, 81530-000, Brazil
eriberto.on@gmail.com
paulo.zannin@gmail.com

Abstract. *The presence of noise interferes negatively with the teaching practice. Thus, suitable acoustic conditions become a relevant issue in classrooms. This study aims to determine the significance of the following classroom conditions: background noise - (A), sound absorption coefficient - (B), confinement - (C), and occupation - (D) on the Reverberation Time (RT) and Speech Transmission Index (STI). Then, based on measurements in 5 classrooms and its validated simulations in the ODEON 11 software, a response matrix was created, totaling 80 virtual rooms using the Design of Experiments. The quantification of the input variable significance was determined using deep neural networks. The results showed that the higher the RT lower was the STI. Furthermore, the following composition explained the percentual variation of the RT: B, D, and C, while for the STI, the conditions were A, B, D. Therefore, in conclusion, the results obtained agree with the current literature.*

Keywords: *Deep Learning, Speech Transmission Index, Reverberation Time, Input Sensitivity*

1. INTRODUCTION

Noise pollution is becoming one of the most problematic burdens that impair people's health in urban settings (YANG et., 2020). There are many ways in which excessive noise levels can interfere in people's daily lives, ranging from noise impairment loss, stress induced by high levels of noise, and various subsequent health problems, such as chronic headaches, high blood pressure, anxiety, and lack of concentration and much more (NAZNEEN; RAZA; KHAN et al., 2020).

In the current literature, noise brings problems to learning (TABUENCA; BÖRNER; KALZ et al., 2021). Excessive noise exposure in classrooms can induce problems in a multi-dimensional aspect, interfering in the teaching-learning practice in both teachers and students (LE CLERCQ et al., 2020). Moreover, high noise levels are particularly harmful to the learning process for Second Language Students and students who possess learning impairment (LOCHLAND, 2020).

Besides, various factors could interfere with the room's acoustic quality (YANG; MAK, 2021). Specific standards evaluate these factors, e.g., ISO 3322-2 (ISO, 2008) sets the commonly measured acoustical descriptors values, including the Reverberation Time (RT) around 0.60 s in classrooms. However, when measuring speech intelligibly, the Speech Transmission Index (STI) should be applied to the detriment of the RT (LIU et al., 2020). The STI is a complex room acoustic descriptor that simultaneously weights the background noise and the reverberation time, which generates a scalar quantifier value derived from the Modulation Transfer Functions. The STI values range from 0 to 1, with one, meaning 100% speech intelligibly at the receiver's position, and zero, total deterioration of the signal, thus not enabling the intelligible speech communication (IEC 60268–16, 2011).

Many normative standards highlighted threshold values that the classroom should comply with, e.g., STI (ISO, 2003, ANSI, 2010) and RT (ISO, 2008). Nevertheless, what is observed in the classrooms is a complex situation since exogenous factors can interfere in the room dynamics, creating a new set of conditions not covered by the standards. For example, in tropical countries with notably elevated temperatures, the ventilation systems induce noise into the classrooms. As a result, the RT do not change while the STI values decrease (MAPP, 2020, CHOI, 2020).

It is well known that a high sound absorption coefficient lowers the reverberation time and that lower background noise levels increase the Speech Transmission Index. Besides, the presence of students in the classroom also increases the sound absorption area. However, Bluyssen et al. (2020) showed that ceiling panels could not provide a reliable solution to classroom noise problems. So, there are many classroom conditions, as background noise - (A), ceiling panel sound absorption coefficient (B), confinement - (C), and student occupation - (D), that interfere with the Reverberation Time and Speech Transmission Index values. Therefore, this work aims to demonstrated that the Multiple Linear Regression

(MLR) is not enough to determine the most significant input values from the A, B, C, and D conditions as mentioned earlier.

For this reason, the Shapley values and the Permutation Importance Algorithm were used to assess the input feature importance on a Deep Neural Networks (DNN) model. Thus, it allowed fitting the literature result with great accuracy. Nevertheless, this work compares three different methods, e.g., Permutation Importance Algorithm, the Shapley Values, to assess the input sensitivity in DNN models. In addition, the Multiple Linear Regression (MLR) method was used as benchmarking.

This paper is divided into sections. Section 2 covers the definition of the reverberation time, and the speech transmission index, and their respective sensitivity regarding the constructive design factors. Section 3 presents the deep neural network's mathematical modeling and its basic notations. Section 4 defines the input feature significance analysis methods used in the DNN model; Section 5 covers the methodology, reporting the training dataset acquisition and the DNN topology structure. Finally, in Section 6, the results and discussions are presented.

2. ROOM ACOUSTICS PARAMETERS

2.1 Reverberation Time

Quantitatively, the ISO 3382-2 standard (ISO, 2008) calculates the TR or T60, as the time required in seconds [s], for the sound pressure level to decay by 60 dB. This decay is based on the extrapolation of a curve fitting process, applying the Least Squares Method. The linear extrapolation coefficient of the linear fit is calculated in the range of -5 dB below the initial level to -35 dB. This 30dB decay range, from -5dB to -35dB, is called T30.

Tocci and Wilkes (2010) recognize that RT is one of the fundamental parameters of acoustics. They define the RT as the time required for the sound level of excitation to undergo successive reflections and absorptions on the enclosure area surfaces, such as on the walls, ceilings, and floor, until the sound pressure level of the excitation ceases. Therefore, RT is related to the average sound absorption coefficient, and the room volume, given that the number of reflections and their intensity depends on the dimensions of the environment.

$$t_r = \frac{0{,}161V}{S_1\alpha_1 + S_2\alpha_2 + \cdots + S_n\alpha_n} \tag{1}$$

here t_r is the reverberation time [s], S_n are the areas of the room's interior surfaces in [m^2], α_n are the sound absorption coefficients of the room's surfaces, and V is the room's volume [m^3]. The reverberation time is a function of frequency.

2.2 Speech Transmission Index

Historically, the metrics related to the qualification of speech intelligibility stems from the balance of various factors. However, a common feature in the previous studies sought to establish a standard human subjective speech perception curve. That is, it sought to determine a psychometric parameter. Thus, two main groups of descriptors were identified in the literature, the objective and the subjective. Subjective measurements refer to the processes for which standardized questionnaires and dictations are used. These are composed of phonetically balanced words, and the listener is asked to identify the heard word. Then, the percentage of correct answers is consolidated, and intelligibility is taken. Especially in this class of subjective descriptors, the reproducibility of the results is a function of the individuals under analysis, of the physiological, psychological conditions, and the most diverse stressors. Therefore, its application is reduced to more restrictive cases (STEENEKEN; HOUTGAST, 2002). As a result, the STI descriptor, which is an objective method, should be applied.

In this work, the STI was measured using the indirect method based on the room impulse response (RIR), with the Maximum Length Sequence (MLS) excitation signal, at 60 dB level. The indirect STI method (IEC, 2011) considers the degradation between the modulated signal generated at the source with the signal received at the listener's position, through the modulation factor, m, with its respective modulation frequency, f_m, according to Eq. 2,

$$m(f_m) = \frac{1}{\sqrt{1 + \left(\frac{2\pi f_m T}{13{,}8}\right)^2}} \frac{1}{1 + 10^{(-SNR/10)}} \tag{2}$$

here, $m(f_m)$ is the modulation reduction factor, f_m is the modulation frequency [Hz], T is the reverberation time [s], SNR is the signal-to-noise ratio [dB]. The modulation frequencies are given in 1/3 octave bands from 0.63 Hz to 12.5 Hz, totaling 14. These frequencies are: 0.63; 0.80; 1.00; 1.25; 1.60; 2.00; 2.50; 3.15; 4.00; 5.00; 6.3; 8.00; 10.0 and 12.5 Hz, according to IEC 60288:2011. The procedure to calculate STI is explained in IEC 60288:2011.

The modulation transfer function, m, represents the difference in amplitudes of the modulated sine waves calculated by the ratio between the output and input signal. The signal deterioration effect comes from the contribution of the background noise and the room's reverberation time.

3. DEEP NEURAL NETWORKS MODELING

Through a neurophysiological inspiration of the connectionist behavior of the biological neurons, McCulloch and Pitts (1943) proposed a mathematical model of an artificial neuron. This primitive model, although simple, formed the basis for the development of more robust Artificial Neural Networks (ANNs) with sophisticated algorithms, such as the Multiple Perceptron Model and the Deep Neural Networks (DNN). Therefore, ANNs constitute a vast group of consolidated methods for data processing, which with statistics, coined the generic term of Machine Learning. Thus, there are applications of ANNs in fields involving classification, regression, approximation, grouping, prediction, and others.

Mathematically, Amari (2017) presented the Multiple Perceptron Model (MLP) as a composition of multiple nested nonlinear pointwise operations that are propagated over the r = 1, …, L- 1 layers, summarized as $y_L = W_L y_{L-1} + \varepsilon$. Those operations seek to fit an output, y_L, given some input parameters as,

$$y_L = W_L y_{L-1} + \varepsilon \tag{3}$$

here, y_L is the MLP output value, W_L is the weight matrix and ε is an error measure. The estimation of the y_L considers the augmented weight matrices through a process that employs the stochastic gradient descendent learning process (RUMELHART; HINTON; WILLIAMS, 1986). Generally, the operation on Eq. 3 is valid to any hidden layer which the output is given as $y = f(x, W) + \varepsilon$, so, the estimation error is given in Eq. 4,

$$e^* = y^* - f(x, W) \tag{4}$$

where, e^* is the true residual error for the output layer, y^* is the target value, and $f(x, W)$ is a general function representing the MLP model. According to Amari (2017), the error backpropagation algorithm updates the weight matrices, given the following error loss error,

$$l(x, y^*, W) = \frac{1}{2} \left| f(x, W) - f(x, W^*) - \varepsilon \right| \tag{5}$$

where, y^* is the target value, W^* is the optimum value for the weight matrices. Finally, the update is given as,

$$\Delta W_r = -\eta \frac{\partial l}{\partial W_r} = -\eta e_r^* y_{r-1} \tag{6}$$

The fundamental purpose of artificial neural networks is to produce generalizations through a training process. In this context, generalization refers to generating estimates for data not used in the training set. Thus, training is a step in which the weights of connections between neurons are adjusted through a training algorithm (RUMELHART; HINTON; WILLIAMS, 1986).

4. INPUT FEATURE IMPORTANCE ASSESSMENT

Artificial Neural Networks are called black-box or grey models, which means that the relationships between the effects of inputs feature over the output's cannot be directly calculated because there are many contributions over the neural network weights matrices. Therefore, the use of deterministic techniques to calculate sensitivity analysis is an arduous task.

Concerning the gradient method, the input significance applies gradient-based rules to perform a perturbation with infinitesimal change in a specific input and assess its effect on the output (DIMOPOULOS; BOURRET; LEK, 1995). For example, through Jacobian matrices, the original model $y = f(x, W) + \varepsilon = f(x)$, the derivatives are,

$$\frac{dy}{dx^T} = \frac{\partial y}{\partial x} = D_n W_n D_{n-1} W_{n-1} \cdots D_i W_i D_1 W_1 \tag{7}$$

here, D_n are the first partial derivative of the activation function, W_n are the weight matrices, D_n.

There are various methods for estimating input sensitivity in deep neural network models. The most popular ones are the first-order derivatives of the ANN model, Permutation Importance, Sharp Values, Garson Coefficients, Profile Method, and Modified Profile Method (OLDEN; JACKSON, 2002, ZANNIN et al., 2018).

4.1 Permutation Importance Algorithm

The permutation importance is a method that uses the input feature partitioning to determine the explanatory power of each j feature. It is an agnostic method, i.e.; it does not make assumptions of the model under evaluation. Instead, it takes a "corrupted" version of the input sample to estimate their respective reference score, encoded as s (ALTMANN et al., 2010). The reference permutation score is the coefficient of determination (R^2) to regression problems after Eq. 8 is applied.

$$i_j = s - \frac{1}{K}\sum_{k=1}^{K} s_{k,j} \tag{8}$$

here i_j is the importance score for the feature f_j, K is the number of repetitions to get the input data of each j feature. According to Breiman (2001), the permutation algorithm effectiveness depends on the evaluation's original model quality. For example, if the original model is not well-adjusted to the training data, it will reflect a lousy input sensitivity algorithm score.

4.2 Shapley values

The use of Shapley values has derived from the game theory. The Shapley values result from a function that ponders all possible combinations of the input features space. This process tries to organize the input data in batches, so each input feature performs multiple integrations, then the resulting subsets will produce values that contain the specific input variable importance (GALE et al., 1953).

As a result, the Shapley value will result in a marginal distribution of input feature importance, and finally, the overall model performance for a given feature is,

$$\phi_J = \frac{1}{M}\sum_{m=1}^{M} \hat{f}\left(x_{+j}^{m}\right) - \hat{f}\left(x_{-j}^{m}\right) \tag{9}$$

here, ϕ_J is Shapley value, $\hat{f}\left(x_{+j}^{m}\right)$ and $\hat{f}\left(x_{-j}^{m}\right)$ are the random model estimation, with the feature $+j$ ad without $-j$ respectively, $\hat{f}$ is the predictive model, i.e., the DNN model in the context of this work, and m is the number of predictive variables.

According to Štrumbelj and Kononenko (2014), one of the significant advantages of using Shapley values is that it does not make any assumptions regarding the feature distribution. Also, there is no requirement for the data to comply with any specific probability distribution.

4.3 Multiple Linear Regression

The Multiple Linear Regression is one of the most basic models used to assess the input feature significance. The MLR approach fits the training dataset into a hyperplane according to Eq. 10,

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j \tag{10}$$

where y is the predicted value from the input variable x_j and β_j is the regressor predictors weight, which assigns the input feature importance. The regressors were estimated using the Least Square Method, and the β_j coefficient is estimated as

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{11}$$

where, $\hat{\beta}$ are the estimated, X is the design matrix shown in Table 2, X^T is the transpose design matrix, $(\cdot)^{-1}$ is the inverse matrix operator, y is the output vector. Finally, the estimated model is $\hat{y} = X\hat{\beta}$, the residual values are $e = y - \hat{y}$.

5. METHODOLOGY

The methodology was structured in four phases: (i) obtaining the experimental dataset to the training of the deep neural network, (ii) applying the input sensitivity methods on the DNN model to assess input variables, (iii) analyses to compare the DNN results regarding the ranking the input variables significance, in front of the Multiple Linear Regression method. Finally, (iv) the results were contrasted with the literature.

Regarding the training dataset acquisition, this work used the results taken by Nascimento (2019), available in the appendices section tables. Nascimento (2019) used a full factorial planning with two levels, 2^k, the minimum (-1) and maximum (+1) levels, and with k = 4 controllable input factors (MONTGOMERY, 1999). The selected factors were the background noise, represented by a value of the NC curve (ISO, 2016), ceiling panel sound absorption coefficient, the openings, which inform whether the doors and windows are open or closed. Finally, the audience occupancy, indicating whether the classroom contains or not an audience. Table 1 shows the factors.

Table 1. The experimental design setting is taken from Nascimento (2019)

Level	**Natural factors**			
	Noise Curve (NC)	Ceiling panel sound absorption coefficient	Openings	Audience occupancy
Maximum	15	0.10	Open	Full
Minimum	40	0.90	Closes	Empty
	Coded factors			
	A	B	C	D
Maximum	-1	-1	-1	-1
Minimum	+1	+1	+1	+1

The contrast matrix was created, containing all the possible combinations between the controllable factors. For example, for the 2^k, with k = 4, there are 16 combinations. Each combination is called a run and represents a row according to Table 2. In total, 80 virtual classrooms were simulated in the ODEON version 11 software.

Table 2. Results from 5 classroom simulations taken from Do Nascimento (2019)

	Input Features				**RT [s] per classroom**					**STI per Classroom**				
Run	**A**	**B**	**C**	**D**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
1	-1	-1	-1	-1	0.63	0.62	0.72	1.67	1.07	0.71	0.71	0.63	0.51	0.61
2	1	-1	-1	-1	0.63	0.62	0.72	1.67	1.07	0.39	0.5	0.44	0.44	0.43
3	-1	1	-1	-1	0.52	0.23	0.42	0.85	0.55	0.76	0.83	0.77	0.74	0.73
4	1	1	-1	-1	0.52	0.23	0.42	0.85	0.55	0.32	0.45	0.43	0.5	0.42
5	-1	-1	1	-1	0.69	0.92	1.19	1.85	1.58	0.7	0.62	0.53	0.49	0.55
6	1	-1	1	-1	0.69	0.92	1.19	1.85	1.58	0.39	0.48	0.41	0.43	0.41
7	-1	1	1	-1	0.41	0.30	0.53	0.83	0.62	0.76	0.8	0.72	0.74	0.7
8	1	1	1	-1	0.41	0.30	0.53	0.83	0.62	0.32	0.47	0.43	0.5	0.41
9	-1	-1	-1	1	0.57	0.37	0.53	1.01	0.88	0.74	0.8	0.73	0.66	0.65
10	1	-1	-1	1	0.57	0.37	0.53	1.01	0.88	0.42	0.49	0.45	0.5	0.43
11	-1	1	-1	1	0.43	0.27	0.40	0.88	0.52	0.78	0.85	0.8	0.77	0.74
12	1	1	-1	1	0.43	0.27	0.40	0.88	0.52	0.31	0.47	0.41	0.48	0.41
13	-1	-1	1	1	0.58	0.50	0.76	1.07	1.18	0.74	0.75	0.66	0.64	0.59
14	1	-1	1	1	0.58	0.50	0.76	1.07	1.18	0.37	0.5	0.45	0.50	0.41
15	-1	1	1	1	0.41	0.39	0.50	0.81	0.56	0.77	0.48	0.75	0.77	0.71
16	1	1	1	1	0.41	0.39	0.50	0.81	0.56	0.31	0.45	0.42	0.36	0.40

The RT was calculated as the mean values from the octave frequencies of 65 Hz, 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz. The STI values were calculated according to the modulation frequencies of 0.63; 0.80; 1.00; 1.25; 1.60; 2.00; 2.50; 3.15; 4.00; 5.00; 6.3; 8.00; 10.0 and 12.5 Hz, according to IEC 60288:2011.

The Deep Neural Network was implemented through the TensorFlow framework with Kera's backend. The model topology was composed of four densely fully connected layers, i.e., input MLP layer with hyperbolic tangent (tanH) the other with Rectified Linear Unit (ReLU) activation function. Since the training dataset showed in Table 2 was already in the bipolar form, any input feature standardization between -1 and +1 was not applied.

The DNN was applied as a regression problem using the training dataset from Table 2. The input feature, denoted by X, was composed of an input feature sample of $x_{4,1}$ dimension and a target dimension of $y_{1,1}$. Thus, the whole input feature dataset had dimensions of $X_{16,4}$, and the target was $Y_{16,1}$, according to each classroom, and RT and STI values are shown in Table 2. The Keras framework for deep learning was employed (CHOLLET, 2018) to implement the training of the ANN models.

The sequential model from Keras/TensorFlow was used with the ADAM optimizer using the Mean Squared Error (MSE) as the loss function. The scikit-learn (PEDREGOSA et al., 2011) wrapper for the Keras regressor was applied as the backend to implement the DNN model. The training dataset composed 90% of the samples, and the 10% reaming was used as the validation split. The maximum number of epochs was 100, and the batch size was 2. Early stopping callback was used to prevent the overfitting.

Table 3. Deep Neural Networks model summary

Layer	Layer name	Activation	Layer Output Shape	Hyperparameter
1	Fully Connected layer (Dense)	tanH	(None, 32)	160
2	Fully Connected layer (Dense)	ReLu	(None, 32)	1056
3	Fully Connected layer (Dense)	ReLu	(None, 2)	66
4	Fully Connected layer (Dense)	linear	(None, 1)	3

The input sensitivity methods described in sections 4.1 and 4.2 were implemented using the scikit-learn package version 0.24.2. Within the scikit-learn framework, the ELI5 package, version eli5 0.11.0, implemented the built-in Permutation Importance Algorithm described in section 4.1. The SHAP (SHapley Additive exPlanations), package version 0.39.0, implemented the Shapley values algorithm according to section 4.2.

6. RESULTS AND DISCUSSION

The results section was divided into three main sections: (i) descriptive analysis of the dataset used to train the deep neural network model; (ii) deep neural network training results for each of the classrooms evaluated and (iii) the sensitivity analysis for the feature importance assessment and literature comparison.

6.1 Descriptive analysis of the training dataset

Figure 1 shows the boxplot for reverberation time and the Speech Transmission Index. It was shown that the reverberation time varied significantly according to the classroom, while the Speech Transmission Index values presented a minor standard deviation.

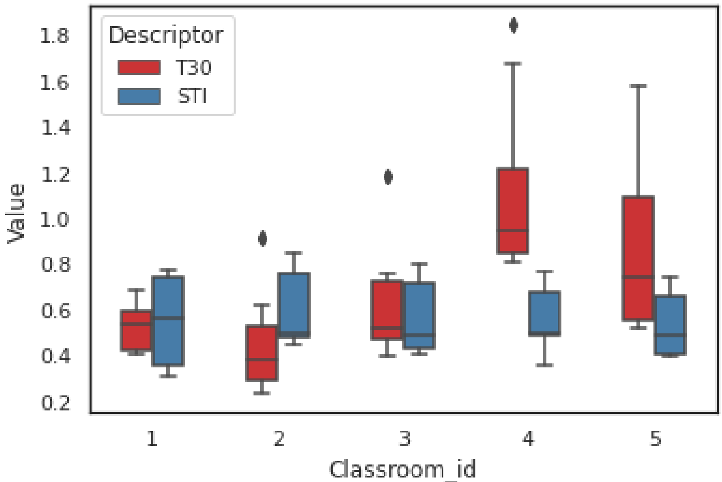

Figure 1. Boxplot of the reverberation time and Speech Transmission Index in all classrooms

The correlations showed that the RT and STI were negatively correlated (R = -0.31 p.value < 0.01). Consequently, low correlation improves the generalization of the deep neural network model since there are no glitches related to multicollinearity or singular values on the training dataset. As a corollary, this is an adequate indicator for employing the methods to assess the sensitivity analysis over the controllable factors.

6.2 Deep Neural Networks training performance

The evaluation of the deep neural network models was straightforward; since the input features were already in the bipolar form, the challenges related to feature standardization and normalization were not presented. Hence, seeking to maximize the generalization and at the same time get good values for the training, the values of the training loss and validation loss were assessed. Therefore, Figure 2 shows the typical pattern of a training process that applied the early stopping method. The early stopping methods stops when the generalization on the validation dataset starts to increase.

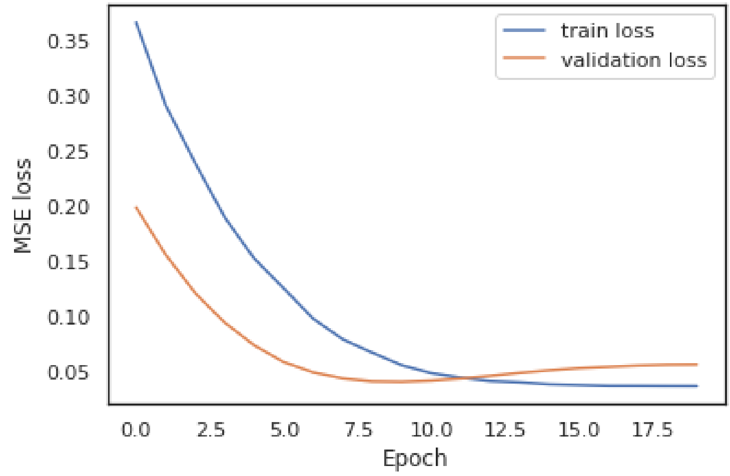

Figure 2. Epochs versus Training Loss Mean Squared Error (MSE)

After the training phase, the accuracy of the models was checked. The first parameter employed was the residual histogram. The residuals are the target values subtracted from the output of the neural network model. With these two values, it was possible to construct a calibration curve and the histogram values. Figure 3 shows the histogram distributions for the residual values and their respective fitted calibration curve. The R-squared of 0.95 validated the model, while the residuals values were small compared to the magnitude of the input feature.

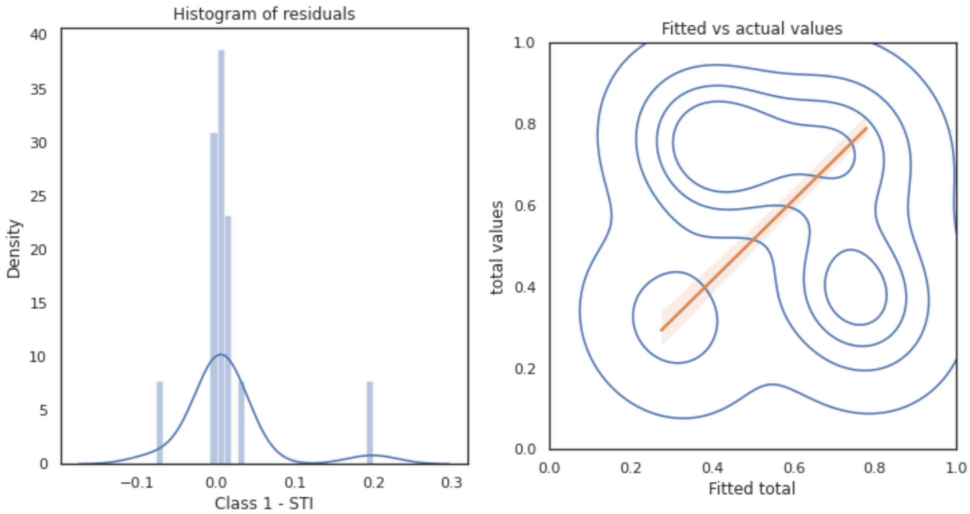

Figure 3. (Left): Histogram of residual values and (Right) the calibration curve of the model

All classrooms and response variables pass to the same training process. After performing the five classrooms' training, the mean value of their R-squared of 0.95. Regarding the residual values, the mean residual value and its standard deviation ratio were about 4%. So, all the models were considered validated.

6.3 Sensitivity Analysis

Montavon, Samek, and Müller (2018) stated that the precision of the input sensitivity methods depends on the DNN model accuracy and the probability distribution of the input features. Nevertheless, when the input features have a high level of nonlinear behavior, performing input sensitivity analysis can resolve into an adequate predictive analysis. Furthermore, as suggested, it is adequate to implement a comparison using some well-defined and well-established approaches in the literature as the analysis regarding the variables' physical.

Although, each method elected in section 4 assesses in a unique approach in the sensitivity of the input feature, e.g., Shapley values, as are generated in a plot where the coefficient varies for each input variable and is assigned a scalar. At the same time, the Permutation Importance Algorithm outputs the maximum magnitude of the entire data set input variable partition. Therefore, Figure 4 shows Shapley values for the STI corresponding to classroom 3.

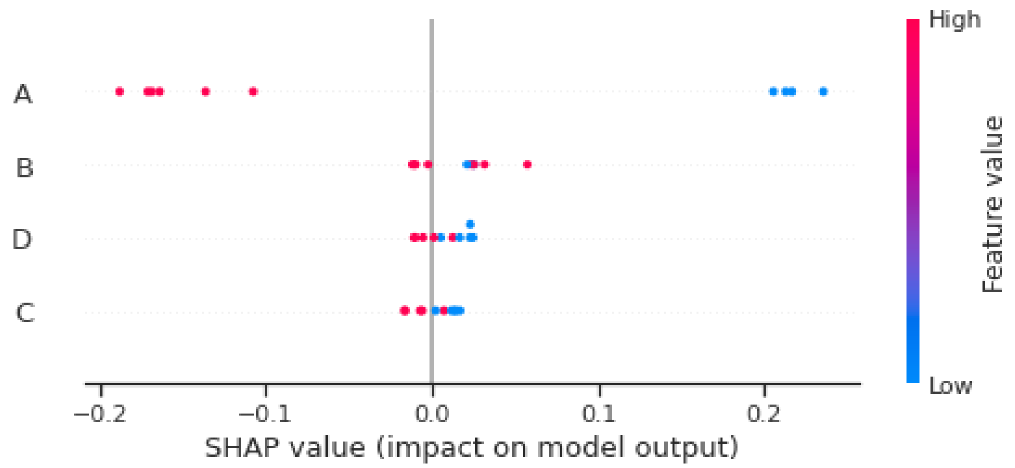

Figure 4. Shapley values for STI in classroom 3

According to Figure 4, the most important input feature was factor A, followed by C, D, and B. Table 4 shows all sensitivity values according to the evaluated classroom and method.

Table 4. Input sensitivity comparison of Permutation Importance, Shapley Value and Multiple Linear Regression

Classroom 1	**Reverberation time**				**Speech Transmission Index**			
Sensitivity Method	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Permutation Importance	0.003	0.026	0.012	0.007	0.072	0.002	0.002	0.001
Shapley Value	0.022	0.152	0.099	0.082	0.271	0.021	0.043	0.020
Multiple Linear Regression	0.000	0.088	-0.008	-0.032	0.196	0.008	0.005	0.005
Classroom 2	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Permutation Importance	0.000	0.069	0.014	0.015	0.032	0.004	0.009	0.008
Shapley Value	0.014	0.210	0.141	0.088	0.207	0.043	0.091	0.077
Multiple Linear Regression	0.010	0.162	0.087	-0.056	0.132	0.002	0.040	-0.010
Sensitivity Method	**Reverberation time**				**Speech Transmission Index**			
Classroom 3	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Permutation Importance	0.000	0.076	0.028	0.015	0.029	0.005	0.001	0.000
Shapley Value	0.044	0.231	0.215	0.117	0.189	0.092	0.021	0.020
Multiple Linear Regression	0.006	0.176	0.121	-0.078	0.139	0.031	0.022	0.015
Sensitivity Method	**Reverberation time**				**Speech Transmission Index**			
Classroom 4	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Permutation Importance	0.000	0.206	0.005	0.067	0.022	0.011	0.003	0.001
Shapley Value	0.059	0.364	0.081	0.221	0.161	0.100	0.015	0.074

Multiple Linear Regression	0.016	0.295	0.033	-0.165	0.107	0.049	0.017	0.014
Sensitivity Method	**Reverberation time**				**Speech Transmission Index**			
Classroom 5	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Permutation Importance	0.000	0.197	0.037	0.013	0.027	0.004	0.000	0.000
Shapley Value	0.047	0.373	0.232	0.112	0.007	0.018	0.063	0.176
Multiple Linear Regression	0.002	0.311	0.118	-0.083	0.125	0.030	0.018	0.002

In Table 4, it was possible to find that the input features A, B, C, and D presented a uniform rank for all three input sensitivity methods for reverberation time and speech transmission index. Furthermore, applying ANOVA analysis for differences in the means, with significance level $\alpha = 0.05$, it was possible to show that the permutation importance and the Shapley values for the input sensitivity analysis methods were equivalent.

Nevertheless, regarding the Multiple Linear Regression results, its mean values to the input sensitivity ranking were dissimilar to the Shapley values method and the Permutation Importance Algorithm. Therefore, one may ascertain that the difference in the rankings between the methods was due to the lower predictive power of the MLR method, thus generating an untrustworthy model. For example, the R^2 value for the MLR was about 0.5, with a residual value not complying with the normal distribution and presented local distortions in the calibration curve, thus resulting in heteroscedasticity.

In conclusion, both the Input Permutation Method and the Shapley values agreed with the literature. Aimed at, when analyzing the reverberation time, it is well known that the sound absorption coefficient of the ceilings of the audience will drastically reduce the number of reflections on the environment, which can directly decrease the reverberation time in that same room (BERANEK, 2006; PASSERO; ZANNIN, 2010). While working with STI values, this analysis is not straightforward since the reverberation time is accounted simultaneously with the background noise in the classroom. However, when analyzing the results from Table 4, the background noise was the leading factor in the results while the sound absolution was not so important. Both results agree with the literature (RENNIES et al., 2014, TANG, YEUNG, 2006)

7. CONCLUSIONS

This work aimed to evaluate the significance analysis of the sensitivity analysis of the Reverberation Time and Speech Transmission Index using deep learning models. The results showed the comparison between two methods using deep learning and MLR method as benchmarking. The results showed that the Permutation Method and Shapley values both agreed with the literature. While the multiple linear regression method was not suitable for assessing significance because it did not present a great value for the Pearson linear correlation coefficient, thus not enabling reliable results.

8. ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

9. REFERENCES

ALTMANN, André et al. Permutation importance: a corrected feature importance measure. Bioinformatics, v. 26, n. 10, p. 1340-1347, 2010.

AMARI, Shun-ichi et al. Dynamics of learning in MLP: Natural gradient and singularity revisited. Neural computation, v. 30, n. 1, p. 1-33, 2017.

AMERICAN NATIONAL STANDARD INSTITUTE (ANSI). ANSI/ASA S12.60-2010/PART 1: Acoustical Performance criteria, Design Requirements, and guidelines for Schools, Part 1: Permanent Schools. Acoustic Society of America, 2010.

BERANEK, Leo L. Analysis of Sabine and Eyring equations and their application to concert hall audience and chair absorption. The Journal of the Acoustical Society of America, v. 120, n. 3, p. 1399-1410, 2006.

BREIMAN, Leo. Random forests. Machine learning, v. 45, n. 1, p. 5-32, 2001.

CHOLLET, F. al. Keras: The python deep learning library. ascl, p. ascl: 1806.022, 2018.

CHOI, Young-Ji. Evaluation of acoustical conditions for speech communication in active university classrooms. Applied Acoustics, v. 159, p. 107089, 2020.

DIMOPOULOS, Yannis; BOURRET, Paul; LEK, Sovan. Use of some sensitivity criteria for choosing networks with good generalization ability. Neural Processing Letters, v. 2, n. 6, p. 1-4, 1995.

DO NASCIMENTO, E. O. Influência dos parâmetros acústicos in situ na avaliação da inteligibilidade da fala em salas de aula via redes neurais artificiais. Dissertação em Engenharia Mecânica. 2019.

GALE, David. A theory of n-person games with perfect information. Proceedings of the National Academy of Sciences of the United States of America, v. 39, n. 6, p. 496, 1953.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 1996-1: Acoustics — Description, measurement and assessment of environmental noise — Part 1: Basic quantities and assessment procedures. Switzerland, 2016.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 3382-2: Acoustics -- Measurement of room acoustic parameters -- Part 2: Reverberation time in ordinary rooms, Switzerland, 2008.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 9921: Ergonomics – Assessment of speech communication. Switzerland, 2003.

LE CLERCQ, Carlijn MP et al. Association of slight to mild hearing loss with behavioral problems and school performance in children. JAMA Otolaryngology–Head & Neck Surgery, v. 146, n. 2, p. 113-120, 2020.

LIU, Hongshan et al. The speech intelligibility and applicability of the speech transmission index in large spaces. Applied Acoustics, v. 167, p. 107400, 2020.

LOCHLAND, P. Intelligibility of L2 speech in ELF. Australian Journal of Applied Linguistics, v. 3, n. 3, p. 196-212, 2020.

MAPP, Peter. Speech Transmission Index (STI): Measurement and Prediction Uncertainty. Uncertainty in Acoustics: Measurement, Prediction and Assessment, p. 217, 2020.

MCCULLOCH, Warren S.; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, v. 5, n. 4, p. 115-133, 1943.

MONTAVON, Grégoire; SAMEK, Wojciech; MÜLLER, Klaus-Robert. Methods for interpreting and understanding deep neural networks. Digital Signal Processing, v. 73, p. 1-15, 2018.

MONTGOMERY, Douglas C. Experimental design for product and process design and development. Journal of the Royal Statistical Society: Series D (The Statistician), v. 48, n. 2, p. 159-177, 1999.

NAZNEEN, Shahla; RAZA, Ali; KHAN, Sardar. Assessment of noise pollution and associated subjective health complaints and psychological symptoms: analysis through structure equation model. Environmental Science and Pollution Research, v. 27, n. 17, p. 21570-21580, 2020.

OLDEN, Julian D.; JACKSON, Donald A. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecological modelling, v. 154, n. 1-2, p. 135-150, 2002.

PASSERO, C. R. M.; ZANNIN, P. H. T. Statistical comparison of reverberation times measured by the integrated impulse response and interrupted noise methods, computationally simulated with ODEON software, and calculated by Sabine, Eyring and Arau-Puchades' formulas. Applied Acoustics, v. 71, n. 12, p. 1204-1210, 2010.

PEDREGOSA, Fabian et al. Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research, v. 12, p. 2825-2830, 2011.

RENNIES, J et al. Listening effort and speech intelligibility in listening situations affected by noise and reverberation. The Journal of the Acoustical Society of America, v. 136, n. 5, p. 2642-2653, 2014.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning representations by back-propagating errors. Nature, v. 323, n. 6088, p. 533-536, 1986.

STEENEKEN, H. J. M; HOUTGAST, T. Phoneme-group specific octave-band weights in predicting speech intelligibility. Speech Communication, v. 38, n. 3-4, p. 399-411, 2002.

ŠTRUMBELJ, Erik; KONONENKO, Igor. Explaining prediction models and individual predictions with feature contributions. Knowledge and information systems, v. 41, n. 3, p. 647-665, 2014.

TABUENCA, Bernardo; BÖRNER, Dirk; KALZ, Marco. Effects of an ambient learning display on noise levels and perceived learning in a secondary school. IEEE Transactions on Learning Technologies, v. 14, n. 1, p. 69-80, 2021.

TANG, Shiu Keung; YEUNG, M. H. Reverberation times and speech transmission indices in classrooms. Journal of Sound and Vibration, v. 294, n. 3, p. 596-607, 2006.

YANG, Da; MAK, Cheuk Ming. Effects of acoustical descriptors on speech intelligibility in Hong Kong classrooms. Applied Acoustics, v. 171, p. 107678, 2021.

YANG, Weijun et al. Evaluation of urban traffic noise pollution based on noise maps. Transportation Research Part D: Transport and Environment, v. 87, p. 102516, 2020.

ZANNIN, P. H. T. et al. Application of artificial neural networks for noise barrier optimization. Environments, v. 5, n. 12, p. 135, 2018.

10. RESPONSIBILITY NOTICE

The authors are the only ones responsible for the printed material included in this paper.