Cognitive and Behavioral Data for Decision Tree-based Diagnosis of Attention-Deficit/Hyperactivity Disorder

Anderson Martins Silva
Graduate Program in Electrical and Computer Engineering
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: 71953371@mackenzista.com.br

Luiz Renato Rodrigues Carreiro
Graduate Program in Developmental Disorders
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: luizrenato.carreiro@mackenzie.br

Mayara Miyahara Moraes Silva
Graduate Program in Developmental Disorders
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: 72054298@mackenzista.com.br

Maria Cristina Triguero Veloz Teixeira
Graduate Program in Developmental Disorders
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: mcris@mackenzie.br

Leandro Augusto da Silva
Graduate Program in Electrical and Computer Engineering
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: leandroaugusto.silva@mackenzie.br

***Abstract*—Attention-Deficit/Hyperactivity Disorder (ADHD) presents in children and adolescents as a persistent pattern of inattention, hyperactivity, and impulsivity that interferes with their development. Computational studies on ADHD focus on measures of brain activity of the participants and a few use standardized cognitive tests or behavioral inventories to assess objective indicators for diagnosis. The paper presents a computational proposal in which the combination of two artificial intelligence methods is used to aid the identification of diagnostic indicators for ADHD. The proposal is to combine a neural network of self-organizing maps to group factors from standardized tests and inventories, and a decision tree to classify the most relevant factors. The study included 127 children and adolescents from 6 to 16 years old, 48 with ADHD diagnosis and 79 without ADHD (control group). The most relevant result of the study was the strong contribution of the Child and Adolescent Behavior Inventory results in the diagnosis of the disorder with great performance in prediction when compared to real data and reliability by Kappa statistics.**

Keywords—Self-Organizing Maps (SOM); Decision Tree; Attention Deficit/Hyperactivity Disorder (ADHD).

I. INTRODUCTION

This work is complementary to and based on the published article BRAININFO 2021 [1] and according to the Diagnostic and Statistical Manual of Mental Disorders, 5th edition - DSM-5 [2]. Attention-Deficit/Hyperactivity Disorder (ADHD) is a persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development. The disorder is characterized by inattention involving, for example, difficulty sustaining attention in tasks or playing activities, a state in which the mind seems elsewhere, even in the absence of any obvious distraction, difficulty to follow through with instructions and failing to finish schoolwork, often forgetful in daily activities, chores, or duties in the workplace, losing things, expressing excessive activity or restlessness, and inability to wait one's turn, always in ways that are excessive for one's age or developmental level. ADHD has its initial expressions in childhood and usually persists into adulthood, resulting in impairments in social, academic, and occupational functioning.

The diagnosis of ADHD is clinical, based on the individual's history and expression of symptoms. Because this diagnosis is often based on reports of symptom severity and because these symptoms are also part of other clinical conditions, the diagnostic difficulty is present in the daily lives of the interdisciplinary teams responsible for the evaluation process [3] [4]. Because of the complexity of the diagnostic evaluation, the American Association of Pediatrics recommends the use of an algorithm, both for evaluation and treatment of children and adolescents with ADHD [5]. To support clinical decision making, neuropsychological, behavioral, and adaptive functioning assessment procedures have often been used in conjunction with neurological assessments [6]. Considering the social importance involved in properly issuing a correct diagnosis of ADHD, in both children and adolescents, studies must be proposed that discuss which are the best indicators of clinical-neurological, neuropsychological, and behavioral-adaptive diagnostic evaluation when children and adolescents present with complaints of inattention and hyperactivity. Furthermore, for appropriate assessments and interventions to be implemented, differential criteria are needed to correctly characterize and identify attention-deficit/hyperactivity among children and adolescents. Comprehensive assessments in this regard allow a better understanding of the complexity of each case for appropriate guidance, design of the therapeutic intervention, and evaluation of the need for educational and emotional support for patients and families [6].

Computational studies can help professionals in diagnostic assessments, especially using machine learning algorithms, Kam et al. [8] used an artificial intelligence algorithm called decision tree for screening ADHD, by monitoring the school activities of 153 children using 3-axial actigraph and obtained results consistent with previous studies. In turn, Lee et al. [9] analyzed the classification of ADHD in children through brain activity measurements. In their work, they used a neural network algorithm called self-organizing maps allowing categorizing characteristics of children with and without clinical indicators of ADHD.

Unlike previous proposals presented in the literature, this work aims to combine two artificial intelligence techniques. In the first step, standardized test results are grouped by means of Self-Organizing Maps (SOM) and, in a second step, the groups with a high level of overlap are analyzed using a decision tree algorithm. This helps discover which attribute is discriminative in the diagnosis of children and adolescents with suspected ADHD.

Besides Section I, that aims to contextualize the work and present the objective, the work is organized into six parts. Section II presents the theoretical framework and justification of the study. Section III presents the proposed use of two artificial intelligence algorithms to aid in the diagnosis. In Section IV, the procedures for developing the study are described, including the computational development with the application of two artificial intelligence techniques. In Section V, the contribution of standardized cognitive tests or behavioral inventories is described, as well as the proposal to solve the diagnostic doubt within the self-organizing maps and then the classification by the decision tree for understanding the characteristics of the diagnosis of the disorder. Finally, in Section VI, we present the conclusion and recommendations for further studies.

II. Related Work

A. Elements of Attention Deficit/Hyperactivity Disorder (ADHD)

ADHD is part of the group of neurodevelopmental disorders beginning in childhood, but a substantial proportion of children with ADHD remain relatively impaired into adulthood [9]. From a cognitive-behavioral point of view, it is characterized by deficits in several cognitive functions, such as attention, especially selective, sustained, alternating, and divided attention, deficits in inhibitory control, processing speed, organization, ability to inhibit distracting information, deficits in cognitive flexibility, hyperactivity behaviors, restlessness, and impulsivity. ADHD affects 5.29% of the world's child population. Of this population, 30% up to 70% maintain symptoms into adulthood [10] [11]. According to DSM-5 [1], ADHD can be classified according to the predominance of symptomatic axes as predominantly inattentive, predominantly hyperactive-impulsive, or combined presentations.

Behavioral patterns are important in the diagnosis of ADHD. Here are some difficulties related by parents regarding the children: listening, obeying, following routine rules, often postponing and forgetting daily activities, following direct instructions, regulating feelings of frustration, exacerbation of motor activity, maybe impulsive in changing activities before they are completed, having difficulty waiting their turn, may have impairments in social relationships. These behaviors may contribute to high-stress in family or school environments [12].

The inequality of symptom axes within the predominantly inattentive subgroup compromises its validity when compared to the combined subgroup. However, individuals considered only inattentive, but with a subclinical level of hyperactivity-impulsivity symptoms (4 or 5 symptoms), have their classification without intensity of the combined ADHD subtype. However, there is still little evidence on the qualitative difference between the subtypes presented, even when inattentive classification is established to individuals with three or fewer hyperactivity-impulsivity symptoms [13]. At different stages of life, ADHD can show itself at about 2:1 in the case of children and 1.6:1 in adults. This shift in prevalence from children to adults hypothetically occurs because children and adolescents can create methods that suppress the disorder as they develop, making them more functional, and hiding some difficulties and symptoms [9].

According to the new classification of the A.P.A [2], the highlights are the changes in the various forms of symptoms, trying to contextualize the criteria diagnosed throughout the individual's life; change in the age of onset of symptoms, from 7 to 12 years of age; change of the term "subtype" for "current presentation"; and removal of autistic spectrum disorders as excluding factors in the diagnosis.

Children who are usually in the preschool phase may present high levels of motor activity, attention deficit, and poor inhibitory control that are behavioral manifestations of ADHD. However, in clinical cases, they are more significant and result in considerable impairment, with high accident rates and poor school performance that can persist into school age in 60 to 80 percent of cases [14]. Approximately 70% of school-age children experience worsening in school activities and impairment in family life and relationships with other children. Generally, inattention symptoms relative to hyperactivity symptoms decelerate with age, i.e., decline slowly between the passage of the child, adolescent, and adult age stages with greater persistence [15]. About 33% of children with ADHD in adulthood no longer have the symptoms, as opposed to the rest who continue to have the disorder or episodes that result in loss. Over a prolonged period, adults with ADHD experience worsening academic and work performances, as well as increased traffic offenses, motor vehicle accidents, and sexual behavior with high risk and the concomitant onset of various psychiatric illnesses, such as anxiety disorder, mood disorders, and substance abuse [6].

Genetic factors and environmental conditions contribute greatly to a complex etiology of ADHD with neurobiological bases established by research. For example, study conducted by Williams [16] points to unusual patterns in the central nervous system of people with ADHD [17]. According to Carreiro et al. [6], the study of neuropsychological endophenotypes is motivated by the etiological complexity and clinical inhomogeneity of ADHD. Endophenotypes, sometimes called ”intermediate phenotypes”, are inherited and gauged traits that can be found in the pathway linking a genotype to complex neuropsychiatric disorders. However, individuals with ADHD may or may not display different forms of patterns following a complex cognitive profile and multiple variations, such as loss or low attentional focus, causing a lack of flexibility, difficulty dealing with distractors, difficulty with self-regulation, behavioral impulsivity, lack of motor coordination, and shuffling of mental information due to difficulty with organization. The signs presented show that the child may have an impairment in the intellectual development in different areas of the brain and are shown by playful interactions, by observation, and by standardization of instruments, becoming fundamental in measuring the complaints of inattention and hyperactivity [18].

Behavioral patterns are important in diagnosing ADHD, parents report that children have difficulty listening for inattention when someone speaks directly, obey, and follow rules and routines, often postpone and forget daily activities, have difficulty following direct instructions, do not accept frustration, have an exacerbation of motor activity, may show some form of impulsivity in an activity before it is completed, are unable to wait, have impaired social relationships, have high-stress in the family or school environments. Observation of several people can increase the accuracy of the diagnosis, enabling in cases of comorbidities and inconsistent symptoms, the differentiated possibility of diagnosis. This creates the need to report in a standardized way, to avoid bias and the possibility of building a representative behavioral profile for evaluation. Currently, the psychology literature has several instruments that are applied to parents and teachers, to extract as much information as possible from the two environments of children [18]. These instruments, which are seen in this work as attributes, have a protocol-based direction to assess ADHD complaints, and the attributes are analyzed by cognitive assessment parameters used, such as the Cancel Attention Test, Trail Making Test, Continuous Performance Test, Wisconsin Letter Test, Wechsler Abbreviated Intelligence Scale and Wechsler Intelligence Scale for Children, as well as the behavior inventories for children and adolescents (CBCL/6-18) and for teachers (TRF/6-18).

Given the importance of collecting various pieces of information in cognitive neuropsychology and behavior analysis, the treatment and multivariate analysis of the data can help us obtain relevant information in understanding ADHD complaints, and the artificial intelligence techniques used become key elements in diagnostic discrimination.

B. Self-Organizing Maps (SOM)

According to Merényi et al. [19], a SOM network provides clustering and visual representation of data in low dimensions. This technique preserves the topological structure of the data in a lattice of neurons. The grid can be defined as a rectangular or hexagonal grid, as in Figure 1, usually two-dimensional, in an ordered manner such that the most similar neurons are grouped with neurons that are close in the grid, and the opposite is true for less similar neurons that are far apart in the grid, providing a topological view of the data. All neurons in the grid must undergo exposure to different realizations of the input dataset to ensure that the self-organization process matures. The algorithm then proceeds to choose synaptic weights initially randomly with small values. Once the grid has been initialized, we have the presence of three essential processes used to construct the self-organizing map.

With the need to understand the characteristics of the combined attributes and facing data with non-linear distribution, it was chosen in this work an unsupervised model, developed by Kohonen in 1982, called SOM, being especially suitable for data assimilation because it has visualization properties such as highlighting [20], but the dataset allowed the use of principal component analysis or cluster analysis of the data. The goal of unsupervised methods is to identify clusters in unlabeled sets of data vectors sharing similarities. This helps to build a cognitive model that realizes the interrelationship of the data [21]. As proposed by Kohonen [22], capturing the features of an input data set, with a nonlinear distribution, is effected by building a complex neural network around a one- or two-dimensional grid of neurons. The ordering of the input data is structured by weight vectors of neurons called prototypes, and is inspired by neurobiology. They were summarized by Kohonen [22] and Kubat [23] as follows:

1) Competition: The generalist model can serve to describe the phenomena of the initial data in a way that collectively orders a complex system, which shows representative statistical features, even with the fully disordered input space [23].

However, the SOM model is automatically associated with the nodes of the rectangular or hexagonal grid, Figure 1, usually two-dimensional, in an ordered fashion and in a way to which the most similar models join closer nodes in the grid. The opposite being true for less similar models that are equidistant, thus this process enables a topological view of the data [24].

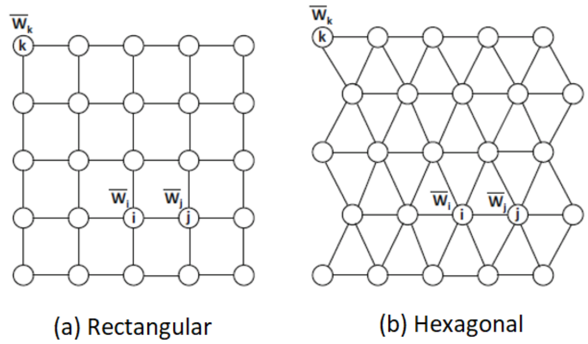

Figure 1. Topological map - rectangular and hexagonal grids ([25, p.451])

The synaptic weight vector is calculated for each *j* neuron of the grid with the same dimension as the input dataset through the inner product between the synaptic weight vector and the input data vector, this function being the basis for choosing the winning neuron. The maximization of this function has mathematical equivalence with the minimization of the Euclidean distance between the synaptic weight and input data vectors.

$$X = [x_1, x_2, ..., x_m]^T \tag{1}$$

X is the input vector of the space *m* transposed.

$$W_j = [w_{j1}, w_{j2}, ..., w_{jm}]^T, \; j = 1, 2, ..., l \tag{2}$$

W_j is the synaptic weight vector of each neuron in the grid. In the competition step, according to Kubat [23], an input vector is randomly selected along with the synaptic weight vector of the neuron *j*, that by making the inner product of the two vectors, being $W_j^T X$ for *j* = 1, 2,..., *l*. Transposing the input vector it is possible to make the selection of the largest product. The topological neighborhood of the excited neurons is centered and since maximizing the inner product of the vectors X and W_j is minimizing their Euclidean distance, by creating an index *i*(X) to identify the neuron that best relates to the input vector X, one can define *i*(X) by:

$$i(X) = \text{arg } \min_j \left\| X - W_j \right\|, \; j = 1, 2, ..., l \tag{3}$$

i(X) is the index that summarizes the competitive process between neurons.

This process summarizes the competition between neurons. Equation (3) shows the *i*(X) is the goal of this process because in this step the identity of the neuron *i* is important and the neuron that satisfies the condition is called the winning neuron for the input vector X.

2) *Cooperation:* The cooperation process starts when the winning neuron is updated around a topological neighborhood of the nearest neurons, being similar around a radius *r*. However, it is necessary to define the topological neighborhood so that only adjacent neurons are updated while having a way that the neighborhood decays smoothly with lateral distance [23].

Given $h_{j,i}$ the topological neighborhood centered around the winning neuron *i* that contains a set of excited neurons, one neuron of this set being represented by *j* and $d_{j,i}$ the lateral distance of the excited neuron *j* by the winning neuron *i*, then one can assume that the topological neighborhood. Equation (4) is a unimodal function of the distance $d_{j,i}$, provided it satisfies the symmetry conditions with respect to the maximum point with $d_{j,i}$=0, and the amplitude decreases with increasing lateral distance $d_{j,i}$ [23].

The basis for cooperation between neighboring neurons is provided by the winner neuron that shows the spatial location of the topological neighborhood of neurons neighboring the winner $h_{j,i(X)}$:

$$h_{j,i(X)} = \exp\left(-\frac{d^2{}_{j,i}}{2\sigma^2}\right) \tag{4}$$

$d_{j,i}$ is side distance and σ is the effective width of the topological neighborhood.

Since the topological neighborhood has some dependence with the lateral distance, as seen in Equation (4), then a one-dimensional grid $d_{j,i}$ is an integer equal to $|j - i|$. However, when it is two-dimensional it is defined by Equation (5) where the discrete vector $\mathbf{r}_j$ is the position of the excited neuron *j* and $\mathbf{r}_i$ is the position of the winning neuron *i*, both being measured in the discrete output space. An interesting feature of SOM is that the size of the topological neighborhood decreases with time and this is done by having the width σ of Equation (4) decrease with time [26].

$$d^2_{j,i} = \left\| \mathbf{r}_j - \mathbf{r}_i \right\|^2 \tag{5}$$

$$\sigma(n) = \sigma_0 \exp\left(-\frac{n}{\tau_l}\right) \; n = 0,\ 1,\ 2,\ ..., \tag{6}$$

σ_0 is the value of σ when starting the SOM and τ_l a time constant. Thus, we have the topological neighborhood defined as a time variable according to Equation (7):

$$h_{j,i(X)}(n) = \exp\left(-\frac{d^2{}_{j,i}}{2\sigma^2(n)}\right) \; n = 0,\ 1,\ 2,\ ..., \tag{7}$$

So one can conclude that when the time *n* increases, the topological neighborhood decreases exponentially, as does the width $\sigma(n)$. This topological neighborhood $h_{j,i(X)}(n)$ was used in the study and will be reference from this point on.

3) Adaptation: In the adaptation phase, for the grid to be self-organizing, the synaptic weight vector W_j of the neuron j of the grid must be changed relative to the input vector $\mathbf{X}$ [23]. However, there is a problem of saturation of the weights at the end of the process, but it can be corrected by changing the Hebbian assumption with a forgetting term $g(y_i)\mathbf{W}_j$ where $\mathbf{W}_j$ is the synaptic weight of the neuron j and $g(y_i)$ is the positive scalar function of the response y_i, the constant term of the Taylor series expansion of function $g(y_i)$ being zero [23].

Neighboring neurons to the winner increase their discriminant function values based on the input dataset, and as appropriate adjustments applied to their synaptic weights improve a subsequent input dataset. As Kohonen [22], the Equation (8) of updating is defined by:

$$W_j(n+1) = W_j(n) + \eta(n)h_{j,i(X)}(n)(X - W_j(n)) \quad (8)$$

n equals epoch, $\eta(n)$ is the learning rate, and $h_{i,j(x)}(n)$ is the neighborhood function.

According to Kohonen [22] there are two phases in the adaptive process: a sorting phase, which organizes the topology of the weight vectors, and a convergence phase, which adjusts the feature map, producing a statistical quantization of the input space. In the sorting phase, the learning rate parameter $eta(n)$ starts with a value of 0.1 and decreases to a value close to 0.01. The neighborhood function must contain almost all the neurons of the grid around the winning neuron *i* with a slow reduction over time and may require 1000 or more iterations. In the convergence phase the number of iterations should be at least 500 times the number of neurons in the lattice, and for good statistical accuracy, the learning rate parameter $eta(n)$ should be close to 0.01 and never zero. The neighborhood function has only the nearest neighbors of the winning neuron, which can be either one or zero neighboring neurons [23].

The originality of the SOM learning presented in the study may contribute to improving the diagnosis of ADHD patients by presenting a set of similar prototypes that represent the combination of attributes and then using a classification algorithm, such as the decision tree, which directly and quickly separates patients prone to the disorder. With this, the study can help health professionals and researchers to develop more accurate tools and reduce the cost of diagnostic tests that currently make their application unfeasible in the public health system.

C. Decision Tree

A decision tree is an Artificial Intelligence algorithm capable of organizing attributes from a dataset in priority, so that it can generate a path that leads to a decision for a classificatory attribute [27] [28]. A decision tree is built using algorithms that variously split a data set into branch-like segments. These segments create an inverted decision tree beginning at a root node at the top of the tree to the leaf at its end. Each object in the study is reflected in the root node and is a simple, one-dimensional display in the decision tree view. The name of the attribute is displayed along with a spread of values that are contained in the attribute. Its display shows all records in the dataset, attributes, and their values are viewed on the analysis object. The development of the decision rule for forming branches under the root node has, in the extraction method, a relationship between the object of analysis and one or more attributes that are used as input attributes to create the branches or segments, being used to estimate the likely value of the target, or outcome attribute or dependent attribute [29].

According to Castro [30], the structure of the decision tree is composed of internal nodes that correspond to an attribute test, each branch represents the test result, and the classes or class distributions are represented by the leaf nodes. The start node represents the root node, and the path from this node to the leaf node is called the classification rule. The decision tree construction can be used to classify an unknown class object, and the estimation is done by testing the attribute values in the tree and traversing until it reaches the leaf node. In this way, the result of the decision tree algorithm becomes a process that is easy to understand and visualize.

The decision tree algorithm as presented by Linoff and Berry [31] is a hierarchical collection of rules that describes how to divide a large collection of objects into successively smaller groups of objects. With each successive split, the members of the resulting followings become more similar to each other. For the selection of attributes to be chosen for division, one can use the information gain (*Gain*), which is one of the best-known methods and has as its generating base the entropy that is a measure of purity [32]. Thus, the gain is the expected reduction of entropy and has as its main function the division of attributes in the data set. Shannon entropy, on the other hand, measures the purity of the dataset [33], being a measure of the heterogeneity of the input data (*S*) relative to its classification (*m*). The expected value of the entropy of the attribute *E(A)* is given by Equation (9) and the expected information of S is given by Equation (11). Thus, the key factor is the use of a gain function that allows the attributes (A) to be compared to select the most relevant one. The chosen attribute is the one that maximizes the information gain which is calculated as being [30] [32]:

With n_{ij} being the number of objects in class C_i in a subset S_j, then the expected attribute information can be:

$$E(A) = \sum_{j=1}^{v} \frac{n_{ij} + ... + n_m j}{n} * I(n_{ij}, ..., n_{mj}) \\ = \sum_{v \in values(A)} p(A_v) Entropy(A_v) \quad (9)$$

The entropy (Shannon's) [33] measures the impurity of the dataset, being a measure of the heterogeneity of the input dataset (S) relative to its classification (c). The *Gain(S,A)* is given by the Equation (10) and the entropy of *S* is given by the equation 11. Thus, the key factor is the use of a gain function that allows the attributes (A) to be compared to select the most relevant one. The attribute chosen is the one that maximizes the information gain which is calculated as being [33]:

$$Gain(S, A) = I(S) - E(A) \quad (10)$$

S is the input dataset, *A* is the attributes, and Θ represents the probability of *A* multiplied by its entropy.

$$I(S) = I(C_1, C_2, ..., C_m) = \sum_{k=1}^{m} -p_i log_2 p_i \quad (11)$$

The information Gain is given by the Equation (10) and represents the expected reduction in entropy when the value of the attribute *A* is known, since the process calculates the gain for each attribute, choosing the attribute with the highest gain to be tested in the set *S*. This process creates the division of objects to form the decision tree, giving rise to the node, labeling the attribute, and creating branches for each attribute value.

III. Proposed Method

The work presents a proposal for an unsupervised learning model as a method used in the identification of the neurons of the grid with greater diagnostic doubt of ADHD, that is, the diagnostic doubt in the neuron shows that it is difficult for both a machine learning algorithm and an expert to make a diagnosis. Thus, the paper brings a proposal to apply a decision tree on the neurons that show overlap to suggest which attributes are more discriminative. To understand this overlapping, the entropy (of Shannon) was calculated with the purpose of measuring the impurity of the neuron with its dataset, that is, the closer the entropy is to one, the greater the impurity of the neuron's dataset. Given this fact, a combination of SOM with the decision tree algorithm, which is a supervised model used in data classification to help identify one or more attributes from standardized assessment tools, such as cognitive tests and behavioral assessment inventories were sought. These tools were used to test the learning of ADHD characteristics.

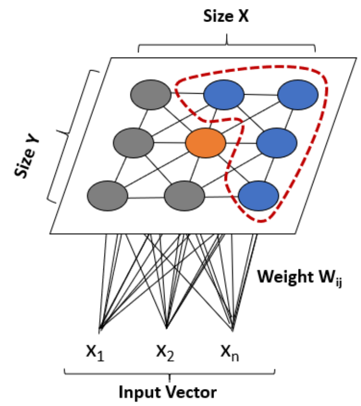

Figure 2. Segmentation of neurons in the self-organizing map

The objective of this decision tree algorithm was to verify the accuracy of the model for the confirmation of cases with ADHD diagnosis by identifying which assessment tools best contributed to this ADHD confirmation.

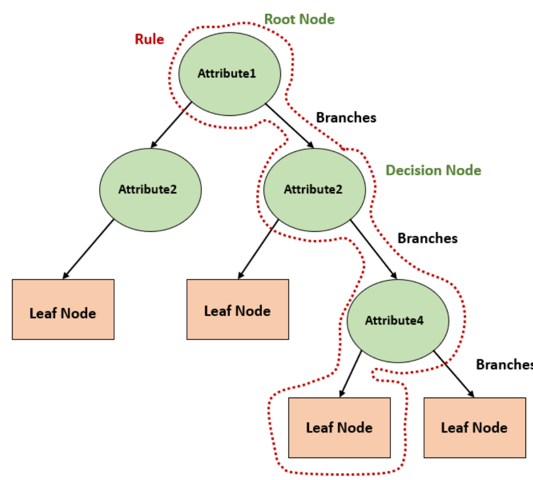

Figure 3. Decision tree structure

Then Kappa statistics is applied to measure the agreement and reliability of the observed diagnostic attributes with the expected diagnostic attribute, which is the measurement of the agreement of an expert's diagnosis compared to the algorithm's diagnosis. The next section will present the methodology developed with the data set and application of the artificial intelligence algorithms to arrive at the results of the work.

IV. Materials and Methods

The study sample consisted of 127 children and adolescents between 6 and 16 years old, 48 with a clinical diagnosis of ADHD and 79 from the control group, with no diagnosis of ADHD. The attributes that make up the neuropsychological tests and behavioral inventories applied in this study are

Attention Cancellation Test (TAC), Trail Making Test (TMT), Wechsler Intelligence Scale for Children (WISC-III), Wechsler Intelligence Scale for Children (WISC-IV), Wechsler Abbreviated Scale of Intelligence (WASI), Child Behavior Checklist for ages 6-18 (CBCL/6-18) and Teacher's Report Form for ages 6-18 (TRF/6-18).

These attributes were normalized by the z-score method [34] to standardize the different scales of the attributes. The normalized data property is used to train the network SOM using the package available in R language [35]. In this library, the functions *somgrid* and *som* are used to parameterize and train the map, respectively. According to Rubbo [36], the map size can be described according to Equations (12) and (13), with n being the number of objects and the constant Cm varying from [-3,3] to generate different map sizes:

$$l_{SOM} = \sqrt{n}/2 + Cm \tag{12}$$

$$Map\ Size = (l_{SOM})^2 \tag{13}$$

For the size of the map topology, the dimension 4x4 was chosen. With this, the hypothesis of the study was to find neurons with a representative density of objects and with a significant class distribution.

With the trained map, the analyses made were the density of objects in each neuron, the distance between neurons, the quality of adjustment of the neurons, the contribution of the attributes in the formation of neurons, and the distribution of the label of each object in each neuron. In addition to the outputs analyzed, the representativeness of the number of objects contained in each neuron with the label attribute was sought in the table generated by the SOM. In this way, the neurons of greater relevance were identified, that is, with larger numbers of objects generated by the SOM algorithm.

From this point on, the entropy algorithm (Shannon's) was used on each neuron in the network to select the neuron with the highest class overlap along with the representativeness of objects that are difficult cases to diagnose. By identifying neurons with overlapping classes, their objects are selected from the database generated by the SOM network for training and validation of the decision tree algorithm. The result of the decision tree brought a hierarchy of attributes in order of discrimination for cases of diagnostic doubt, and the validation of the algorithm shows the performance of the classification.

A. Rating Performance Evaluation

Table I shows the confusion matrix that was used to analyze the classification performance of the decision tree. The table indicates the prediction of the positive and negative scenarios, as well as current true and false scenarios [37]:

- TN is the correct number of negative predictions;
- FP is the number of false positive predictions;
- FN is the number of false negative predictions;
- TP is the correct number of positive predictions.

Table I. Confusion matrix.

	Predicted Negative	Predicted Positive
Current False	TN	FP
Current True	FN	TP

From the confusion matrix, it is possible to measure the performance of the algorithm by calculating the accuracy, as follows:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{14}$$

$$Error = (FP + FN)/(TP + TN + FP + FN) \tag{15}$$

B. Kappa Statistics

The measurement of agreement or Kappa coefficient (K) was used to comparatively measure the ADHD diagnosis of children submitted to the inventories corresponding to the disorder by the ADHD diagnosis predicted in the decision tree. The Kappa coefficient can be calculated with the results of the confusion matrix by Equation (16) [38], [39]:

$$kappa = P(O)\ \text{-}\ P(E)/1\ \text{-}\ P(E) \tag{16}$$

Where P(O) is the observed probability of agreement (sum of the concordant answers divided by the total); P(E) is the expected probability of agreement (sum of the expected values of the concordant answers divided by the total). According to Silva [38], Kappa is a measure of interobserver agreement that evaluates the degree of agreement, as well as whether it is beyond what is expected given chance. It is a maximum unit value measure that corresponds to the absolute agreement and values close to zero or negative, indicating no or lack of agreement between the attributes being judged.

To simplify the methodological description developed in the study, we present a flowchart that briefly describes the process through the programming created to obtain the result and can be seen in Figure 4.

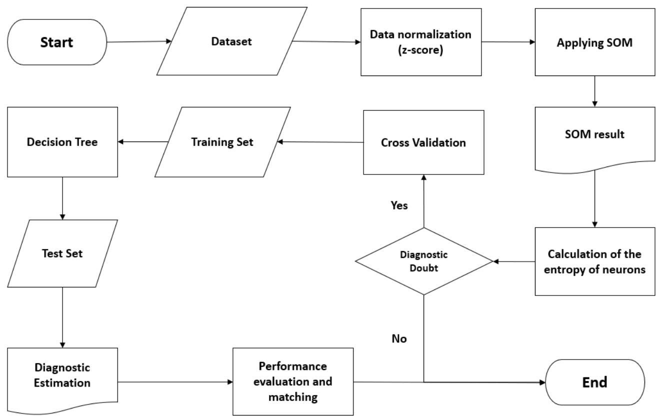

Figure 4. Dataset modeling flowchart

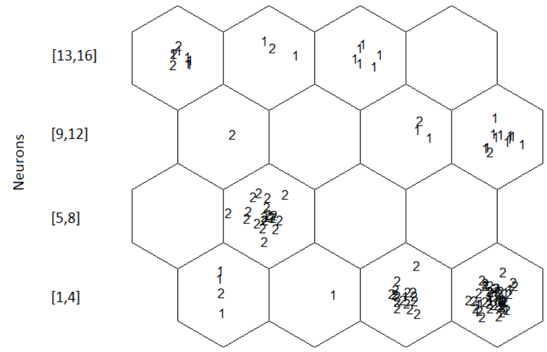

Figure 6. Scattering of objects diagnostic within neurons

Table II. Comparative diagnosis by the neuron dimension 4X4.

	Diagnostic	1	2	Total
neuron	1	3 (6.2%)	1 (1.3%)	4 (3.1%)
	2	1 (2.1%)	0 (0.0%)	1 (0.8%)
	3	1 (2.1%)	16 (20.3%)	17 (13.4%)
	4	16 (33.3%)	35 (44.3%)	51 (40.2%)
	6	0 (0.0%)	20 (25.3%)	20 (15.7%)
	9	0 (0.0%)	1 (1.3%)	1 (0.8%)
	11	2 (4.2%)	1 (1.3%)	3 (2.4%)
	12	11 (22.9%)	1 (1.3%)	12 (9.4%)
	13	6 (12.5%)	3 (3.8%)	9 (7.1%)
	14	2 (4.2%)	1 (1.3%)	3 (2.4%)
	15	6 (12.5%)	0 (0.0%)	6 (4.7%)
Total		48 (100.0%)	79 (100.0%)	127 (100.0%)

After this stage, it is possible to better understand the model's contribution to the understanding of Attention Deficit Hyperactivity Disorder, as well as to the diagnostic evaluation of patients. The next section presents the results obtained in this work.

V. Results

The training result of the SOM network can be seen in two different visualizations, depicted in Figures 5 and 6. Figure 5 presents the attributes, common to the trials, graphically distributed in each neuron. The sizes indicate the contribution that each attribute has to the formation of the neuron. Note that neighboring neurons have similarities among the attributes. In Figure 6, the diagnosis, an attribute that is not used in training the SOM, is projected on the map, allowing visualization of which neurons have the overlap of class 1 (group diagnosed with ADHD) and 2 (control group without ADHD). The network could not separate the diagnosed cases in neuron 4.

Table II presents for each neuron the percentage of objects of each class. Neuron 4 is the one with the highest concentration of objects (40%) and overlapping classes in the whole dataset.

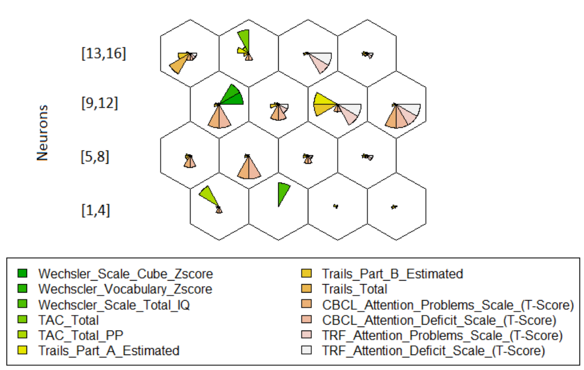

Figure 5. Contribution of the attributes in the formation of the neuron

The result of the decision tree with the data mapped onto neuron 4 can be seen in Figure 7. The result shows that the Child Behavior Checklist for ages 6-18 attribute, specifically the probability of attention problems scale (T-score) [40] [41] neurons had the highest discrimination.

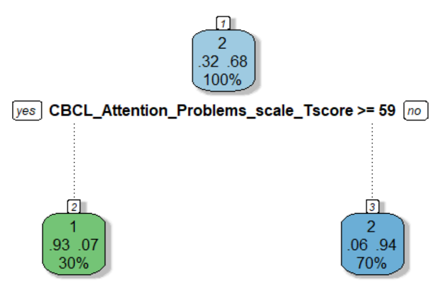

Figure 7. Decision tree of neuron 4

Finally, Figure 8 allows you to visualize all six attributes with greater discrimination for complex cases relative to the integration of clinical evaluation and evaluation using tests for a confirmation of the diagnosis.

The decision tree is a supervised algorithm, has as response attribute an estimated of the class attribute and the result can

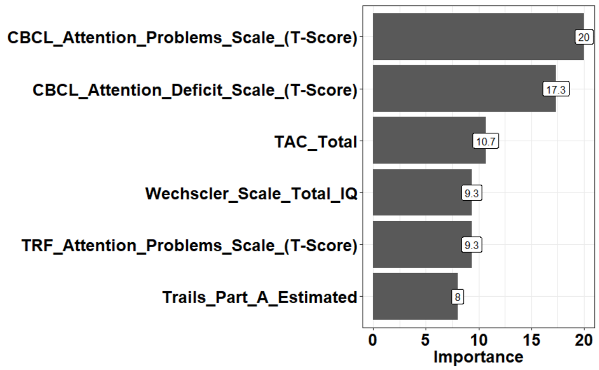

Figure 8. Importance of the attributes in neuron 4 by decision tree

be seen in Table III. Based on the confusion matrix generated, it is observed that of the 35 positive diagnoses for ADHD collected by applying the inventories, there are 33 positive diagnoses estimated by the algorithm. Based on the numbers presented, one can calculate the accuracy of the algorithm, as well as use the Kappa statistic to measure the agreement between the observed objects and the estimated objects [39]. The result with the 4x4 grid generated an accuracy of 88% with a good Kappa reliability of 72%, with the p-value equal to $2.68e^{-7}$.

Table III. Confusion matrix

	Predicted Negative	**Predicted Positive**
Current False	12	4
Current True	2	33

The results presented by the SOM and the decision tree corroborate the three dimensions analyzed and bring a new perspective to cases of doubt in diagnosing ADHD. However, the next section will discuss how the study could help through the mathematical understanding of the interpretation of neurodevelopmental disorder and the possibility of future work using newly supervised or/and unsupervised computational approaches to improve on groups with overlapping diagnoses of ADHD.

VI. Conclusion and Future work

Data from the behavioral assessment inventory [6] can generally be more susceptible to respondent bias because it is based on the answers of the subject. This bias is less so when using cognitive tests which are assessment measures applied directly to the person. Mathematical understanding and model generation is likely to become more difficult using only behavioral inventories. Since ADHD demands the use of both types of measures, in this study both tools were used to apply the decision tree. In the study, it was possible to group the children with and without ADHD by SOM, which made it possible to understand from the perspective of each grouping what was most important in their formation. The self-organizing map contributed especially to the formation of groups and the understanding of clusters with class overlapping, which is the proposal of this work. In this case of overlapping to diagnose a disorder, the decision tree was used to classify the attributes that contributed to the formation of the ADHD group. With this, the predominance of characteristics that helped in the understanding of ADHD in children and adolescents in the study was observed.

The application of the decision tree identified six attributes, namely two of cognitive assessment and four of behavioral assessment, that showed relevant discrimination to make the diagnosis. The Child Behavior Checklist for ages 6-18 attribute the one that showed the highest discriminative power. However, the incidence of low T-scores on the attention problems scale and attention deficit scale does not necessarily imply that the child has ADHD. The results presented showed the difficulty and complexity of finding indicators that define ADHD, as already signaled by some authors [6] [7] [9] [42] [43]. Importantly, the diagnosis of ADHD is a clinical diagnosis that considers the measurement of behavioral correlates of attentional deficits and indicators of hyperactivity and impulsivity in more than one environment. With the Child Behavior Checklist for ages 6-18 attribute being a parent-reported measure, the validity of these two scales for identifying ADHD will likely be confirmed. However, when disregarding the scales, one should consider the evaluations made with the cognitive tests that directly make cognitive measurements and are essential to decide the diagnosis of ADHD. In this study, the tests that contributed the most to this decision tree were the Attention Cancellation Test (ACT) and the Trail Making Test (TMT).

The study presented as a relevant factor the case of overlapping diagnoses of neurons when using the SOM and, in conjunction with the decision tree, was able to separate 88% of the cases. This way, future works can collaborate with the technique addressed in the study through supervised data procedures. These tools can help in making comparisons between results of standardized tests aiming to reduce possible biases of behavioral evaluations based on informants' reports. Future studies can test the same decision tree on larger samples to see if the attributes that showed higher accuracy are maintained. By doing so, the best indices of cognitive and behavioral assessment instruments that contribute to the increased accuracy of ADHD diagnosis may be identified. Since this study controlled for no comorbidities in the ADHD group, it is recommended for future studies to use sample groups with and without ADHD comorbidities from other psychiatric and neurodevelopmental conditions. This type of sample may allow the testing of new and more complex models due to the natural overlap of signs and symptoms between ADHD and some of these comorbidities.

VII. Acknowledgment

We thank the Academic Excellence Program of the Coordination for the Improvement of Higher Education Personnel (CAPES-PROEX), Process number 1133/2019, (CAPES-PROSUC) in mode II, the National Council for Scientific and Technological Development (CNPq, Cases 307730/2017-4 and 307443/2019-1), the Mackenzie Research Fund (MACK-PESQUISA) of Mackenzie Presbyterian University, and the Foundation for Research Support of the State of São Paulo (FAPESP, Cases 2018/01063-0 and 2019/20757-1) for financial support and this study was financed in part by the Academic Excellence Program of the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES-PROEX) - Finance Code 88887.581390/2020-00.

References

[1] A.Silva, L.Carreiro, M. Silva, M. Teixeira and L. Silva, ”Combining Self-Organizing Maps and Decision Tree to Explain Diagnostic Decision Making in Attention-Deficit/Hyperactivity Disorder” The Sixth International Conference on Neuroscience and Cognitive Brain Information BRAININFO 2021. IARIA, Jul. 2021, pp. 5-10.

[2] American Psychiatric Association, DSM-5: Diagnostic and Statistical Manual of Mental Disorders. Artmed publisher, 2014.

[3] E. M. Mahone and M. B. Denckla, ”Attention-deficit/hyperactivity disorder: a historical neuropsychological perspective,” Journal of the International Neuropsychological Society: JINS, Oct. 2017, vol. 23, no. 9-10, pp. 916-929, doi: 10.1017/S1355617717000807.

[4] E. Aretouli, ”How neuropsychology can inform our understanding of preschool ADHD: Clinical and research implications,” Applied Neuropsychology: Child, Apr-Jun. 2019, vol. 8, no. 2, pp. 174-181, doi: 10.1080/21622965.2017.1421463.

[5] M. L. Wolraich, J. F. Jr. Hagan, C. Allan, E. Chan, D. Davison, M. Earls, S. W. Evans, S. K. Flinn, T. Froehlich, J. Frost, J. R. Holbrook, C. U. Lehmann, H. R. Lessin, K. Okechukwu, K. L. Pierce, J. D. Winner and W. Zurhellen, ”Clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents,” Pediatrics, Mar. 2020, vol. 145, no. 3, doi: 10.1542/peds.2019-2528.

[6] L. R. R. Carreiro, J. S. Schwartzman, C. N. Cantiere, A. F. Ribeiro, N. A. Silva, M. A. F. Martin, C. M. Chiquetto, G. S. Baraldi, M. M. C. Mariani, M. F. F. Seraceni and M. C. T. V. Teixeira, ”Interdisciplinary neuropsychological, behavioral and clinical assessment protocol for children and adolescents with complaints of inattention and hyperactivity,” Psychology: theory and practice, [online], Dec. 2014, vol. 16, no. 3, pp. 155-171. Available from: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-36872014000300012&lng=pt&nrm=iso

[7] S. Chandana and K. Vijayalakshmi, ”An Approach to Measure and Improve the Cognitive Capability of ADHD Affected Children Through EEG Signals,” 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT), Mumbai, India, Jul. 2018, pp. 314-318, doi: 10.1109/ICALT.2018.00079.

[8] H. J. Kam, Y. M. Shin, S. M. Cho, S. Y. Kim, K. W. Kim and R. W. Park, ”Development of a decision support model for screening attention-deficit hyperactivity disorder with actigraph-based measurements of classroom activity,” Applied clinical informatics, Nov. 2010, vol. 1, no. 4, p. 377, doi: 10.4338/ACI-2010-05-RA-0033.

[9] S. H. Lee, B. Abibullaev, W. Kang, Y. Shin and, J. An, ”Analysis of attention deficit hyperactivity disorder in EEG using wavelet transform and self organizing maps,” ICCAS 2010, Oct. 2010, pp. 2439-2442, doi: 10.1109/ICCAS.2010.5670255.

[10] G. Polanczyk, M. S. Lima, B. L. Horta, J. Biederman and L. A. Rohde, ”The worldwide prevalence of adhd: a systematic review and metaregression analysis,” American journal of psychiatry, Am Psychiatric Assoc, Jun. 2007, vol. 164, no.6, pp. 942–948, doi: 10.1176/ajp.2007.164.6.942.

[11] V. Simon, P. Czobor, S. Bálint, A. Mészáros and I. Bitter, ”Prevalence and correlates of adult attention-deficit hyperactivity disorder: meta-analysis,” The British Journal of Psychiatry, Cambridge University Press, Mar. 2009, vol. 194, no.3, pp. 204–211, doi: 10.1192/bjp.bp.107.048827.

[12] E. B. P. Benczik and E. B. Casella, ”The impact of Attention Deficit Hyperactivity Disorder (ADHD) on family relationships,” Journal of Psychopedagogy, Jan-Mar. 2015, vol. 32, no. 97, pp. 93-103, 2015.

[13] E. G. Willcutt, J. T. Nigg, B. F. Pennington, M. V. Solanto, L. A. Rohde, R. Tannock, S. K. Loo, C. L. Carlson, K. McBurnett and B. B. Lahey, ”Validity of dsm-iv attention deficit/hyperactivity disorder-symptom dimensions and subtypes,” Journal of abnormal psychology, AmericanPsychological Association, Nov. 2012, vol. 121, no. 4, pp. 991-1010, doi: 10.1037/a0027347.

[14] M. Cherkasova, E. M. Sulla, K. L. Dalena, M. P. Pondé and L. Hechtman, ”Developmental course of attention deficit hyperactivitydisorder and its predictors,” Journal of the Canadian Academy of Child & AdolescentPsychiatry, Feb. 2013, vol. 22, no. 1, pp. 47-54.

[15] S. Dalsgaard, M. K. Humlum, H. S. Nielsen and M. Simonsen, ”Common danish standards in prescribing medication for childrenand adolescents with adhd,” European child & adolescent psychiatry, Springer, Sep. 2014, vol. 23, no. 9, pp. 841–844, doi: 10.1007/s00787-013-0508-5.

[16] N. M. Williams, I. Zaharieva, A. Martin, K. Langley, K. Mantripragada, R. Fossdal, H. Stefansson, K. Stefansson, P. Magnusson, O. O. Gudmundsson, O. Gustafsson, P. Holmans, M. J. Owen, M. O'Donovan and A. Thapar, ”Rare chromosomal deletions and duplications in attention-deficithyperactivity disorder: a genome-wide analysis,” The Lancet, Elsevier, Oct. 2010, vol. 376, no. 9750, pp. 1401–1408, doi: 10.1016/S0140-6736(10)61109-9.

[17] D. Puper-Ouakil, N. Ramoz, A. M. Lepagnol-Bestel, P. Gorwood and M. Simonneau, ”Neurobiology of attention deficit/hyperactivity disorder,” Pediatric research, Nature Publishing Group, May. 2011, vol. 69, no. 8, pp. 69–76, doi: 10.1203/PDR.0b013e318212b40f.

[18] L. R. R. Carreiro, M. C. T. V. Teixeira, M. M. M. Silva, I. T. Paes, R. A. C. B. Novaes and A. P. R. Micieli, ”Neuropsychological Assessment in ADHD: contributions to the identification of cognitive difficulties and school counseling,” In: C. A. H. Amato, D. Brunoni and P. L. Boggio, (Ed.). DEVELOPMENTAL DISORDERS: Interdisciplinary Studies. 1.ed. São Paulo: Memnon, 2018. vol. 1, c. 10, pp. 142–153. ISBN 978-85-7954-146-9.

[19] E. Merényi, M. J. Mendenhall, and P. O'driscoll, ”Advances in self-organizing maps and learning vector quantization,” Advances in Intelligent Systems & Computing, Jan. 2016, vol. 295, pp. 370, https://doi.org/10.1007/978-3-319-28518-4.

[20] J. Vesanto and E. Alhoniemi, ”Clustering of the self-organizing map,” IEEETransactions on neural networks, IEEE, May. 2000, vol. 11, no. 3, pp. 586–600, doi: 10.1109/72.846731.

[21] A. Rauber, D. Merkl and M. Dittenbach, ”The growing hierarchical self-organizingmap: exploratory analysis of high-dimensional data,” IEEE Transactions on NeuralNetworks, IEEE, Nov. 2002, vol. 13, no. 6, pp. 1331–1341, doi: 10.1109/TNN.2002.804221.

[22] T. Kohonen, ”Self-organized formation of topologically correct feature maps,” Biologicalcybernetics, Springer, Jan. 1982, vol. 43, no. 1, pp. 59–69, doi: 10.1007/bf00337288.

[23] M. Kubat, ”Neural networks: a comprehensive foundation by Simon Haykin,” Macmillan, 1994, ISBN 0-02-352781-7. The Knowledge Engineering Review, Feb. 1999, vol. 13, no. 4, pp. 409-412.

[24] T. Kohonen, ”Essentials of the self-organizing map.Neural networks,” Elsevier, 2013, vol. 37, pp. 52–65, doi: 10.1016/j.neunet.2012.09.018.

[25] C. C. Aggarwal, ”Deep Reinforcement Learning. In: Neural Networks and Deep Learning,” Springer, Cham, 2018, pp. 373-417, doi: 10.1007/978-3-319-94463-0.

[26] H. Ritter, ”Neural computation and self-organizing maps: an introduction,” Reading, MA: Addison-Wesley, 1992, pp. 141-161.

[27] M. J. Berry and G. S. Linoff, ”Data mining techniques: for marketing, sales, and customer relationship management,” John Wiley & Sons, 2004.

[28] I. H. Witten, E. Frank, M. A. Hall, and C.J. Pal, ”Data Mining: Practical machine learning tools and techniques,” 2005, vol. 2, pp.4.

[29] B. D. Ville and P. Neville, ”Decision Trees for Analytics: Using SAS Enterprise Miner,” SAS Institute Cary, NC, 2013, pp. 1-14.

[30] L. N. D. Castro and D. G.Ferrari, ”Data Mining Introduction,” Saraiva, 2016, pp. 344–360.

[31] G. S. Linoff and M. J. Berry, ”Data mining techniques: for marketing, sales, andcustomer relationship management,” John Wiley Sons, 2011.

[32] A. Colin, ”Building decision trees with the ID3 algorithm,” Dr Dobbs Journal, 1996, vol. 21, no. 6, pp. 107-+.

[33] T. Mitchel, ”Machine learning,” mcgraw hill education (ise editions), 1997, pp. 870-877.

[34] A. E. Curtis, T. A. Smith, B. A. Ziganshin and J. A. Elefteriades, ”The mystery of the Z-score,” AORTA Journal, Aug. 2016, vol. 4, no. 4, pp. 124-130, doi:10.12945/j.aorta.2016.16.014.

[35] R. Wehrens and J. Kruisselbrink, ”Flexible self-organizing maps in kohonen 3.0,” Journal of Statistical Software, Nov.2018, vol. 87, no. 1, pp. 1-18.

[36] M. Rubbo and L. A. Silva, ”Filtering-based instance selection method for overlapping problem in imbalanced datasets,” J, Jul. 2021, vol. 4, no. 3, pp. 308–327, ISSN 2571-8800, https://doi.org/10.3390/j4030024.

[37] S. Ruuska, ”Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle,” Behavioural processes, Aug. 2018, vol. 148, pp. 56-62, doi: 10.1016/j.beproc.2018.01.004.

[38] R. Silva and A. Paes, ”Inside Statistics: Kappa concordance test,” publication/uuid/3E5F4C37-E639-43D6-89C9-96597CA6AB40, Educ. Contin. health Einstein, 2012, vol. 10, no. 4, pp. 165–166.

[39] H. C. Kraemer, ”Kappa coefficient,” Wiley StatsRef: Statistics Reference Online, WileyOnline Library, Jun. 2014, pp. 1–4, doi: 10.1002/9781118445112.stat00365.pub2.

[40] H. Budak, S. E. Tasabat, ”A modified t-score for feature selection,” Anadolu Üniversitesi Bilim Ve Teknoloji Dergisi A-Uygulamalı Bilimler ve Mühendislik, 2016, vol. 17, no. 5, pp. 845-852, doi: 10.18038/aubtda.279853.

[41] T. M. Nolan, L. Bond, R. Adler, L. Littlefield, P. Birleson, K. Marriage, A. Mawdsley, R. Salo and B. J. Tonge, ”Child Behaviour Checklist classification of behaviour disorder,” Journal of paediatrics and child health, Oct. 1996, vol. 32, no. 5, pp. 405-411, doi: 10.1111/j.1440-1754.1996.tb00939.x.

[42] R. A. Barkley and K. R. Murphy, ”Attention-deficit hyperactivity disorder: A clinical workbook,” Guilford Press, 2006.

[43] F. L. Cibrian, G. R. Hayes, and K. D. Lakes, ”Research Advances in ADHD and Technology,” Synthesis Lectures on Assistive, Rehabilitative and Health-Preserving Technologies, 2020, vol. 9, no. 3, pp. i-156, doi: 10.2200/S01061ED1V01Y202011ARH015.