Comparison of Machine Learning predictive methods to diagnose the Attention Deficit/Hyperactivity Disorder levels using SPECT

Comparação de métodos preditivos de Machine Learning para diagnosticar os níveis do Transtorno de Déficit de Atenção/Hiperatividade usando imagens SPECT

Comparación de métodos predictivos de Machine Learning para diagnosticar los niveles de Trastorno por Déficit de Atención/Hiperactividad utilizando imágenes SPECT

Marcilio de Oliveira Meira
ORCID: https://orcid.org/0000-0002-8547-1607
Federal University of Rio Grande do Norte, Brazil
E-mail: marcilio_meira@hotmail.com
Anne Magaly de Paula Canuto
ORCID: https://orcid.org/0000-0002-3684-3814
Federal University of Rio Grande do Norte, Brazil
E-mail: anne.canuto@ufrn.br
Bruno Motta de Carvalho
ORCID: https://orcid.org/0000-0002-9122-0257
Federal University of Rio Grande do Norte, Brazil
E-mail: bruno@dimap.ufrn.br
Roberto Levi Cavalcanti Jales
ORCID: https://orcid.org/0000-0003-2993-3704
Nuclear Clinic of Natal, Brazil
E-mail: sbbn_ufrn@hotmail.com

Abstract
ADHD (attention deficit hyperactivity disorder) is a neurodevelopmental disorder characterized by harmful levels of inattention, disorganization, and/or hyperactivity-impulsivity. In childhood, these symptoms often overlap with those of other disorders, and they tend to persist into adulthood, interfering with relationships and academic and work life. Diagnosis, traditionally made by assessing the patient, i.e., testing and listening to relatives and teachers, has already been aided by neuroimaging. However, the visual analysis of such images to make a psychiatric diagnosis is a complex and sometimes time-consuming task. For this reason, computer-aided diagnostic tools have increasingly evolved that, when combined with machine learning (ML) techniques, can accelerate, facilitate, and maximize the accuracy of diagnoses. Nevertheless, research evaluating ML models for classifying ADHD considering severity using images of the brain SPECT (Single Photon Emission Computed Tomography) is still very sparse. For this reason, this article aims to evaluate the performance of the ML methods: k-NN (k-Nearest Neighbors), Naive Bayes, Decision Tree, MLP (Multilayer Perceptron) and SVM (Support Vector Machine) in the classification of ADHD. The main goal of this analysis is to check whether the subjects have the disorder or not, and to classify the severity of those who have it using SPECT images. A database was created from SPECT images and diagnostic reports. After pre-processing these data, the best hyperparameters for the ML methods were searched, trained/tested and finally statistically compared. The best results were obtained with SVM and k-NN, with 98% accuracy. Although ADHD diagnosis by neuroimaging is not yet a standard clinical procedure, we argue that this study can contribute to ADHD diagnosis research and support methods for the development of CAD (computer-aided diagnosis) systems.
Keywords: ADHD assisted diagnosis; Computer-aided diagnosis; Machine learning; Nuclear medicine; SPECT.

Resumo
O TDAH (transtorno de déficit de atenção e hiperatividade) é um transtorno do neurodesenvolvimento caracterizado por níveis prejudiciais de desatenção, desorganização e/ou hiperatividade-impulsividade. Na infância, esses sintomas muitas vezes se sobrepõem aos de outros transtornos e tendem a persistir na vida adulta, interferindo nos relacionamentos e na vida acadêmica e profissional. O diagnóstico, tradicionalmente feito pela avaliação do paciente, ou seja, testando e ouvindo familiares e professores, já tem sido auxiliado pela neuroimagem. No entanto, a análise visual de tais imagens para fazer um diagnóstico psiquiátrico é uma tarefa complexa e às vezes demorada. Por esse motivo, têm evoluído cada vez mais ferramentas de diagnóstico auxiliadas por computador que, quando combinadas com técnicas de aprendizado de máquina (ML), podem acelerar, facilitar e maximizar a precisão dos diagnósticos. No entanto, pesquisas avaliando modelos de ML para classificar o TDAH considerando a gravidade usando imagens do

cérebro SPECT (Tomografia Computadorizada por Emissão de Fóton Único) ainda são muito escassas. Por esse motivo, este artigo tem como objetivo avaliar o desempenho dos métodos de ML: *k*-NN (*k*-Nearest Neighbors), Naive Bayes, Decision Tree, MLP (Multilayer Perceptron) e SVM (Support Vector Machine) na classificação do TDAH. O principal objetivo desta análise é verificar se os sujeitos têm ou não o transtorno e classificar a gravidade daqueles que o têm usando imagens SPECT. Um banco de dados foi criado a partir de imagens SPECT e relatórios de diagnóstico. Após o pré-processamento desses dados, os melhores hiperparâmetros para os métodos de ML foram pesquisados, treinados/testados e por fim comparados estatisticamente. Os melhores resultados foram obtidos com SVM e *k*-NN, com 98% de acurácia. Embora o diagnóstico de TDAH por neuroimagem ainda não seja um procedimento clínico padrão, argumentamos que este estudo pode contribuir para a pesquisa do diagnóstico de TDAH e apoiar métodos para o desenvolvimento de sistemas CAD (computer-aided diagnosis).
Palavras-chave: Diagnóstico assistido de TDAH; Diagnóstico auxiliado por computador; Aprendizado de máquina; Medicina nuclear; SPECT.

Resumen
El TDAH (trastorno por déficit de atención con hiperactividad) es un trastorno del neurodesarrollo caracterizado por niveles nocivos de falta de atención, desorganización y/o hiperactividad-impulsividad. En la infancia, estos síntomas a menudo se superponen con los de otros trastornos y tienden a persistir en la edad adulta, interfiriendo con las relaciones y la vida académica y laboral. El diagnóstico, tradicionalmente realizado valorando al paciente, es decir, testeando y escuchando a familiares y profesores, ya ha sido ayudado por la neuroimagen. Sin embargo, el análisis visual de tales imágenes para hacer un diagnóstico psiquiátrico es una tarea compleja y, a veces, lenta. Por esta razón, las herramientas de diagnóstico asistidas por computadora han evolucionado cada vez más y, cuando se combinan con técnicas de aprendizaje automático (ML), pueden acelerar, facilitar y maximizar la precisión de los diagnósticos. Sin embargo, la investigación que evalúa los modelos ML para clasificar el TDAH considerando la gravedad utilizando imágenes del cerebro SPECT (tomografía computarizada por emisión de fotón único) es todavía muy escasa. Por ello, este artículo tiene como objetivo evaluar el desempeño de los métodos ML: *k*-NN (*k*-Nearest Neighbors), Naive Bayes, Decision Tree, MLP (Multilayer Perceptron) y SVM (Support Vector Machine) en la clasificación del TDAH. El principal objetivo de este análisis es comprobar si los sujetos tienen el trastorno o no, y clasificar la gravedad de los que lo tienen mediante imágenes SPECT. Se creó una base de datos a partir de imágenes SPECT e informes de diagnóstico. Después de preprocesar estos datos, se buscaron, entrenaron/probaron y finalmente se compararon estadísticamente los mejores hiperparámetros para los métodos de ML. Los mejores resultados se obtuvieron con SVM y *k*-NN, con un 98% de precisión. Aunque el diagnóstico de TDAH por neuroimagen aún no es un procedimiento clínico estándar, argumentamos que este estudio puede contribuir a la investigación del diagnóstico de TDAH y apoyar métodos para el desarrollo de sistemas CAD (diagnóstico asistido por computadora).
Palabras clave: Diagnóstico asistido por TDAH; Diagnóstico asistido por computadora; Aprendizaje automático; Medicina nuclear; SPECT.

1. Introduction

According to CDC (CDC, Data and Statistics About ADHD, 2021), ADHD (attention-deficit/hyperactivity disorder) is one of the most common neurodevelopmental disorders in childhood, often persisting into adulthood. This disorder is characterized by problems in directing attention, controlling impulsive behavior, or being overly active. It occurs in most cultures and affects about 5% of children and 2.5% of adults. The Diagnostic and Statistical Manual of Mental Disorders or DSM (American Psychiatric Association, 2013) establishes three possible levels of "current severity" of ADHD: mild, moderate, and severe, depending on the degree of impairment or symptoms observed in the patient. According to De Silva, et al. (2019), early diagnosis minimizes long-term effects and helps in the development of intellectual skills. Clinical diagnosis involves several steps without a single test. It usually includes a checklist to assess symptoms and consideration of the child's history from the perspective of parents and teachers (CDC, What is ADHD? 2021).

Kautzky, et al. (2020) warn of a likely challenge to ADHD diagnosis. This is based primarily on behavioral symptoms rather than objective biomarkers because of overlap with symptoms of other disorders. Another difficulty is studying ADHD in adults and trying to retrospectively assess symptoms in childhood. One of the solutions lies in the use of neuroimaging, where the use of EEG (electroencephalogram) and MRI (magnetic resonance imaging) in ADHD diagnosis, also using machine learning (ML) methods, is well established (Pulini, et al., 2019). ADHD significantly affects quality of life, especially from a parent's perspective. Individuals with ADHD have lifelong impairment in psychosocial, educational, and

neuropsychological functioning. It is very important that the disorder is identified early to prevent and treat it effectively to improve the quality of life of those affected (Biederman, et al., 2012; Danckaerts, et al., 2010).

Regarding the areas of the brain whose functions are impaired by ADHD, Amen and Blake (1997), using images of the brain SPECT (Single Photon Emission Computed Tomography), demonstrated that blood flow to the prefrontal cortex is decreased in children and adolescents with ADHD. Kaya, et al. (2003) demonstrated that individuals with ADHD may have significant hypoperfusion in the medial and lateral temporal cortex in the right hemisphere. Goldberg, et al. (1999) found that ADHD is associated with significant dysfunction of the frontal and temporal lobes. Santra and Kumar (2014) also observed hypoperfusion in these regions and found evidence of normalization of prefrontal activity in post-therapy scans available after successful treatment. Several other brain regions are also being studied in ADHD patients, such as the parietal, cingulate gyrus, cerebellum, caudate, and thalamus. Nevertheless, the frontal region predominates in research (Santra & Kumar, 2014).

Nuclear medicine imaging PET (positron emission tomography) and SPECT have been used to diagnose ADHD (Kautzky, et al., 2020; Vázquez-Abad, et al., 2020). Nuclear medicine involves administering a radiopharmaceutical to the patient, the radioactive energy of which is sufficient to penetrate the patient's body and reach a radiation detector that converts it into images (O'malley, et al., 2020). According to Jales and Santos-Filho (2020), PET and SPECT are powerful weapons for the professionals who use them. According to Daniel Amen, MD, adding neuroimaging to patients' medical histories leads to more targeted treatments (Amen, 2012). Technetium-99 is commonly used in SPECT imaging due to its relatively short half-life in the form of the radiopharmaceuticals HMPAO (HexaMetilPropilenAminaOxima) and ECD (Ethyl Cysteinate Dimer).

For all of them, the manual evaluation of all these image details by a simple visual analysis is not a trivial task to make a diagnosis. In Alzheimer's disease (AD), for example, visual diagnosis in early stages is a difficult task that requires experienced specialists (Chaves, et al., 2009). In Parkinson's disease (PD), very subtle changes can be detected on magnetic resonance imaging of the brain by visual analysis (Haller, et al., 2012). According to Illán, et al. (2012), semiquantitative parameters can be used in the studies of PD with SPECT images, where the accuracy of quantification is important to monitor disease progression and therapeutic effects. All these aspects motivate the development of automated quantification techniques that provide similar performance to human operators. Therefore, to assist physicians in their diagnostic work with medical images, it is essential to use a computer-aided diagnosis (CAD) technique. In CAD systems, ML methods have been increasingly used in clinical research with promising results (Dubreuil-Vall, et al., 2020).

There are several studies that evaluate automatic diagnostic methods based on the analysis of images of the brain SPECT (Horn, et al., 2009; Segovia, et al., 2010; Illán, et al., 2012). According to Horn, et al. (2009), this type of tool can help physicians in their daily practice, especially when visual assessment is inconclusive. Segovia, et al. (2010) emphasize that CAD tools are desirable. Illán, et al. (2012) believe that the requirements for a diagnostic tool are potentially met by machine learning techniques for developing CAD systems.

Machine learning (ML) is a set of methods that can automatically detect patterns in data and use them to predict future data (Murphy, 2012). It is a process that induces a function approximation (hypothesis) from a set of sample data (experience) provided to it (Faceli, et al., 2021). After the methods have been trained and learned the hypothesis induction rule, they need to be evaluated. There are two major ML approaches: predictive or supervised and descriptive or unsupervised. In this study, only the predictive methods were used: k-Nearest Neighbors (k-NN), Naive Bayes (NB), Decision Trees (DT), Artificial Neural Networks (ANN) and Support Vector Machines (SVM). The k-NN algorithm classifies a new object based on training examples that are close to it. NB is based on computing probabilities under the assumption that the attribute values of an instance are independent of its class. DT recursively divides a decision problem into subproblems whose solutions can be combined in the form of a tree. ANNs are distributed computing systems consisting of densely connected units (neurons) arranged in layers that process mathematical functions. SVM searches for a hyperplane that best separates instances by their

classes. For more information on these ML methods, see Murphy (2022).

Whether diagnosing ADHD or another psychiatric disorder based on neuroimaging and CAD, research faces similar challenges, but not always in an appropriate manner. Some difficulties can be highlighted: the number of available samples related to the high dimensionality of the image attributes, the imbalance between the frequencies of the classes, and the lack of appropriate methods to validate the results. In particular, about the evaluation of ML methods for automatic ADHD diagnosis, including classification of the severity levels of the disorder, using brain images SPECT, is a lack of previously published studies, suggesting that this is a topic that requires further investigation. Therefore, this is the central topic of this study. Our aim was to evaluate and compare the performance of different ML methods using SPECT imaging modality, in classifying ADHD patients and determining severity for those who suffer from it.

The article is organized as follows: The next subsections provide an overview of related research; Section 2 explains relevant details of our methodology; Section 3 discusses the results found; and Section 4 concludes.

1.1 Related research

Research related to this study is then analyzed, first in the context of diagnosing various psychiatric disorders and then specifically in the context of ADHD.

1.1.1 Mental illnesses classification by ML methods and SPECT images

Table 1 lists several studies using classic ML methods to investigate automated diagnostic methods for various psychiatric disorders using a series of images of the brain SPECT. There are studies on Alzheimer's disease, frontotemporal dementia, Parkinson's disease, cocaine addiction, autism, and disorders with amnestic symptoms. These studies were analyzed and some key aspects are highlighted in Table 1: the disease studied, the size of the dataset, the distribution of classes, the ML methods used, and whether a statistical test was used to validate the results.

Table 1 - Related research in the diagnosis of various psychiatric illnesses.

Author	Diseases	Number of subjects	Class Distribution	ML Methods	Statistical Validation
Horn, et al. (2009)	Alzheimer and frontotemporal dementia	173	47% AD 53% FTD	Logistic regression, linear discriminant analysis, SVM, *k*-NN, multilayer *perceptron* and kernel logistic PLS	No
Chaves, et al. (2009)	Alzheimer	79	51% NOR 49% ATD	SVM	No
Segovia, et al. (2010)	Alzheimer	91	45% Normal 55% AD	SVM	No
Abdi, et al. (2012)	Alzheimer and frontotemporal dementia	104	34% Normal 35% AD 31% FTD	MUBADA, SVM, PCA-LDA, PLS-DA	No
Haller, et al. (2012)	Parkinson	40	43% PD 23% Others	SVM	Yes
Illán, et al. (2012)	Parkinson	208	48% Control 52% PS	SVM, *k*-NN, NM	No
Padilla, et al. (2012)	Alzheimer	97	42% NOR 58% AD	SVM	No
Palumbo, et al. (2014)	Parkinson	90	38% non-PD 62% PD	SVM	No
Oliveira & Castelo-Branco (2015)	Parkinson	654	32% Control 68% PD	SVM	No
Salas-Gonzalez, et al. (2016)	Alzheimer	70	Not specified	SVM	No
Prashanth, et al. (2016)	Parkinson	584	31% Normal 69% PD	Naive Bayes, SVM, Boosted Trees, Random Forests	Yes
Mete, et al. (2016)	Cocaine addiction	162	43% Control 57% Dependent	SVM	No
Höller, et al. (2017)	Disorders with amnestic symptoms	220	18% AD 31% DCI 32% aMCI 19% aSCC	Linear SVM	No
Rondina, et al. (2017)	Alzheimer	38	47% Control 53% AD	SVM and Multiple kernel learning (MKL)	Yes
Ferreira, et al. (2017)	Alzheimer	38	47% Control 53% AD	SVM	Yes
Tagare, et al. (2017)	Parkinson	658	32% Control 68% PD	Logistic Lasso	No
Amen, et al. (2017)	Autism	928	50% HC 50% ASD	Random Forests, Logistic regression, SVM	No
Oliveira, et al. (2018)	Parkinson	652	32% Control 68% PD	SVM, *k*-NN and Logistic regression	Yes
Castillo-Barnes, et al. (2018)	Parkinson	388	50% Control 50% PD	Performance-weighted ensemble classification model	Yes
Jin, et al. (2019)	Parkinson	600	33% Control 67% PD	Relief-SVM	Yes
Nicastro, et al. (2019)	Parkinson	578	36% Control 64% PD	SVM	Yes
Segovia, et al. (2019)	Parkinson	189	50% Control 50% PD	SVM	Yes
Hsu, et al. (2019)	Parkinson	202	03% Normal 11% HYS I 13% HYS II 26% HYS III 43% HYS IV 03% HYS V	Logistic regression and SVM	No
Huang, et al. (2020)	Parkinson	634	27% Normal 20% Nearly normal 13% Potentially Abnormal 40% Abnormal	SVM, Naive Bayes, random forests, XGBoost, logistic regression, linear discriminant analysis	No
Shiiba, et al. (2020)	Parkinson	200	50% NC 50% PD	SVM	No
Castillo-Barnes, et al. (2020)	Parkinson	386	50% HC 50% PD	SVM, Naive Bayes, MLP	No

Source: Author's own.

From the "Diseases" column in Table 1, it appears that the scientific community has paid more attention to Parkinson's disease (PD) and Alzheimer's disease (AD). This is understandable considering that AD is among the 10 diseases

that cause the most deaths in the United States of America (Xu, et al., 2010), and the number of PD cases is expected to reach 1,238,000 by 2030 (Marras, et al., 2018). However, this should not be allowed to crowd out research into other diseases that affect mental health, as there is a wide variety of diseases and each one affects and limits patients' quality of life.

The number of samples from the datasets used in research, represented by "Number of subjects" in Table 1, is a relevant factor for accurate ML methods (Martinez-Murcia, et al., 2017). For this reason, this aspect was analyzed, and it was found that the number of examples is generally low, which can be justified by the fact that the dataset is protected or has limited access. When only a small number of instances are available, it is advisable to use a data augmentation technique (Goodfellow, et al., 2016) to systematically create "artificial data" and add them to the training set. This is because the more data available for training, the better the generalization of the model. None of the 26 papers analyzed used data augmentation, even those with fewer than 100 instances.

From "Class Distribution" in Table 1, it appears that the research generally dealt with unbalanced databases that contained classes more frequently than others, such as in Tagare, et al. (2017), where there were 68% PD subjects but only 32% healthy subjects. This is a common problem in classification problems that needs to be worked around so that the performance of the ML algorithms is not compromised. If ignored, it leads to favoring the classification of new data from classes with a larger number of samples (majority classes). Therefore, the papers in Table 1 were analyzed and found that at least half of them, even those with a relevant number of subjects, ignored this problem. This may have affected the interpretation of their results.

Regarding the ML techniques analyzed in Table 1 in the column "ML Methods", a wide use of SVM (Support Vector Machine) for the classification of mental illness can be seen. According to Cascianelli, et al. (2016), this could be due to its good generalizability or to the fact that its results are comparable to and often superior to those of other ML methods (Faceli, et al., 2021). Nevertheless, we defend studies that compare the performance of more than one classifier type. This would help expand the study of ML techniques for diagnosing mental illness with SPECT neuroimaging.

Regarding the performance of the classifiers, the papers were analyzed and those that used a statistical test were identified in the column "Statistical Validation" (Table 1). It was noted that some of the studies use it only as a pre-processing technique for feature selection. Nevertheless, when comparing two or more classifiers, it is necessary to check whether there are statistically relevant differences between their results. In this issue, it was found that many studies had not statistically validated their results. Others had only tested the statistical significance of accuracy using a permutation test (Wilcox, 2017).

1.1.2 ADHD classification by ML methods and neuroimaging

In this section, Table 2 presents a selection of papers that have examined models for automating ADHD diagnosis based on neuroimaging, machine learning (ML), and deep learning (DL). DL is a particular subtype of ML approaches that use learning to represent the world as a nested hierarchy of concepts (GoodFellow, et al., 2016).

Table 2 - Research related to ADHD diagnosis with neuroimaging.

Author	Image Modality	Number of Subjects	Class Distribution	ML or DL Classifier	Accuracy	Statistical Validation
Peng, et al. (2013)	sMRI	110	50% HC 50% TDAH	ML (ELM and SVM)	90%	Yes
Johnston, et al. (2014)	sMRI	68	50% TDAH 50% Control	ML (SVM)	93%	Not specified
Kuang, et al. (2014)	fMRI	947	62% HC 38% TDAH	DL (DBN)	54%	Not specified
Iannaccone, et al. (2015)	fMRI	36	50% TDAH	ML (SVM)	78%	Not specified
Hao, et al. (2015)	fMRI	947	62% HC 38% TDAH	DL (DBN) and ML (SVM)	72%	Not specified
Deshpande, et al. (2015)	fMRI	947	62% HC 38% TDAH	DL (FCC)	90%	Not specified
Du, et al. (2016)	fMRI	216	55% TDAH	ML (SVM)	95%	Not specified
Ghiassian, et al. (2016)	fMRI and sMRI	769	36% TDAH	ML (MHPC)	70%	Not specified
Qureshi, et al. (2017)	fMRI and sMRI	159	33% TDC 33% ADHDI 33% ADHDC	ML (ELM and SVM)	85%	Yes
Zou, et al. (2017)	fMRI and sMRI	730	35% TDAH	DL (3D CNN)	69%	Not specified
Sen, et al. (2018)	fMRI and sMRI	558	50% TDAH	ML (SVM)	69%	Yes
Zhang-James, et al. (2020)	sMRI	3377	49% Control 51% TDAH	ML (classifiers committee)	62%	Yes
Kautzky, et al. (2020)	PET	38	57% Control 43% TDAH	ML (Random Forests)	82%	Not specified
Vázquez-Abad, et al. (2020)	SPECT	1673	6% HC 94% TDAH	DL (CNN)	99%	Not specified

Source: Author's own.

When analyzing the image type according to the "Image Modality" column in Table 2, it is noticeable that the use of functional magnetic resonance (fMRI) and structural magnetic resonance (sMRI) is widespread. We hypothesize that the availability of public datasets such as ADHD-200 and ENIGMA-ADHD is an important factor in the choice of this imaging modality. All studies in Table 2 that used MRI with more than 500 samples ("Number of Subjects" column) used one of these two databases. sMRI images allow analysis of brain volume and anatomy, whereas fMRI measures brain activity by detecting fluctuations in blood oxygenation (Sen, et al., 2018). The paucity of work in Table 2 using nuclear medicine suggests that further investigation into automated ADHD diagnosis using ML/DL methods, PET, or SPECT imaging is needed.

It should be noted again how important it is to work with a data set in which the distribution of classes is balanced. The "Class Distribution" column in Table 2 shows that there is research that has not solved this problem and leads to strange results. For example, the Vázquez-Abad, et al. (2020) model classifies very confidently when examples of the ADHD class are presented (99% accuracy). However, when HC class (control) examples are presented, it achieves only 79%. The authors suggest that this discrepancy is due to the small number of HC-class examples. Nevertheless, we believe that the use of a class balancing technique, such as SMOTE (Chawla, et al., 2002), may have helped to solve this problem.

As in Table 1, in the column "ML or DL Classifier" in Table 2, SVM was the most commonly used ML method. However, Peng, et al. (2013) and Qureshi, et al. (2017) evaluated the performance of SVM and ELM (extreme learning machine), a variant of a single-layer artificial neural network (Huang, et al., 2006). ELM performed better than SVM. Looking now at the performance of the methods ML and DL (column "Accuracy" in Table 2), given the discrepancies in the number of samples and the different methodological approaches in research, it is not possible to make a statement about which method is better, ML or DL, when it comes to automatic diagnosis of ADHD using neuroimaging. Finally, the column "Statistical Validation" in Table 2 shows that most studies did not report or ignored the statistical tests used to evaluate the performance of their classifier models.

2. Methodology

This section details our methodology, which follows a sequence adopted from KDD, Knowledge Discovery from Data

(Fayyad, et al., 1996). The steps were: data acquisition, database preparation, pre-processing, finding the best hyperparameters, training and testing models, and statistical validation of results. Our algorithms have used the Python[1] programming language from the Anaconda[2] distribution and the *sci-kit learn*[3] machine learning library.

2.1 Data acquisition

The brain SPECT was recorded in 236 patients with an intravenous injection of 20mCi 99-mTc/HMPAO. This method was similar to that described by Mena (2009), in which images were acquired using a Siemens Corp. dual-head ECAM system. Dual image reconstruction was performed using Segami Corporation's OASIS software. One was without attenuation correction for images of lateral, anteroposterior, and superior cerebral cortex. The second with attenuation correction using a Chang coefficient of 0.1 was applied to parasagittal images and images of the inferior brain, from which the cerebellum was removed. This allowed us to easily examine the lower aspects of the occipital and temporal lobes and the basal ganglia.

The result of the acquisition was a set of 2D images, and a diagnostic report prepared by a specialist for each of the 236 patients. Figure 1 shows an example of a brain SPECT, acquired from a patient with bipolar disorder, ADHD, obsessive-compulsive disorder (OCD), and schizoaffective disorder. For each image, 8 views of the cerebral cortical perfusion (right medial, right lateral, anterior, superior, left medial, left lateral, posterior, and inferior) and 6 views of the basal nuclei were recorded.

Figure 1 - SPECT obtained with HMPAO from a patient with bipolarity, ADHD, OCD and schizoaffective disorder.
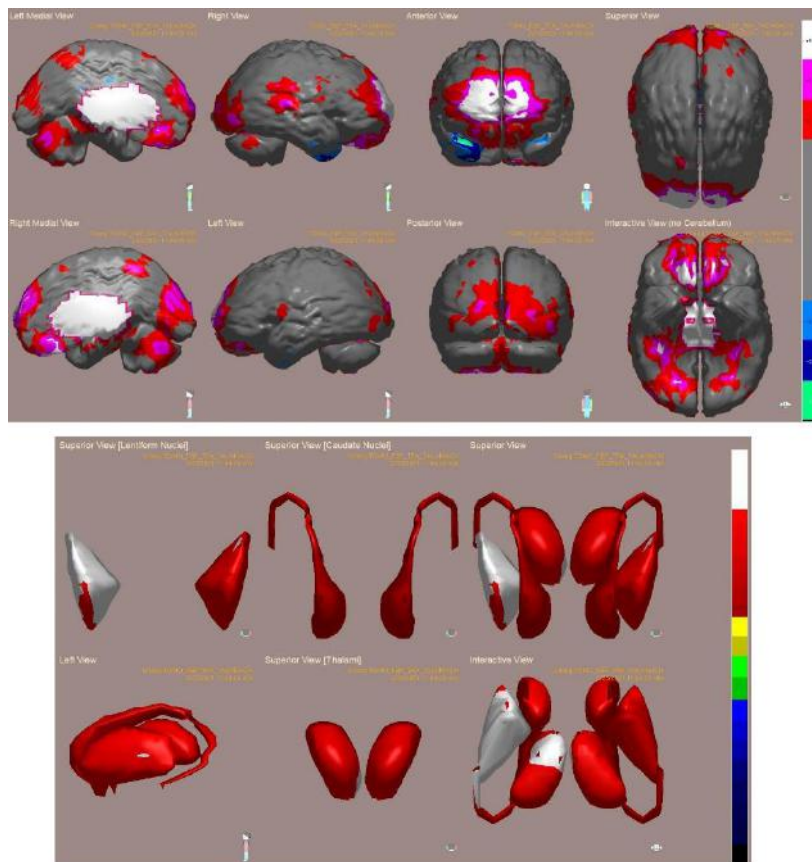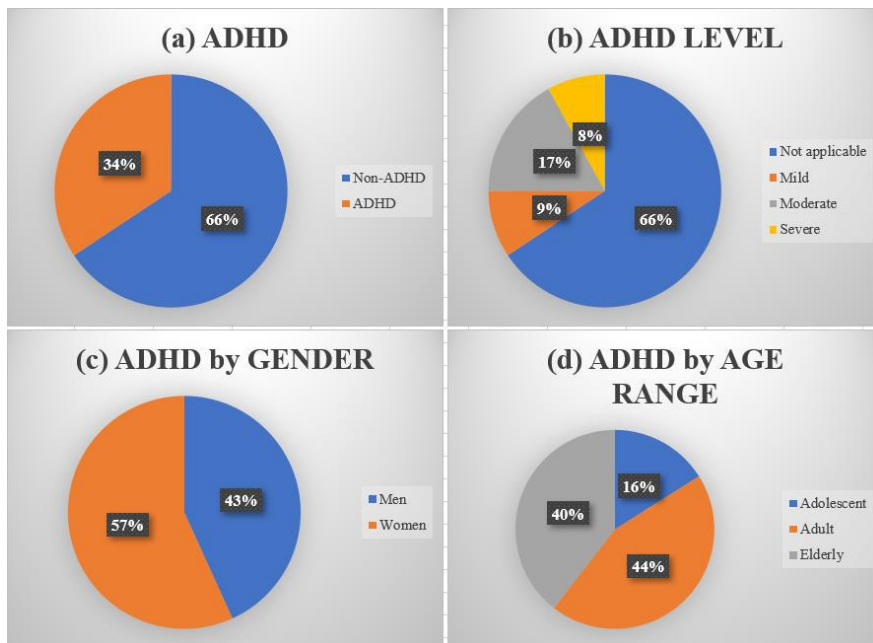

Source: Author's own.

2.2 Database preparation

Of the 236 subjects, 81 had an ADHD diagnosis and 155 had no ADHD diagnosis. The original distribution of the data is shown in Figure 2, where you can see the percentages of: (a) subjects who have the disorder and those who do not; (b) severity of ADHD, with those who do not have it classified as "Not Applicable"; (c) ADHD by sex; and (d) ADHD by age group.

The pixels of the images from SPECT were specified as input attributes for our database. For the output attribute, a single attribute named ADHD LEVEL was created with 4 discrete values (classes): 0 for individuals not diagnosed with ADHD; 1 for individuals with mild ADHD; 2 for moderate ADHD; and 3 for severe ADHD. Our strategy was to perform two tasks: Classifying ADHD and simultaneously assigning the severity level to individuals who have the disorder. The labeling of the baseline attributes was based on the diagnostic reports.

Finally, as mentioned earlier, there are a variety of studies that differ on which regions of brain function are affected by ADHD. The only region that is consistent in the research is the frontal region. Therefore, only the 3rd view (anterior view) was cropped from this group of images (Figure 1), leaving only one image per patient (Figure 3). The pixels of the resulting image were still preprocessed before being used in training/testing.

Figure 2 - Original dataset distribution.

Source: Author's own.

Figure 3 - 2D view of the frontal cortex extracted from the SPECT of an ADHD patient.
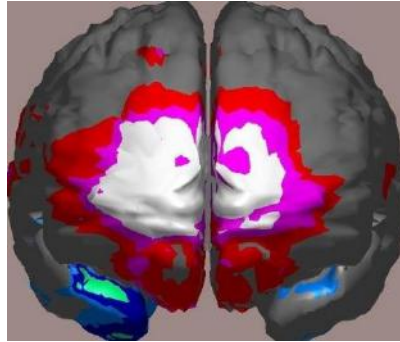
Source: Author's own.

2.3 Pre-processing

Some data quality issues may affect the process of hypothesis induction in the ML methods. Therefore, it is necessary to apply pre-processing techniques such as data cleaning, transformations, and dimensionality reduction (Bishop, 2006). We consider that our original database contains a small number of samples to achieve a good generalization of the models. In addition, it is evident from the percentages that the classes were originally unbalanced (Figure 2-b). Therefore, a pre-processing strategy had to be used to solve both problems, i.e., to increase the number of samples and to achieve a balance between classes. Therefore, the technique SMOTE was used, the Synthetic Minority Over-sampling Technique (Chawla, et al., 2002). This helped to balance the number of samples between the majority and minority classes. It was chosen to increase the number of minority class instances to match the number of majority classes. After applying SMOTE, all classes had a frequency of 25% and increased from 236 to 620 samples.

However, we recommend that images that need to be *standardized* or *normalized* be processed before applying SMOTE. Therefore, pixel standardization and normalization in our algorithms were performed immediately after loading the images so that SMOTE could be applied.

The next step was feature selection, because each image had a size of 611 x 519 pixels, which would result in 317,109 input attributes to be computed by the ML algorithms. Consequently, it was necessary to reduce the number of attributes to improve the performance of the induced model, reduce the computational cost, and make the results more understandable. Therefore, Incremental Principal Component Analysis (IPCA) was used instead of PCA, Principal Component Analysis (Murphy, 2012, p. 387). Our dataset was too large to fit in memory and had to be decomposed. IPCA creates a low-rank approximation for the input data using a memory size that is independent of the number of input data samples. After applying IPCA, the dimensionality dropped to 100 input attributes for each of the 620 samples

2.4 Search for best hyperparameters

Most ML algorithms have hyperparameters that affect their performance, processing time, or memory consumption. The choice of these parameters, which can be manual or automatic, affects the quality of the model and its ability to generalize to new entries (Goodfellow, et al., 2016).

Automatic selection was used in this study. The *GridSearchCV* class, from the *sci-kit learn framework*, was instantiated and given the parameters: a classifier object, a specific parameter array for each classifier type, the performance evaluation strategy, and a *StratifiedKFold* object for the stratified cross-validation strategy. For each of the ML methods this process was performed to determine the best parameter configuration. The nested cross-validation strategy (Cawley & Talbot, 2010) was used and then the *accuracy* and *F-measure*, a harmonic mean of *precision* and *recall*, were calculated (Olson &

Delen, 2008, p. 138). The percentage of instances and the image resolution rate were also varied to investigate how data reduction might affect the results.

2.5 Training and testing

Once the best hyperparameter settings for each classifier type were determined, the training phase began. A procedure was implemented to train each ML method. To evaluate performance, *cross-validation* was chosen, a learning strategy in which the dataset is divided into a fixed number of subsets (*folds*) and each fold is used once as test data while the rest is used for training. 10 subsets were used (*tenfold cross-validation*), but in such a way that each subset maintains an approximately balanced distribution of classes (*stratified cross-validation*).

In addition, the training and cross-validation were performed in a loop with 30 iterations. According to Witten et al., (2011), this generally strengthens the mitigation of biases caused by a particular sample selected for validation. In the end, the average of the results was calculated using the metrics of *accuracy* and *F-measure*.

2.6 Statistical validation

Demšar (2006) reported that the ML community is increasingly aware of the need for statistical validation of published results. He reasoned that this is due to the maturity of the field, the more frequent use of ML in applications, and the availability of new ML frameworks that facilitate the implementation and comparison of algorithms. García and Herrera (2008) also explored the use of statistical comparisons of classifiers across multiple datasets, extending Demšar's work.

Given our problem context, where the same dataset was used for all algorithms (pairing of samples), and the measurement level of the performance variables (ordinal), the *Friedman test* (Friedman, 1937) was applied, a nonparametric statistical test. The defined H0 hypothesis was "there are no differences between the samples", meaning that there are no statistically relevant differences between the performance of the models. Otherwise, an alternative hypothesis would be adopted. This would lead us to use a paired test, for example, comparing two classifiers using the *Nemenyi test* (Nemenyi, 1963).

To facilitate analysis of the results of the paired tests, the CD diagram was used (Demšar, 2006), a simple diagram that provides the *critical differences* for the *Nemenyi test*. This diagram presents the order of algorithms, the magnitude of differences between them (in terms of rankings) and the importance of observed differences.

3. Results

The discussion begins with a commentary on the best hyperparameters found in an automated way for our data context. In Table 3, you can see the list of evaluated values and the best configuration found for each method. Although Naive Bayes has no parameters, both *sci-kit learn* implementations were tested: *GaussianNB* and *BernoulliNB*, and the best result was obtained with *GaussianNB*. In this parameter search, the percentage of instances and the image resolution rate were varied and found that the more instances, the better the results. Decreasing the image resolution only slightly improved the results for most methods, except for *k*-NN. Therefore, 100% of the instances and the original resolution of the images were chosen for training/testing and statistical validation.

Table 3 - List of best discovered hyperparameters.

Method	Parameter Array	Best Configuration
SVM	C = [0.1, 1, 10, 100, 1000], kernel = ("linear", "rbf")	C = 100 and kernel = "rbf"
k-NN	N_neighbors = [1, 3, 5, 7, 9, 11]	n_neighbors = 1
MLP	solver = ["adam", "sgd", "lbfgs"], max_iter = [6000, 8000, 10000], hidden_layer_sizes = [7, 11, 15, 19, 23, 27], momentum = [0.1, 0.3, 0.8, 0.9], learning_rate_init = [0.01, 0.001, 0.0001]	solver = "sgd", max_iter = 8000, hidden_layer_sizes = (23,) momentum = 0.8, learning_rate_init = 0.01
Decision Tree	criterion = ["gini","entropy"] max_depth = [2, 5, 6, 10], min_samples_split = [2, 3, 4], min_samples_leaf = [1, 2, 3, 4]	criterion = "gini", max_depth = 10, min_samples_split = 3, min_samples_leaf = 1
Naive Bayes	-	-

Source: Author's own.

In an effort to reduce as much as possible the problems of our dataset, such as the small number of instances and the unbalanced classes, the technique SMOTE was added to the pre-processing phase. However, to better understand how this technique affects the results, the models were first trained with the original data distribution and then SMOTE was applied to evaluate performance using the metrics of *accuracy* and *F-measure*.

The results are presented in Table 4. From the "Accuracy" and "F-measure" columns, it can be seen that the use of SMOTE improves the accuracy of the methods by about 33% on average. MLP, for example, improved its accuracy by 39%. For this reason, only the results of classes balanced by SMOTE were included in our analyzes. It can also be observed that the differences between the values of the "Accuracy" and "F-measure" columns (*right side* of Table 4) are extremely small. This could be explained by the balance of classes achieved by SMOTE. Accordingly, **accuracy** was chosen as the performance measure for comparing the classifiers.

The methods are ranked by "Accuracy" (bold values) in Table 4. The three best performances were obtained by SVM, k-NN, and MLP classifiers, followed by Decision Tree (DT) and Naive Bayes (NB). It is believed that the first three methods have better generalization than the other two.

Table 4 - Results of ML Methods Performance.

| Method	Performance with Original Distribution Data				Performance with Balanced Classes by SMOTE			
	Accuracy	Std. Dev.	F-measure	Std. Dev.	Accuracy	Std. Dev.	F-measure	Std. Dev.
SVM	0.7222	0.0646	0.6793	0.0644	**0.9780**	0.0171	0.9778	0.0175
k-NN	0.7099	0.0736	0.6860	0.0709	**0.9769**	0.0169	0.9767	0.0172
MLP	0.6919	0.0836	0.6684	0.0769	**0.9595**	0.0233	0.9588	0.0240
Decision Tree	0.7351	0.0727	0.7308	0.0682	**0.9134**	0.0329	0.9121	0.0320
Naive Bayes	0.6639	0.0776	0.6631	0.0731	**0.8529**	0.0418	0.8525	0.0419

Source: Author's own.

Using the *Friedman test*, it was confirmed that the observed differences in accuracy were statistically significant. Therefore, based on the accuracy, the mean ranks for each classifier were calculated and the CD diagram (Critical Differences diagram) was created for the *Nemenyi test*. This diagram helped to describe the differences more accurately and to understand them better. From the CD diagram (Figure 4), it can be seen that connected classifiers: SVM and k-NN; MLP and DT; and DT and NB are not significantly different. In any case, from the CD diagram, it can be concluded that the performances of SVM

and *k*-NN are statistically superior to those of MLP, DT, and NB. Thus, the importance of statistical validation as a suitable approach for comparing the results of different ML methods has been confirmed.

Regarding the two best classifiers, SVM and *k*-NN, it can be seen in Table 3 that the best SVM version for our experiments was the one using the radial basis function (RBF) kernel (Lazzaro & Montefusco, 2002). It can also be seen from Table 4 and Figure 4 that the SVM and *k*-NN results are very close. In fact, *k*-NN and SVM with RBF kernel perform similarly. However, SVM is more efficient than *k*-NN because SVM only needs to remember the examples that make up the support vectors.

Figure 4 - CD diagram comparing all classifiers using Nemenyi test.
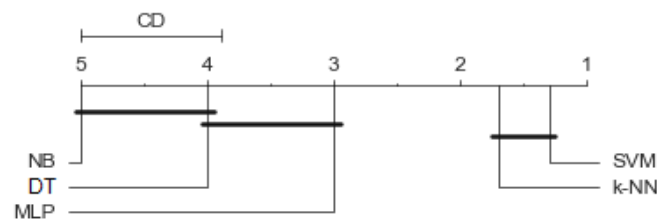
Source: Author's own.

4. Final considerations

In this work, the performance of five machine learning (ML) methods was evaluated: *k*-NN, Decision Trees (DT), Naive Bayes (NB), MLP, and SVM in classifying ADHD patients using brain images SPECT. There is strong evidence that the ML methods can provide satisfactory models that can help in ADHD diagnosis, including severity level classification. SVM and *k*-NN were the best ML methods for this problem. The high accuracy rates of the models generated may reflect correct decisions about our methodological procedures. We took care of solving the data problems, found the best fit of the hyperparameters, and applied training/testing procedures appropriate for our problem context. Analogous to Vázquez-Abad, et al. (2020), we can suggest that ADHD severity level classification applications developed with our models would potentially be more accurate, faster, and less expensive, which would improve patient treatment.

It is expected that this work will be useful for researching computational methods for building CAD applications for automatic diagnosis of ADHD. In addition to comparing the performance of different machine learning methods, methodological directions can also be found, such as hyperparameter values. Nevertheless, much more research and refinement of this research is needed. Future improvement could be to include the classification of ADHD subtypes, to consider the comorbidities of ADHD, and to consider the heterogeneity of this disorder (Pulini, et al., 2019).

Acknowledgments

References

Abdi, H., Williams, L. J., Beaton, D., Posamentier, M. T., Harris, T. S., Krishnand, A., & Devous, M. D. (2012, January 01). Analysis of regional cerebral blood flow data to discriminate among Alzheimer's disease, frontotemporal dementia, and elderly controls: a multi-block barycentric discriminant analysis (MUBADA) methodology. *Journal of Alzheimer's Disease : JAD, 31*(3), 189-201. 10.3233/jad-2012-112111

Amen, D. (2012). Daniel Amen, MD: the impact of brain imaging on psychiatry and treatment for improving brain health and function. Interview by Karen Burnett. *Alternative Therapies in Health and Medicine, 18*(2), 52–58.

Amen, D., Sarabi, M., Willeumier, K., Taylor, D., Raji, C., Meysami, S., & Raghavendra, C. (2017, April 10). Functional SPECT neuroimaging using machine learning algorithms distinguishes autism spectrum disorder from healthy subjects. *Syst Integr Neurosci, 3*(3), 1-9. 10.15761/JSIN.1000160

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* (5th ed.). Arlington, VA: American Psychiatric Association.

Biederman, J., Petty, C. R., Woodworth, K. Y., Lomedico, A., Hyder, L. L., & Faraone, S. V. (2012). Adult outcome of attention-deficit/hyperactivity disorder: a controlled 16-year follow-up study. *The Journal of Clinical Psychiatry, 73*(7), 941-950. 10.4088/JCP.11m07529

Bishop, C. M. (2006). *Pattern Recognition And Machine Learning.* New York, NY: Springer.

Cascianelli, S., Scialpi, M., Amici, S., Forini, N., Minestrini, M., Fravolini, M., & Palumbo, B. (2017). Role of Artificial Intelligence Techniques (Automatic Classifiers) in Molecular Imaging Modalities in Neurodegenerative Diseases. *Current Alzheimer Research, 14*(2), 198-207. 10.2174/1567205013666160620122926

Castillo-Barnes, D., Martinez-Murcia, F. J., Ortiz, A., Salas-Gonzalez, D., RamÍrez, J., & Górriz, J. M. (2020). Morphological Characterization of Functional Brain Imaging by Isosurface Analysis in Parkinson's Disease. *International Journal of Neural Systems, 30*(9), p. 2050044. 10.1142/S0129065720500446

Castillo-Barnes, D., Ramírez, J., Segovia, F., Martínez-Murcia, F. J., Salas-Gonzalez, D., & Górriz, J. M. (2018, August 14). Robust Ensemble Classification Methodology for I123-Ioflupane SPECT Images and Multiple Heterogeneous Biomarkers in the Diagnosis of Parkinson's Disease. *Frontiers in Neuroinformatics, 12*, 53. 10.3389/fninf.2018.00053

Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research, 11*, 2079-2107. 10.5555/1756006.1859921

CDC. (2021, September 23). *Data and Statistics About ADHD*. Centers for Disease Control and Prevention: https://www.cdc.gov/ncbddd/adhd/data.html

CDC. (2021, September 23). *What is ADHD?* Center for Disease Control and Prevention: https://www.cdc.gov/ncbddd/adhd/facts.html

Chaves, R., Ramírez, J., Górriz, J., López, M., Salas-Gonzalez, D., Álvarez, I., & Segovia, F. (2009, September 15). SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting. *Neuroscience Letters, 461*(3), 293-297. 10.1016/j.neulet.2009.06.052

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002, January). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*(1), 321–357. 10.5555/1622407.1622416

Danckaerts, M., Sonuga-Barke, E. J., Banaschewski, T., Buitelaar, J., Döpfner, M., Hollis, C., & Coghill, D. (2010, February). The quality of life of children with attention deficit/hyperactivity disorder: a systematic review. *Eur Child Adolesc Psychiatry, 19*(2), 83-105. doi:10.1007/s00787-009-0046-3

De Silva, S., Dayarathna, S., Ariyarathne, G., Meedeniya, D., & Jayarathna, S. (2019). A Survey of Attention Deficit Hyperactivity Disorder Identification Using Psychophysiological Data. *International Journal of Online and Biomedical Engineering, 15*(13), 61-76. 10.3991/ijoe.v15i13.10744

Demšar, J. (2006, December 01). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research, 7*, 1-30. 10.5555/1248547.1248548

Deshpande, G., Wang, P., Rangaprakash, D., & Wilamowski, B. (2015). Fully Connected Cascade Artificial Neural Network Architecture for Attention Deficit Hyperactivity Disorder Classification from Functional Magnetic Resonance Imaging Data. *IEEE Transactions on Cybernetics, 45*(12), 2668-2679. 10.1109/TCYB.2014.2379621

Du, J., Wang, L., Jie, B., & Zhang, D. (2016, April 23). Network-based classification of ADHD patients using discriminative subnetwork selection and graph kernel PCA. *Computerized Medical Imaging and Graphics : the Official Journal of the Computerized Medical Imaging Society, 52*, pp. 82-88. 10.1016/j.compmedimag.2016.04.004

Dubreuil-Vall, L., Ruffini, G., & Camprodon, J. A. (2020, April 9). Deep Learning Convolutional Neural Networks Discriminate Adult ADHD From Healthy Individuals on the Basis of Event-Related Spectral EEG. *Frontiers in Neuroscience, 14*, p. 251. 10.3389/fnins.2020.00251

Faceli, K., Lorena, A. C., Gama, J., & De Carvalho, A. C. (2021). *Inteligência Artificial: uma Abordagem de Aprendizado de Máquina.* Rio de Janeiro: LTC.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine, 17*(3), 37. From Data Mining to Knowledge Discovery in Databases

Ferreira, L. K., Rondina, J. M., Kubo, R., Ono, C. R., Leite, C. C., Smid, J., & Buchpiguel, C. A. (2017, October 02). Support vector machine-based classification of neuroimages in Alzheimer's disease: direct comparison of FDG-PET, rCBF-SPECT and MRI data acquired from the same individuals. *Revista Brasileira de Psiquiatria (Sao Paulo, Brazil: 1999), 40*(2), 181-191. 10.1590/1516-4446-2016-2083

Friedman, M. (1937, December). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J Am Stat Assoc, 32*(200), 675-701. 10.2307/2279372

García, S., & Herrera, F. (2008). An extension on "Statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. *Journal of Machine Learning Research, 9*(89), 2677-2694.

Ghiassian, S., Greiner, R., Jin, P., & Brown, M. R. (2016, December 28). Using Functional or Structural Magnetic Resonance Images and Personal Characteristic Data to Identify ADHD and Autism. *Plos One, 11*(12), e0166934. 10.1371/journal.pone.0166934

Goldberg, M., Mena, I., & Miller, B. (1999, July 1). Frontal and temporal lobe dysfunction in autism and other related disorders: ADHD and OCD. *Alasbimn Journal*. http://www.alasbimnjournal.net/a/80

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* Cambridge, Massachusetts: The MIT Press.

Haller, S., Badoud, S., Nguyen, D., Garibotto, V., Lovblad, K., & Burkhard, P. (2012, December). Individual detection of patients with Parkinson disease using support vector machine analysis of diffusion tensor imaging data: initial results. *American Journal of Neuroradiology, 33*(11), 2123-2128. 10.3174/ajnr.A3126

Hao, A. J., He, B. L., & Yin, C. H. (2015). Discrimination of ADHD children based on Deep Bayesian Network. *IET International Conference on Biomedical Image and Signal Processing (ICBISP 2015)*, 1-6. 10.1049/cp.2015.0764

Höller, Y., Bathke, A. C., Uhl, A., Strobl, N., Lang, A., Bergmann, J., & Staffen, W. (2017, September 7). Combining SPECT and Quantitative EEG Analysis for the Automated Differential Diagnosis of Disorders with Amnestic Symptoms. *Frontiers in Aging Neuroscience, 7*(9), 290. 10.3389/fnagi.2017.00290

Horn, J., Habert, M., Kas, A., Malek, Z., Maksud, P., Lacomblez, L., & Fertil, B. (2009, October). Differential automatic diagnosis between Alzheimer's disease and frontotemporal dementia based on perfusion SPECT images. *Artificial Intelligence in Medicine, 47*(2), 147-158. 10.1016/j.artmed.2009.05.001

Hsu, S. Y., Lin, H. C., Chen, T. B., Du, W. C., Hsu, Y. H., Wu, Y. C., & Chen, H. Y. (2019, April). Feasible Classified Models for Parkinson Disease from 99mTc-TRODAT-1 SPECT Imaging. *Sensors (Basel, Switzerland), 19*(7), 1740. 10.3390/s19071740

Huang, G., Zhu, Q., & Siew, C. (2006, December). Extreme learning machine: Theory and applications. *Neurocomputing, 70*(1-3), 489-501. 10.1016/j.neucom.2005.12.126

Huang, S. F., Wen, Y. H., Chu, C. H., & Hsu, C. C. (2020). A Shape Approximation for Medical Imaging Data. *Sensors (Basel, Switzerland), 20*(20), 5879. 10.3390/s20205879

Iannaccone, R., Hauser, T. U., Ball, J., Brandeis, D., Walitza, S., & Brem, S. (2015, October). Classifying adolescent attentiondeficit/hyperactivity disorder (ADHD) based on functional and structural imaging. *European Child & Adolescent Psychiatry, 24*(10), 1279-1289. 10.1007/s00787-015-0678-4

Illán, I. A., Gorrz, J. M., Ramirez, J., Segovia, F., Jimenez-Hoyuela, J. M., & & Ortega Lozano, S. J. (2012, October). Automatic assistance to Parkinson's disease diagnosis in DaTSCAN SPECT imaging. *Medical Physics, 39*(10), 5971-5980. 10.1118/1.4742055

Jales, R. L., & Santos-Filho, S. D. (2020). Discussion of Review on Protocols Treatments of Arachnoid Cysts Explored by Nuclear Medicine. *EC Clinical and Medical Case Reports, 3*(6), 147-152.

Jin, L., Zeng, Q., He, J., Feng, Y., & & Wu, Y. (2019, January 1). A ReliefF-SVM-based method for marking dopamine-based disease characteristics: A study on SWEDD and Parkinson's disease. *Behavioural Brain Research, 356*(1), pp. 400-407. 10.1016/j.bbr.2018.09.003

Johnston, B. A., Mwangi, B., Matthews, K., Coghill, D., Konrad, K., & Steele, J. D. (2014, October). Brainstem abnormalities in attention deficit hyperactivity disorder support high accuracy individual diagnostic classification. *Human Brain Mapping, 35*(10), 5179-5189. 10.1002/hbm.22542

Kautzky, A., Vanicek, T., Philippe, C., Kranz, G. S., Wadsak, W., Mitterhauser, M., & Lanzenberger, R. (2020, April 07). Machine learning classification of ADHD and HC by multimodal serotonergic data. *Translational Psychiatry, 10*(1), p. 104. doi:Machine learning classification of ADHD and HC by multimodal serotonergic data

Kaya, G. C., Pekcanlar, A., Bekis, R., Ada, E., Miral, S., Emiroğlu, N., & Durak, H. (2002, December). Technetium-99m HMPAO brain SPECT in children with attention deficit hyperactivity disorder. *Annals of Nuclear Medicine, 16*(8), 527-531. 10.1007/BF02988629

Kuang, D., Guo, X., An, X., Zhao, Y., & He, L. (2014). Discrimination of ADHD Based on fMRI Data with Deep Belief Network. In C. Springer (Ed.), *ICIC 2014: Intelligent Computing in Bioinformatics*, *8590*, pp. 225-232. 10.1007/978-3-319-09330-7_27

Lazzaro, D., & Montefusco, L. B. (March de 2002). Radial basis functions for the multivariate interpolation of large scattered data sets. *Journal of Computational and Applied Mathematics, 140*(1-2), 521-536. 10.1016/S0377-0427(01)00485-X

Marras, C., Beck, J. C., Bower, J. H., Roberts, E., Ritz, B., Ross, G. W., & Group, P. F. (2018, July 10). Prevalence of Parkinson's disease across North America. *npj Parkinson's Disease, 4*(1), p. 21. 10.1038/s41531-018-0058-0

Martinez-Murcia, F. J., Górriz, J. M., J., R., Illán, I. A., Segovia, F., Castillo-Barnes, D., & Salas-Gonzalez, D. (2017, 14 November). Functional Brain Imaging Synthesis Based on Image Decomposition and Kernel Modeling: Application to Neurodegenerative Diseases. *Frontiers in Neuroinformatics, 11*, p. 65. 10.3389/fninf.2017.00065

Mena G, I. (2009, July). Neurospect: Imagenología Funcional em Psiquiatría. *Alasbimn Journal, 11*(45).

Mete, M., Sakoglu, U., Spence, J. S., Devous, M. D., Sr, H. T., & Adinoff, B. (2016). Successful classification of cocaine dependence using brain imaging: a generalizable machine learning approach. *BMC Bioinformatics, 17(Suppl 13)*, 357. 10.1186/s12859-016-1218-z

Murphy, K. P. (2012). *Machine Learning: a Probabilistic Perspective.* Cambridge, Massachusetts: MIT Press.

Nemenyi, P. B. (1963). *Distribution-free Multiple Comparisons.* Princeton, New Jersey: Princeton University.

Nicastro, N., Wegrzyk, J., Preti, M. G., Fleury, V., Van de Ville, D., Garibotto, V., & Burkhard, P. R. (2019, July). Classification of degenerative parkinsonism subtypes by support-vector-machine analysis and striatal 123I-FP-CIT indices. *Journal of Neurology, 266*(7), 1771-1781. 10.1007/s00415-019-09330-z

Oliveira, F. P., & Castelo-Branco, M. (2015, April). Computer-aided diagnosis of Parkinson's disease based on [(123)I]FP-CIT SPECT binding potential images, using the voxels-as-features approach and support vector machines. *Journal of Neural Engineering, 12*(2), 026008. 10.1088/1741-2560/12/2/026008

Oliveira, F. P., Faria, D. B., Costa, D. C., Castelo-Branco, M., & Tavares, J. M. (2018, June). Extraction, selection and comparison of features for an effective automated computer-aided diagnosis of Parkinson's disease based on [123I]FP-CIT SPECT images. *European Journal of Nuclear Medicine and Molecular Imaging, 45*(6), 1052-1062. 10.1007/s00259-017-3918-7

Olson, D., & Delen, D. (2008). *Advanced Data Mining Techniques.* Heidelberg: Springer-Verlag Berlin Heidelberg.

O'Malley, J. P., Ziessman, H. A., & Thrall, J. H. (2020). *Nuclear Medicine and Molecular Imaging: The Requisites* (5th ed.). Elsevier Health Sciences.

Padilla, P., López, M., Górriz, J. M., Ramírez, J., Salas-González, D., & Álvarez, I. (2012, February). NMF-SVM Based CAD Tool Applied to Functional Brain Images for the Diagnosis of Alzheimer's Disease. *IEEE Transactions on Medical Imaging, 31*(2), 207-216. 10.1109/TMI.2011.2167628

Palumbo, B., Fravolini, M. L., Buresta, T., Pompili, F., Forini, N., Nigro, P., & Tambasco, N. (2014, December). Diagnostic accuracy of Parkinson disease by support vector machine (SVM) analysis of 123I-FP-CIT brain SPECT data: implications of putaminal findings and age. *Medicine (Baltimore), 93*(27), e228. doi:10.1097/MD.0000000000000228

Peng, X., Lin, P., Zhang, T., & Wang, J. (2013, November). Extreme learning machine-based classification ofADHD using brain structural MRI data. *Plos One, 8*(11), e79476. 10.1371/journal.pone.0079476

Prashanth, R., Dutta Roy, S., Mandal, P. K., & Ghosh, S. (2016, June). High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning. *International Journal of Medical Informatics, 90*, 13-21. 10.1016/j.ijmedinf.2016.03.001

Pulini, A. A., Kerr, W. T., Loo, S. K., & Lenartowicz, A. (2019, February). Classification accuracy of neuroimaging biomarkers in attention-deficit/hyperactivity disorder: effects ofsample size and circular analysis. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging, 4*(2), 108–120. 10.1016/j.bpsc.2018.06.003

Qureshi, M., Oh, J., Min, B., Jo, H. J., & Lee, B. (2017, April). Multi-modal, Multi-measure, and Multi-class Discrimination of ADHD with Hierarchical Feature Extraction and Extreme Learning Machine Using Structural and Functional Brain MRI. *Frontiers in Human Neuroscience, 11*, 157. 10.3389/fnhum.2017.00157

Rondina, J. M., Ferreira, L. .., de Souza Duran, F. L., Kubo, R., Ono, C., Leite, C. C., & Busatto, G. F. (2017, November). Selecting the most relevant brain regions to discriminate Alzheimer's disease patients from healthy controls using multiple kernel learning: A comparison across functional and structural imaging modalities and atlases. *NeuroImage Clinical, 17*, 628-641. doi:10.1016/j.nicl.2017.10.02

Salas-Gonzalez, D., Segovia, F., Martínez-Murcia, F. J., Lang, E., Gorriz, J., & Ramırez, J. (2016). An Optimal Approach for Selecting Discriminant Regions for the Diagnosis of Alzheimer's Disease. *Current Alzheimer Research, 13*(7), pp. 838-844. 10.2174/1567205013666160415154852

Santra, A., & Kumar, R. (2014, October). Brain perfusion single photon emission computed tomography in major psychiatric disorders: From basics to clinical practice. *Indian Journal of Nuclear Medicine, 29*(4), 210–221. 10.4103/0972-3919.142622

Segovia, F., Górriz, J. M., Ramírez, J., Martínez-Murcia, F. J., & Castillo-Barnes, D. (2019, November). Assisted Diagnosis of Parkinsonism Based on the Striatal Morphology. *International Journal of Neural Systems, 29*(9), 1950011. 10.1142/S0129065719500114

Segovia, F., Górriz, J. M., Ramírez, J., Salas-González, D., Álvarez, I., López, M., & Padilla, P. (2010, April 19). Classification of functional brain images using a GMM-based multi-variate approach. *Neuroscience Letters, 474*(1), 58–62. 10.1016/j.neulet.2010.03.010

Sen, B., Borle, N. C., Greiner, R., & Brown, M. R. (2018, April). A general prediction model for the detection of ADHD and Autism using structural and functional MRI. *Plos One, 13*(4), e0194856. 10.1371/journal.pone.0194856

Shiiba, T., Arimura, Y., Nagano, M., Takahashi, T., & & Takaki, A. (2020, January). Improvement of classification performance of Parkinson's disease using shape features for machine learning on dopamine transporter single photon emission computed tomography. *Plos One, 15*(1), e0228289. 10.1371/journal.pone.0228289

Tagare, H., DeLorenzo, C., Chelikani, S., Saperstein, L., & Fulbright, R. K. (2017, May). Voxel-based logistic analysis of PPMI control and Parkinson's disease DaTscans. *Neuroimage, 152*, 299-311. 10.1016/j.neuroimage.2017.02.067

Vázquez-Abad, F. J., Bernabel, S., Dufresne, D., Sood, R., Ward, T., & Amen, D. (2020). Deep Learning for Mental Illness Detection Using Brain SPECT Imaging. *Medical Imaging and Computer-Aided Diagnosis. 633*, pp. 17-26. Singapore: Springer. 10.1007/978-981-15-5199-4_3

Wilcox, R. (2017). *Introduction to Robust Estimation and Hypothesis Testing* (4th ed.): Academic Press.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical machine Learning Tools and Techniques* (3rd ed.). Amsterdam: Morgan Kaufmann Series in Data Management Systems Morgan Kaufmann.

Xu, J., Kochanek, K. D., Sherry, L., Murphy, B. S., & Tejada-Vera, B. (2010, May). Deaths: final data for 2007. *National Vital Statistics Reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System, 58*(19), 1-19.

Zimmermann, R. (2017). *Nuclear Medicine: Radioactivity for Diagnosis and Therapy* (2nd ed.). EDP Sciences.

Zou, L., Zheng, J., Miao, C., McKeown, M. J., & Wang, Z. J. (2017). 3D CNN Based Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Functional and Structural MRI. *IEEE Access, 5*, 23626–23636. 10.1109/ACCESS.2017.2762703