Autonomous University of Queretaro

Department of Engineering

Master Science in Artificial Intelligence

Implementation of artificial intelligence algorithm for EEG signals classification.

A Thesis submitted in fulfilment of requirements for the degree of Master of Science of Autonomus University of Queretaro.

by:

Iván Alejandro García Amaya

Supervised by:

Dr. Marco Antonio Aceves Fernández

Querétaro, Qro. November 3, 2021

Autonomous University of Queretaro

Department of Engineering

Master Science in Artificial Intelligence

Implementation of artificial intelligence algorithm for EEG signals classification.

A Thesis submitted in fulfilment of requirements for the degree of Master of Science of Autonomus University of Queretaro.

by:

Iván Alejandro García Amaya

Supervised by Dr. Marco Antonio Aceves Fernández

Dr. Marco Antonio Aceves Fernández
President

Dr. Efrén Gorrostieta Hurtado
Secretary

Dr. Juan Manuel Ramos Arreguín
Vocal

Dr. Jesús Carlos Pedraza Ortega
Alternate

Dr. Saúl Tovar Arriaga
Alternate

Abstract

Attention Deficit Hyperactivity Disorder is one of the problems that affect a considerable part of society, in this research work a couple of methodologies are proposed to address this problem through the classifications of electroencephalogram signals. The first methodology consists of filtering the electroencephalogram signals to remove the greatest amount of noise and to extract the beta band. Then it is necessary to extract the Absolute Power of the electroencephalogram signals from the selected channels (Cz, C3, C4, Fz, Pz). The Absolute Power generates a smaller data set with the help of a Genetic Algorithm which selects the most representative attributes to be able to represent them in 2D images. In this case, with the help of the Continuous Wavelet Transform technique, a Power Spectrum represented in 2D images is obtained. The three Deep Learning Models that are implemented are ResNet-152, the ResNeXt-101 and the GoogLeNet. Similar results are observed with the implementation of these models; with the ResNet-152 model an accuracy of 86.6% is shown as a result, with the ResNeXt-101 model shows an accuracy of 83.92% and with the GoogLeNet model an accuracy of 85.74%. The second methodology for the approach to the classification of people diagnosed with Attention Deficit and Hyperactivity Disorder and people of Control making use of the electroencephalogram, proposes a filtering phase first where the noise from the signals is removed, while extracting the beta band and five different measurs are extracted: Standard Deviation, Variance, Entropy, Absolute Power and Relative Power. Each of these measurs provides relevant information. Two algorithms are implemented which are Logistic Regression and the Maximum Likelihood, these models are trained with 70% of the data set and 30% are used for validation that is randomly selected, this process is repeated several times and the best results obtained are 80.30% of accuracy in the Logistic Regression model and 84.40% in the Maximum

Likelihood model, greater extraction of characteristics translates into a greater amount of information, the correct extraction of information allows us to implement algorithms such as Logistic Regression and Maximum likelihood.

Acknowledgements

Firstly, I would like to gratefully acknowledge my supervisor Dr Marco Antonio Aceves Fernández for his enthusiastic supervision, valuable input, and guidance throughout this work.

I would also like to thank my second supervisor Dr Jesus Carlos Pedraza Ortega and Dr Saúl tovar Arriaga for their helpful advice and suggestions.

Finally. I would like to thank my family for their support, encouragement, and education, especially to my mother and my son for their understanding and support in this chapter of our lives

List of Abbreviations

ADHD Attention-Deficit and Hyperactivity Disorder

EEG Electroencephalogram

TSC Time-Series Classification

CWT Continous Wavelet Transform

SVM Support Vector Machine

MEFM Mixture of Expert Fuzzy Models

ANCOVA Analysis of Covariance

ANOVA Analysis of Variance

LDA Linear Discriminant Analysis

DLDA Direct Linear Discriminant Analysis

DISR Double Input Symmetrical Relevance

mRMR minimum Redundancy Maximum Relevance

PSD Power Spectral Density

PSDEDs Power Spectrum Density Energy Diagrams

LOFC Low Order Functional Connectivity

RP Recurrent Plot

OSA Obstructive Sleep Apnea

GA Genetic Algorithm

ReLU Rectified Linear Unit

ICA Independent Components Analysis

PCA Principal Components Analisys

MLP Multi Layer Perceptron

FD Fractal Dimension

ResNet Residual Network

CNN Convolutional Neural Networks

SCP Slow Cortex Potential

CNV Contingent Negative Variance

List of Nomenclatures

σ Standard deviation

μ Mean

σ^2 Variance

$\Psi(t)$ Complex-valued wavelet Morlet

$x(t)$ EEG signals

Z Z-Score

Contents

List of Figures

List of Tables

En ese instante gigante, he visto millones de actos deleitables o atroces; ninguno me asombró como el hecho de que todos ocuparan el mismo punto, sin superposición y sin trasparencia. Lo que vieron mis ojos fue simultáneo: lo que transcribiré, sucesivo, porque el lenguaje lo es. Algo, sin embargo, recogeré. (In that single gigantic instant I saw millions of acts both delightful and awful; not one of them amazed me more than the fact that all of them occupied the same point in space, without overlapping or transparency What my eyes beheld was simultaneous, but what I shall now write down will be successive, because language is successive. Nonetheless, I'll try to recollect what I can). Jorge Luis Borges, The Aleph, page 13. translate by Norman Thomas di Giovanni.

CHAPTER 1

Introduction

1.1 Background.

Attention Deficit Hyperactivity and Disorder (ADHD) have a great influence on brain activity, and the study of this process can be categorized mostly into 3 groups [1]. The first group focuses on the study of event-related potential (ERP), in which implement different types of tests and in these studies a difference is reported in the levels of activity in the parietal and frontal lobes, a second group focuses on Slow Cortex Potential (SCP). In this type of research it is mentioned that the Contingent Negative Variance (CNV) in subjects diagnosed with ADHD is lower than in control subjects, finally, a third group that focuses on data mining and neurofeedback, use EEG signals and emphasize that a subject diagnosed with ADHD has an increase in the power of the δ and θ waves and the lack of attention generates a decrease in the power of the α and β band.

In this research, the proposed approach is focused on the third group mentioned above [1], where it is intended to find the variation in the power of the different waves, this through different techniques which we will address throughout this investigation.

For the classification of EEG signals, the use of Support Vector Machines (SVM) has been applied to generate different results of accuracy, Helgadttir use 17 channels according 10-20 international system and SVM model with 79% of accuracy [2], Tenev applied 4 SVM models obtaining an accuracy of 82.3% [3], Abibuallaev applied power spectral density and SVM with a result of 98.25% , Reza Yahoobi has the two works with the highest accuracy, SVM with Stochastic Fractal Search with 98.25% of accuracy and Mixture of Expert Fuzzy Models (MEFM) with 99.01% [1] [4].

The Cz channel was used as the only channel in different works, using its θ/β ratio, Ogrim applied the ANCOVA statistical tool obtaining an accuracy of 58% [5], Snyder on the other hand applied ANOVA obtaining 89% [6], Monastra achieves better results applying ANOVA with 91% of accuracy [7].

Also Sadatnezhad applied Linear Discriminant Analysis (LDA) in addition to other classification techniques such as Direct Linear Discriminant Analysis (DLDA), the results obtained depend on certain conditions such as eyes open or eyes closed and noise ratio, with 10% of this and open eyes get 80.92% $\pm$ 16.23 of accuracy [8], Ahmadlou use one-way ANOVA to find the most discriminative features and LDA model for classification that give them an accuracy of 84.2% [9], Magee used 3 different clusters in these the Linear Regression model was implemented, mentions that collectively these three equations give an accuracy of 87% [10].

Mohammadi implemented a Multilayer Preceptron, to select the best characteristics, he implemented two different techniques Double Input Symmetrical Relevance (DISR) and minimum Redundancy Maximum Relevance (mRMR), obtained 93.65% accuracy[11], the repository generated in the research is the same as that used in this work, it is addressed in greater depth in the dataset section.

Author	Year	Measures	Classification Algorithm	Accuracy
Vicent Monastra	2001	θ / β ratio	ANOVA	91.00%
Christopher Magee	2005	Realtive and absolute δ, θ, α and β power	Logistic Regression	87.00%
Steven Snyder	2008	θ / β ratio	ANOVA	89.00%
Khadijeh Sadtnezhad	2011	Fractal-based, band power and wavelet coefficients	Fisher Linear Discriminant Analysis	80.92%
Geir Ogrim	2012	Absolute θ and β power	ANCOVA	85.00%
Mehran Ahmadlou	2012	Inter-electrode synchronization	Linear Discriminant Analysis	88.20%
Bedakh Abibullaev	2012	Ratio θ/α θ/β power, relative θ and relative β	Support Vector Machine	97.00%
Aleksandar Tenev	2013	Absolute δ, θ, α and β power	Support Vector Machine	82.30%
Halla Helgadóttir	2015	Multi-channels, optimal features	Support Vector Machine	79.00%
Reza Mohammadi	2016	Multi-channels fractal dimension	Multi-Layer Perceptron	93.65%
Reza Yaghoobi	2017	Multi-channels, optimal features	Mixture of Expert Fuzzy Model	99.01%
Reza Yaghoobi	2018	RQA, power and entropy	Support Vector Machine	98.25%
Our Work	2021	Absolute and Relative β power, Entropy, Standard Deviation , Variance	ResNet-152, ResNeXt-101, GoogLeNet, Maximum Likelihood	86.60%

Table 1.1: Previous work in the classification of EEG signals.

It is important to address the use of residual networks in the classification of EEG signals, regardless of the purpose of these investigations, there are some tools to address this issue. Bizopoulos made use of Signal2Image (S2Is), to convert signals to images, implemented different models, LeNet , AlexNet, VGGnet, ResNet, DenseNet and their variations, as shown in table 1.2, with the use of ResNet152 an accuracy of 74.1 % [12].

Junayed made use of power spectral density (PSD) to detect different sleep stages with the use of the ResNet model, obtaining a qualification accuracy of 87.8% in women and 83.7% in men. [13].

Shalbaf aimed to identify Schizophrenia using EEG signals, he applied continuous wavelet transform (CWT) method to convert the signals into images, then he made use of 4 different pre-trained CNNs models: AlexNet, ResNet-18, VGG-19 and Inception-v3 and then they used the output of these models like inputs of support vector machine (SVM) classifier with an accuracy of 98.60 % [14].

Yunyuan in his work for classification on people with epilepsy, the EEG signals are denoised by wavelet transform and then analyzed by power spectrum density (PSD), the two-dimensional images generated from PSD, are called power spectrum density energy diagrams (PSDEDs), These reflect the energy information of different frequency bands and as classification algorithms, he implements Inception-ResNet-v2, Inception-v3, and ResNEt152 obtaining an accuracy of 91.9%, 91.3%, 95% respectively [15].

Duan in his work for the diagnosis of Alzheimer using EEG signals, used Low Order Functional Connectivity (LOFC) as input for a ResNet-18 model with an accuracy of 98.33 % [16].

Hatami in his work makes a deep analysis in the use of Recurrent Plot (RP) to transform Time-Series Signals (TSC) to 2D images and use these as inputs in deep CNN classifier using the UCR archive dataset [17].

Author	Year	Application	2D Technique	Model	Accuracy
Paschalis Bizopoulos	2019	Epilepsy	S2Is	ResNet-152	74.10%
Md Junayed Hasan	2020	Different sleep stages	PSD	ResNet	85 ± 2.8 %
Yunyuan Gao	2020	Epilepsy	PSD	ResNet-152	95 %
Ahmad Shalbaf	2020	Schizophrenia	CWT	ResNet-18	97.60 ± 1.29 %
Feng Duan	2020	Alzheimer	LOFC	ResNet-18	98.33 %
Shaoqiang Wang	2020	Epilepsy	-	ResNeXt-50	91.50 %
Qiren Chen	2020	Obstructive Sleep Apnea	PSD	Mr-ResNeXt	94.23 %

Table 1.2: Previous work in the classification of EEG signals using ResNet-ResNeXt model.

In the use of the ResNeXt model, for the classification of EEG signals, there are several works, for example for the classification of epilepsy [18], where consistent results have been obtained. In this work, the EEG signals are used as training input in the model, leaving aside the techniques to convert them into 2D images, showing good results since it mentions that by applying this technique, it achieves a correct classification with 91.50% accuracy. On the other hand, work carried out by [19] the Mr-ResNeXt model was implemented to address the problem of detecting Obstructive Sleep Apnea (OSA), which he mentions is one of the most frequent sleep problems, the work makes use of the Spectrum to convert the Time Series to 2D images, resulting in an accuracy of 91.91%

1.2 Hypothesis.

It is feasible to implement models of the Artificial Intelligence for the classification of EEG signals in ADHD subjects using the different features.

1.3 Objectives.

- Acquire dataset of subjects diagnosed with ADHD.
- Pre-process data.
- Select models of Artificial Intelligence.
- Train-Evaluate and Test Model.
- Compare with existing works the feasibility of the present contribution.

CHAPTER 2

Pre-processing

2.1 Proposed Methodology.

The figure shows a flow diagram of the methodology implemented in this work. Each of the stages will be addressed in greater depth in the present document, a deeper description of the implemented database is made, then the steps applied for a pre-processing, the methodology implemented to convert Time Series to 2D images, the Deep Learning models, train validation and test of these models.

The second methodology proposes a different approach, like the first methodology, the signals are filtered, the β band is extracted, only 5 channels are used (C3, C4, Cz, Fz, Pz), the Absolute Power feature is extracted, in addition to the Standard Deviation , Variance, Entropy and Relative Power, this generates a data set that is used to train the Logistic Regression and Maximum Likelihood models.

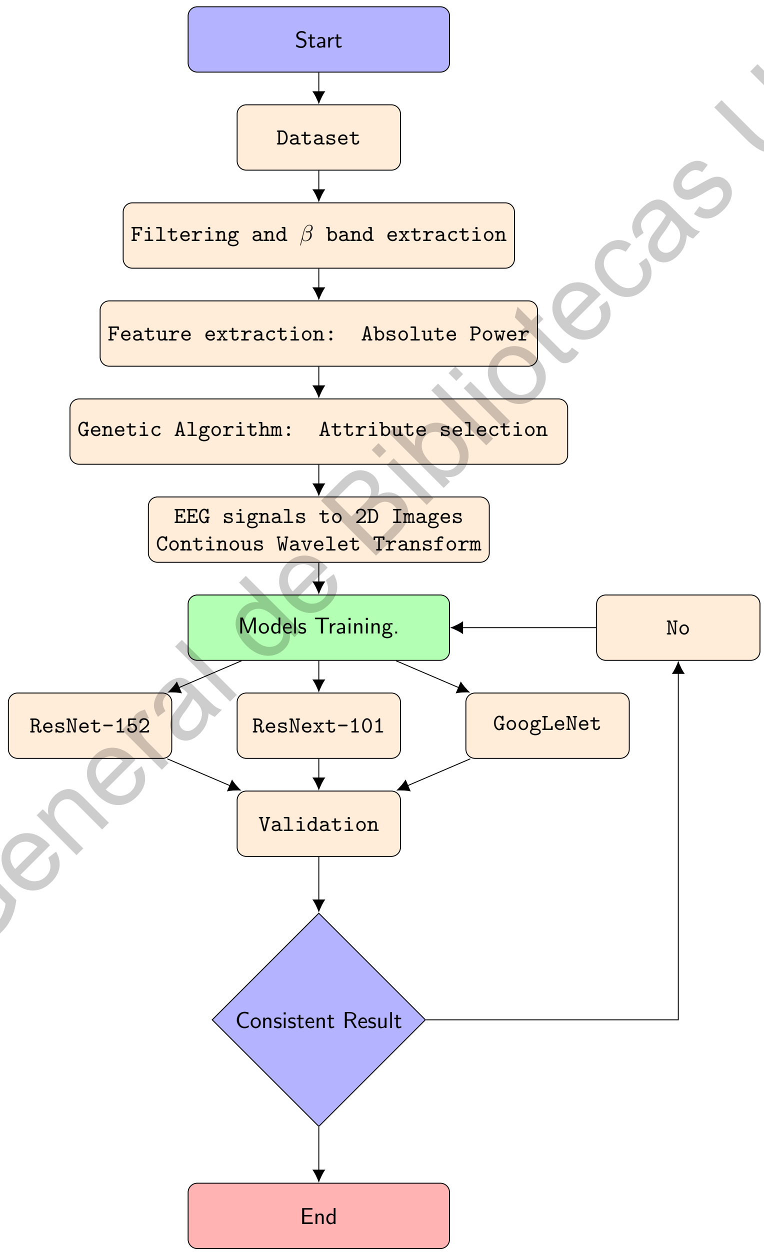


Figure 2.1: Proposed Methodology.

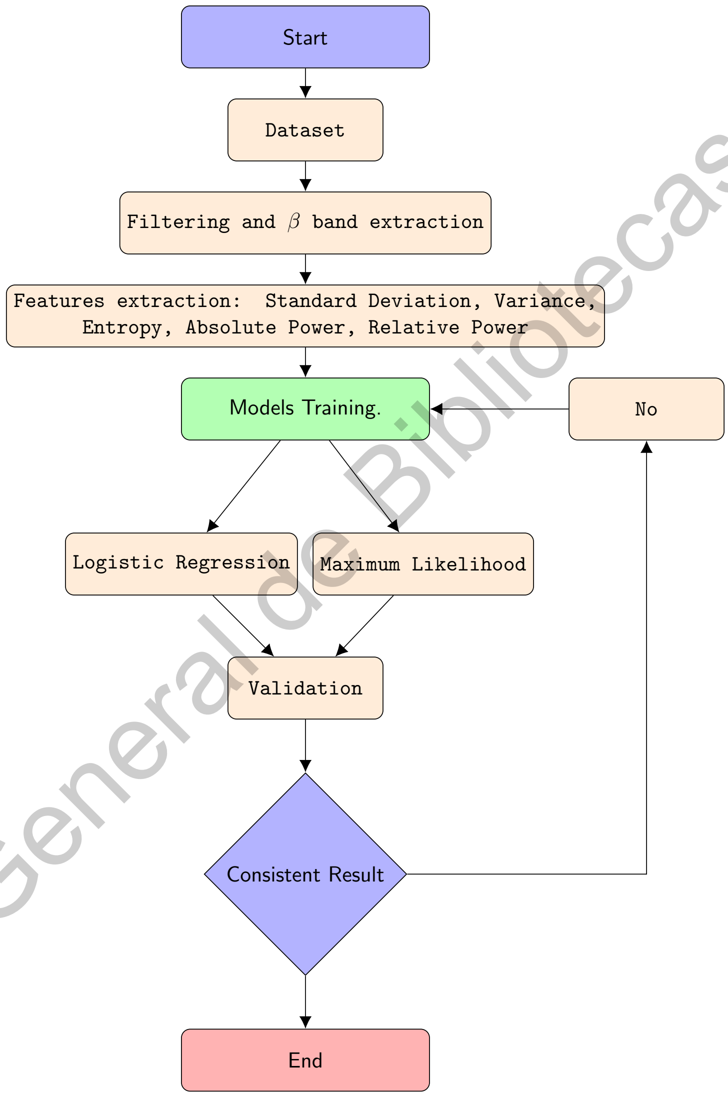


Figure 2.2: Proposed Methodology 2.

2.2 Dataset.

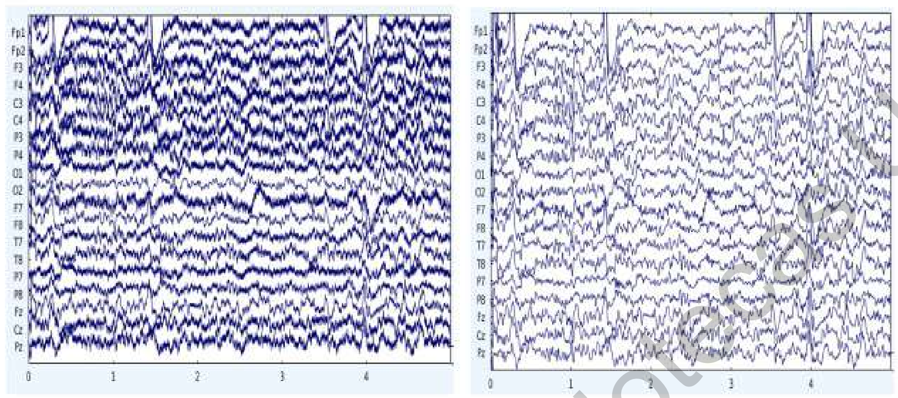
In this work, the data set used is a set of EEG signals, these time series are made up of 60 people diagnosed with ADHD (9.62 ± 1.75 years) and 60 healthy control people (9.85 ± 1.77 years),[11] , correspond to 19 channels based on the 10-20 standard (Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2), with two reference electrodes A1, A2 placed on the earlobes [20].

2.3 Filtering.

Filtering can be done using EEGLAB in Matlab, for example, removing non-physiological artifacts, physiological artifacts, among others.

It is necessary to add certain characteristics such as Sampling rate, in this case the 128Hz rate, in addition to adding the location of the respective channels, in order to have a clearer representation of the data.

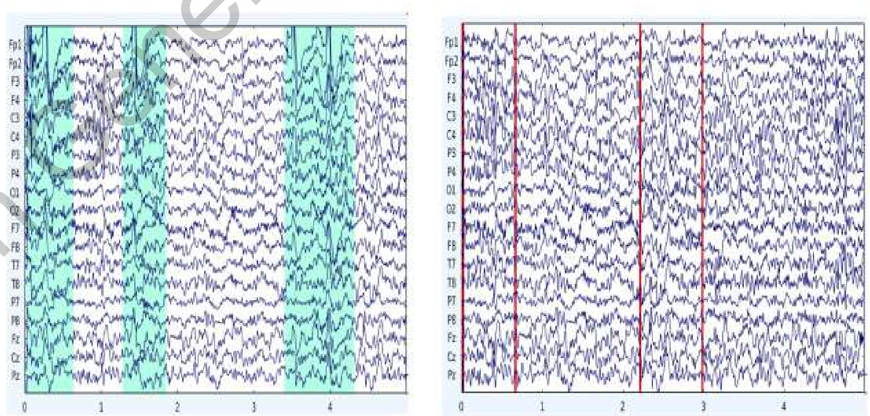
One of the non-physiological artifacts may include a 1-60 Hz filter to remove the embedded noise buried in the signal fabricated by the power supply as show on figure 2.3



a) b)

Figure 2.3: a) Unfiltered original signals. b) Signals after applying the 1-60 Hz filter.

The figure 2.4, shows the parts that are removed by excess noise.



a) b)

Figure 2.4: a) Signals selected for be removed. b) Signals after selected parts removed.

2.4 EEG signals

The EEG signals can be decomposed into different frequency bands; δ (0.5-4Hz), θ (4-8Hz), α (8-13), and β (13-30), to obtain their respective band power filter by bandpass filters [21]. These features were obtained from the Fz, Pz, Cz, C3, and C4 channels, in addition statistical features such as standard deviation, variation and mean [1].

In statistics it is known that the measure of the variability of a set of numbers is the standard deviation, and is given by the equation:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{2.1}$$

Where N is the number of elements in the data set x_i the element at index i and μ the mean given by

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2.2}$$

We know that the variance is the square of the standard deviation and is given by the equation:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \tag{2.3}$$

The figure 2.5 shows two different signals from the same channel (Fz) from two different subjects, one from ADHD and the other from Control, we can observe a difference in the standard deviation between them.

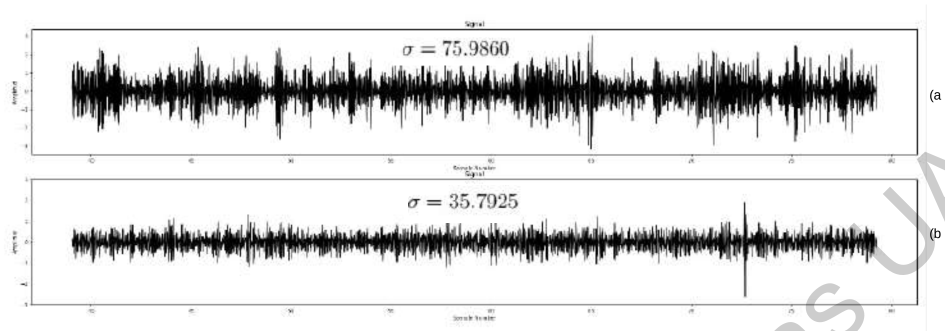


Figure 2.5: a) Signal of the Fz channel of the data set 109 Control b) Signal of the Fz channel of the data set v27 ADHD.

2.4.1 Power Spectral Entropy

The spectral complexity can be quantify by power spectral entropy that is a information entropy of an uncertain system, assume a set of signals X, the value of X is [22]

$$X = \{x_1, x_2, ..., x_n\} \tag{2.4}$$

The corresponding probability is

$$P = p_1, p_2, ..., p_n 0 \leq p_i \leq 1, i = 1, 2, ..., n \tag{2.5}$$

Under consideration

$$\sum_{i=1}^{n} p_i = 1 \tag{2.6}$$

The information entropy of the system is given by

$$H = -\sum_{i=1}^{n} p_i ln p_i \tag{2.7}$$

2.4.2 Average Power.

The average power in $y(n)$ is

$$E\{|y(n)^2|\} = y_y(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_y(e^{jw})dw \tag{2.8}$$

then we have

$$E\{|y(n)^2|\} \approx \frac{\triangle w}{2\pi} P_x(e^{jw_0}) \tag{2.9}$$

then, $P_x(e^{jw})$ may be viewed as a density function that describes how the power in $x(n)$ varies with w [23].

2.4.3 Relative Power.

The relative power of its respective band given band / sum of power from 12 to 30 Hz is calculated by:

$$RP(f_1, f_2) = \frac{P(f_1, f_2)}{P(12, 30)} 100\% \tag{2.10}$$

where P (.) indicates power, RP (.) relative power, and f1, f2 indicate low and high frequency respectively [24].

2.4.4 Normalize data.

The max-min normalization is given by:

$$X_{scaled} = \frac{x - min(x)}{max(x) - min(x)} \tag{2.11}$$

2.4.5 Data distribution.

In the figure 2.6, the distribution of the characteristics extracted from the EEG signals is observed, these belong to the two different groups are ADHD and Control.

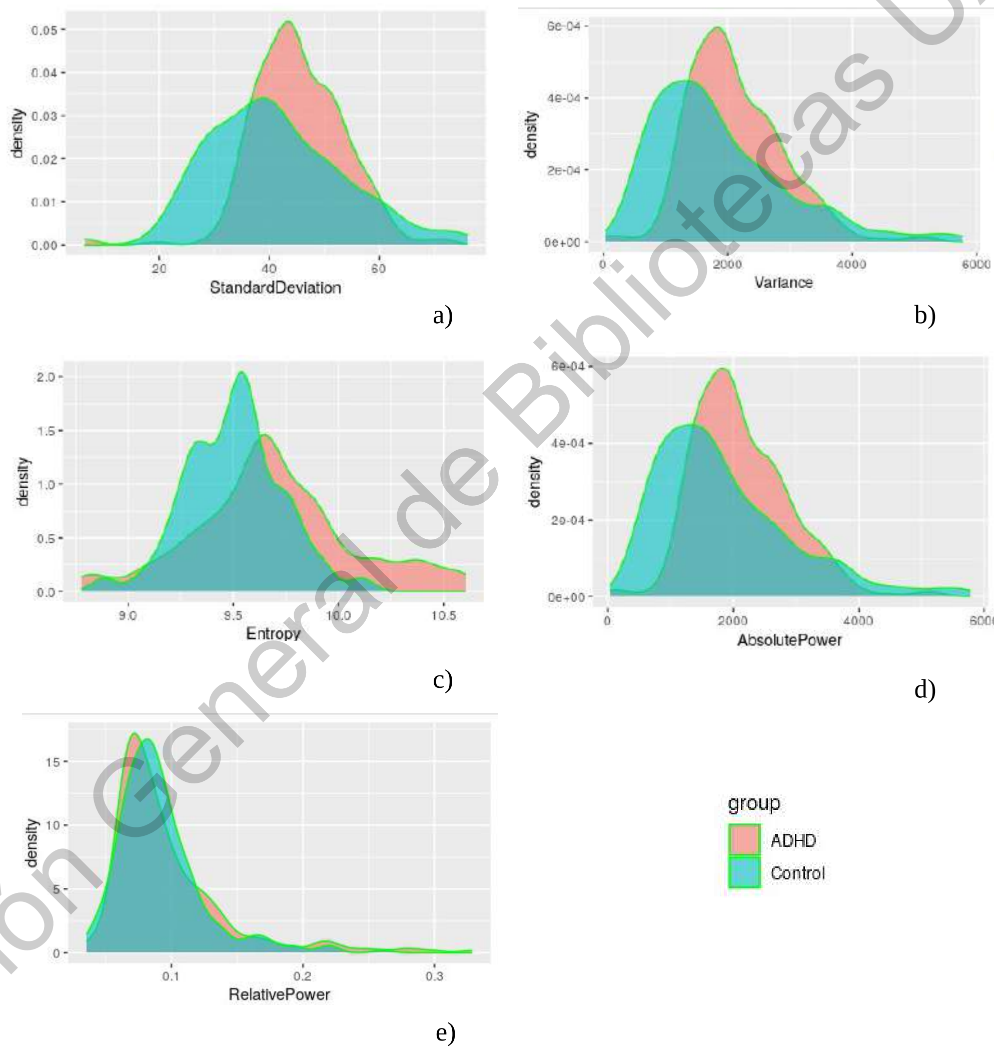


Figure 2.6: a) Density distribution of Standard Deviation . b) Density distribution of Variance . c) Density distribution of Entropy. d) Density distribution of Absolute Power . e) Density distribution of Relative Power.

2.5 Genetic Algorithm

The use of genetic algorithms has been successfully applied to solve different combinatorial problems, where it is investigated to find an approximation to the best solution [25] [26]. This research has implemented a GA to find the characteristics that generate the best performance in the classification algorithm. The implemented chromosomes consist of ADHD and Control characteristics, through the components of this algorithm such as an Initial Population, followed by a Random Selection, Crossover and Genetic Competition as an Elitism technique to ensure the best cromosomes are selected, obtaining a new generation of better solutions and finally repeating this process a given number of iterations in order to obtain an approximation to the possible best solutions. A combinatorial optimization problem can be solved by making an approximation with a genetic algorithm, in this work a simple algorithm is proposed.

The problem is to find x_1 and x_2 such that we maximize the function $f(x)$ given by.

$$f(x) = x_1 - x_2 \tag{2.12}$$

where x_1 is the features of Control subjects and x_2 the features of ADHD subjects.

2.5.1 Initial Population.

The possible number of solutions it's given by

$$Ps = \frac{1}{N_a} \sum_{i=1}^{n} x_{1i} \tag{2.13}$$

Where Na is the numbers of Control attributes per Chromosome, subsequently a Chromosome can be described as seen on equation 2.14

$$F = [a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4] \text{ Where } a_i \in x_1 \text{ and } b_j \in x_2 \tag{2.14}$$

2.5.2 Cross Over.

The cross over technique consist in for the two principal chromosomes, create all possible combinations, evaluate it and select the two best without repeated values.

$$F_1 = [a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4]$$

$$F_2 = [a_5, a_6, a_7, a_8, b_5, b_6, b_7, b_8]$$

$$S_1 = [a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8]$$

$$S_2 = [b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8]$$

After this, is necessary take S_1 , S_2 and evaluate it then back a sort chromosome in decedent order, the first 4 of S_1 will be the best and the other 4 the worst, and oppositive S_2 the first will be the worst and the last the best, finally we take the best of them by selecting B_1, B_2.

2.6 Continuous Wavelet Transformation.

If $x(t)$ is a EEG signal, t, then its CWT is defined as [27] :

$$C(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \Psi^* \frac{t-b}{a} dt \tag{2.15}$$

Algorithm 1: Genetic Algorithm.

input : Data set.

output: Best attributes.

$X2 \leftarrow \{\}$;

$X \leftarrow InitialPopulation$;

$l \leftarrow length(InitialPopulation)$;

$Generations \leftarrow 100$;

for $j \rightarrow Generations$ **do**

    for $i \rightarrow l$ **do**

        $F_1, F_2 \leftarrow Selection(X)$;

        $S_1, S_2 \leftarrow CrossOver(F_1, F_2)$;

        $B_1, B_2 \leftarrow Competition(F_1, F_2, S_1, S_2)$;

        $X_j \leftarrow B_1, B_2$;

    end

    $X \leftarrow X2$;

end

$best\_attributes \leftarrow X$;

where $a, b \in \mathbb{R}$, $a \neq 0$, and $\mathbb{R}$ is the set of real numbers, a is the dilation parameter called 'scale' and b is the location parameter of the wavelet, $\Psi(t)$ is the wavelet function called the "mother wavelet", superscript "*" denotes the complex conjugate of the function, and $\frac{1}{\sqrt{a}}$ is used to normalize the energy such stays at the same level for different values of a and b.

Complex-valued wavelet Morlet function is defined by:

$$\Psi(t) = \pi^{-\frac{1}{4}} e^{iw_0 t} e^{-\frac{t^2}{2}} \tag{2.16}$$

Where $\Psi(t)$ is the wavelet function that depends on a non dimensional time parameter t and i denotes the imaginary unit. this wavelet function forms two exponential functions modulating a Gaussian envelope of unit width, where the parameter w_0 is the non-dimensional frequency parameter.

The figure 2.7 shows the Power Spectrum of the Control and ADHD subject v109, specifically of the Fz channel.

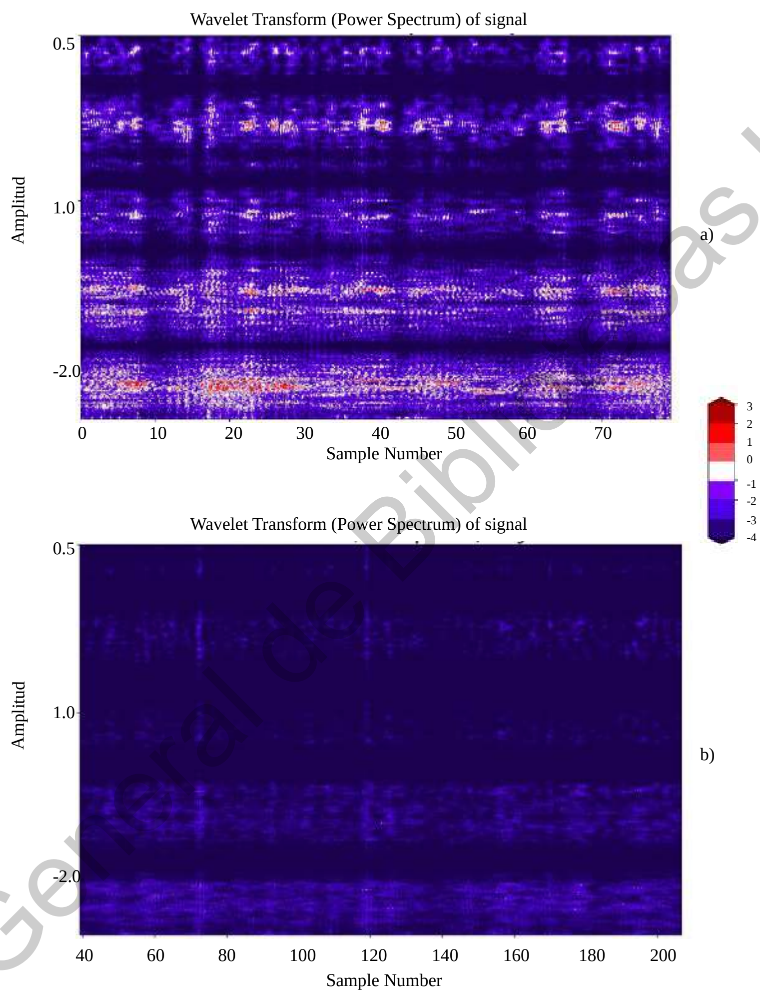


Figure 2.7: a) CWT Power Spectrum Control. b) CWT Power Spectrum for ADHD patients.

CHAPTER 3

Models

3.1 Deep Learning Models

The models implemented in this research are three: ResNet-152, ResNeXet-101, and GoogLeNet, these share certain characteristics, such as the execution environment, in this case, Google Collaboratory that offers a GPU (K80s, T4s, P4s, and P100s), They are all pre-trained models in Pytorch, they also share values such as learning rate = 0.001, batch size = 4, SGD Optimizer.

3.1.1 Pretrained networks.

PyTorch offers the most relevant pre-trained models, these models are pre-trained with the ImagenNet-12 dataset [28], this dataset is a very large one, approximately 14 million images maintained by Stanford University. (http://imagenet.stanford.edu).

3.2 Residual Network 152 (ResNet-152).

This is a model of 152 layers of deep [29], introduce the deep residual nets that is easier to optimize, this model won the first place on the ILSVRC 2015 classification task, has 11.3 billion Floating point operations per second (FLOPs), the figure 3.1, is a representation of the ResNet-152 architecture.

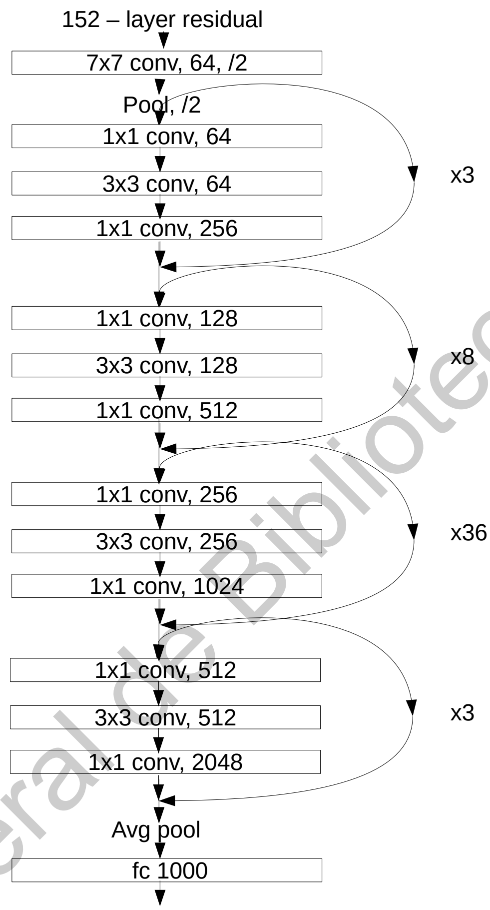


Figure 3.1: Topology ResNet-152 [29].

3.2.1 Residual learning.

It is necessary to consider $H(x)$ as the underlying mapping [29], of a set of blocks, if x denotes the inputs of the first layers, an approximation of the residual function can be made by

$$H(x) - x \tag{3.1}$$

This assuming that the inputs and outputs are of the same dimension. So instead of the stacked layers approaching $H(x)$, they can be allowed to approach a residual function.

$$F(x) := H(x) - x \tag{3.2}$$

The original function is:

$$F(x) + x \tag{3.3}$$

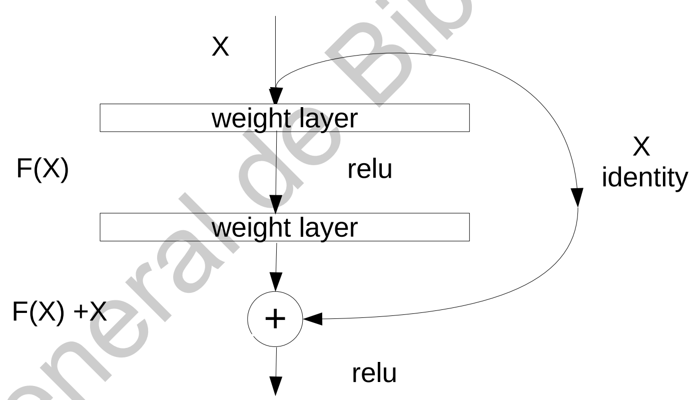


Figure 3.2: Residual learning: a building block [29].

3.3 ResNeXt-101(101_32x8d).

This model has 101 deep layers [30] and exposes a new strategy of dimension that is called "cardinality", which describes the size of the set of transformations, which is an important factor in the dimension of depth and width, this proposes that the increasing cardinality is more effective than going deeper or wider, this model won the second place in ILSVRC 2016 classification task.

A family of Inception models that achieve convincing accuracy with low theoretical complexity is presented [31] [32] [33]. These models have an important common property called split-transform-marge strategy. In this model, the input is divided into fewer lower-dimension embeddings by 1x1 convolutions and transformed by a set of functional filters 3x3, 5x5 and so on, united by concatenation.

The VGG-nets [34] propose stock building blocks for deep networks all in the same shape and ResNet [29] stock modules of the same topology. The ResNeXt architecture adopts VGG/ResNets strategies.

Figure 3.3, shows a great similarity with the figure 3.1, except for some differences. For instance, C = 32, this means, cardinality = 32, this can be seen in greater detail in the figure 3.4, in addition to showing a difference in the input and output channels, these being greater than those of the ResNet-152 and the difference in depth that one has 152 layers whilst the ResNeXt-101.

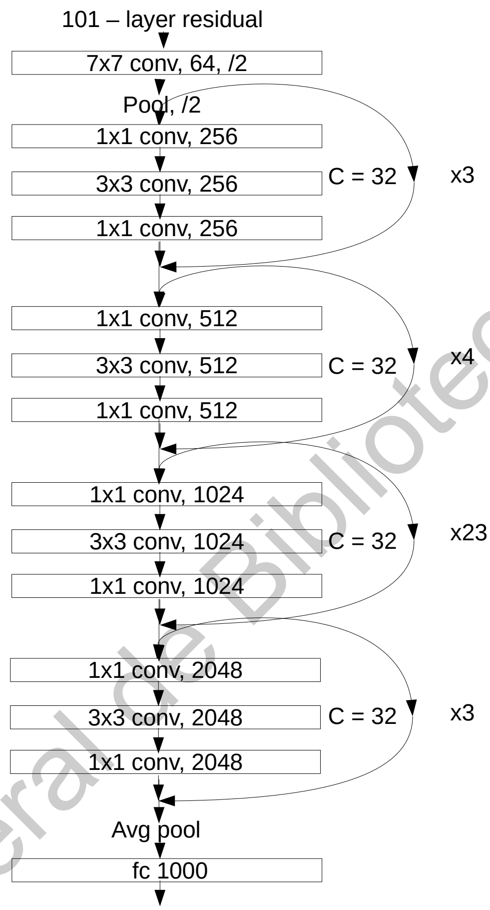


Figure 3.3: Topology ResNeXt-101 [30].

3.3.1 Inception model

As mentioned above, cardinality = 32, means that we have a parallel distribution of 32 paths as seen in figure 3.4 section c), we also observe that the number of channels is 8, as its name mentions (ResNeXt101 (101_32x8d)) is a multiplication of 8x32 = 256.

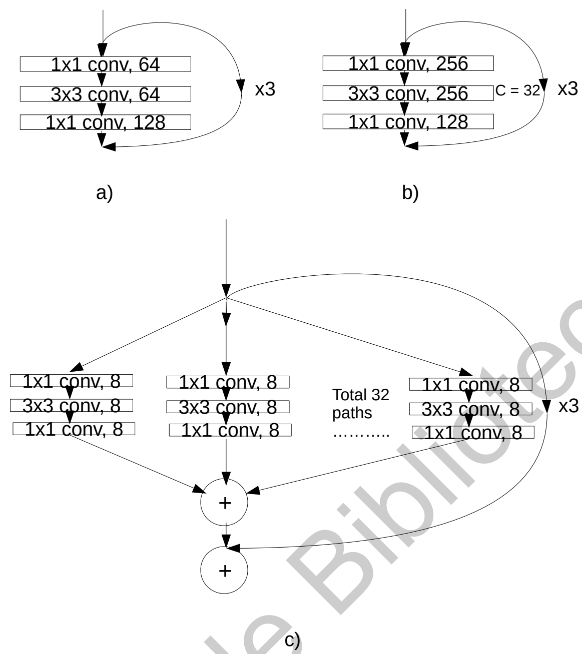


Figure 3.4: a) Block from ResNet-152, b) Block from ResNeXt-101 c) Parallel ResNeXt-101 [30].

3.4 GoogLeNet

The last model is the GoogLeNet [32] and is a pre-trained modle in Pytorch like the ones seen in the last section with learning rate = 0.001, batch size = 32 , SGD optimizer, in Google Colab that often include GPU Nvidia, K80s, T4s, P4s and P100s.

GoogLeNet is a 22 layers deep network, the complete architecture can be seen in greater detail in [32] , but one of the characteristic features of this neural network is the use of Inception module like the one observed in the figure 3.5.

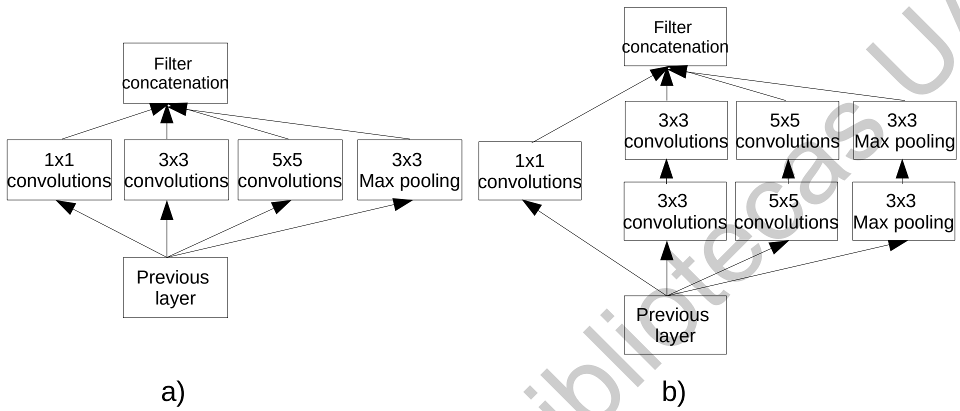


Figure 3.5: a) Inception module, naive version , b) Inception module with dimension reductions [32].

The architecture GoogleNet is composed of 9 different inceptions modules, it also shows us that certain datasets are a bit tricky, this is of great help when implementing this model in our dataset as the power spectrum images, in some cases it is difficult to visually recognize the difference [32].

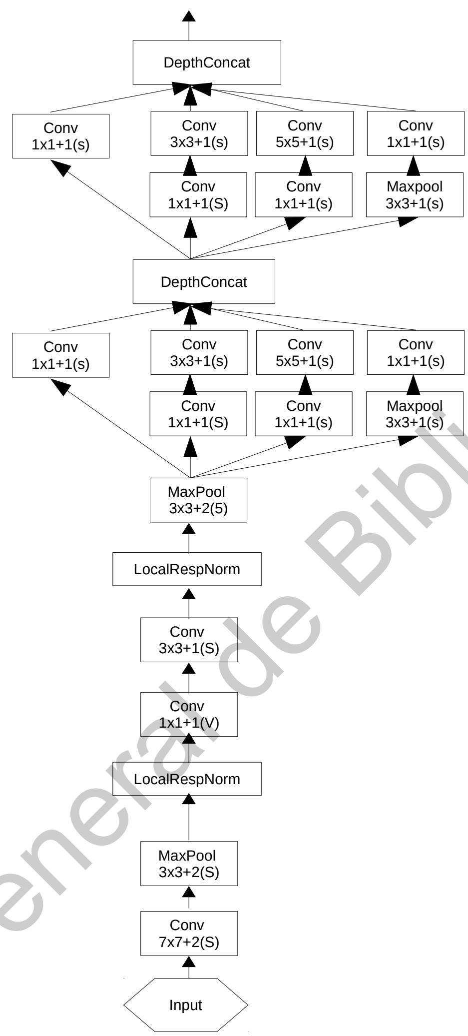


Figure 3.6: GoogLeNet Part1 [32].

3.5 Model training, validation and testing.

In the table 3.1, the different data sets are shown, which correspond to different individuals, column 1 shows the name of the data set used, column 2 the corresponding channel, and column 3 shows the section in which was implemented, this means training, validation or testing, the table, shows only data from people with ADHD.

It is important to mention that for this work three different models are implemented, the table shown below is used for all the models.

Data	Channel
v24pO.txt	Pz
v34pO.txt	C3
v29pO.txt	Fz
v31pO.txt	Fz
v34pO.txt	C4
v10pO.txt	Fz
v28pO.txt	Pz
v21pO.txt	Cz
v32pO.txt	Cz
v25pO.txt	Pz
v15pO.txt	Fz
v21pO.txt	Pz
v32pO.txt	C3
v28pO.txt	C4
v34pO.txt	Cz
v10pO.txt	Cz
v30pO.txt	Pz
v25pO.txt	Fz
v29pO.txt	C3
v35pO.txt	Cz
v34pO.txt	Pz
v35pO.txt	C4
v27pO.txt	Pz
v25pO.txt	C3
v30pO.txt	C3
v21pO.txt	C3
v19pO.txt	C4
v27pO.txt	C3
v15pO.txt	C3
v19pO.txt	Pz
v27pO.txt	C4
v19pO.txt	C3
v27pO.txt	Cz
v15pO.txt	Pz
v12pO.txt	Fz
v19pO.txt	Cz
v27pO.txt	Fz
v34pO.txt	Fz
v25pO.txt	C4
v21pO.txt	C4
v31pO.txt	C3
v21pO.txt	Fz

Table 3.1: Data sets used in the train, validation, and test of subjects diagnosed with ADHD.

Similar to the table 3.1, the table 3.2 shows similar columns, the difference is that the latter shows the data used from Control subjects.

Data	Channel
v109pO.txt	Fz, Pz
v111pO.txt	Pz
v60pO.txt	Fz, Pz
v42pO.txt	Cz
v59pO.txt	Pz
v60pO.txt	C3, Pz
v55pO.txt	Fz, Cz
v44pO.txt	Fz, C4
v58pO.txt	Pz
v44pO.txt	Pz
v58pO.txt	C4
v60pO.txt	C4
v47pO.txt	Fz
v60pO.txt	Cz
v55pO.txt	C3
v58pO.txt	Fz
v58pO.txt	Cz
v47pO.txt	C4
v47pO.txt	Cz
v58pO.txt	C3
v47pO.txt	C3
v59pO.txt	Fz
v53pO.txt	C4
v55pO.txt	C4
v42pO.txt	C3
v52pO.txt	Cz
v47pO.txt	Pz
v53pO.txt	Fz
v52pO.txt	Pz
v45pO.txt	C4
v52pO.txt	C4
v42pO.txt	C4
v53pO.txt	C3
v53pO.txt	Cz
v109pO.txt	C4
v42pO.txt	Pz
v44pO.txt	C3
v44pO.txt	Cz
v109pO.txt	Cz
v55pO.txt	Pz
v111pO.txt	Fz
v111pO.txt	Cz
v111pO.txt	C3

Table 3.2: Data set used in the training, validation, and test of Control subjects.

3.5.1 Error metrics.

The performance of the models was evaluated using the corresponding metrics [35];

Accuracy is the fraction of correct predictions, it is formally given by equation 3.12.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{3.4}$$

Precision is the fraction of relevant predictions between the retrieved predictions given by equation 3.13.

$$Precision = \frac{tp}{tp + fp} \tag{3.5}$$

Recall (sensitivity) is the fraction of relevant predictions that were retrieved (equation 3.14).

$$Recall = \frac{tp}{tp + fn} \tag{3.6}$$

F1 is the harmonic mean of precision and recall (equation 3.15).

$$F1 = (2)\frac{(Precision)(Recall)}{Precision + Recall} \tag{3.7}$$

and F2-score that is more used in the area of healthcare is given by:

$$F2 = (1 + 2^2)\frac{PrecisionxRecall}{(1 + 2^2)(Precision + Recall)} \tag{3.8}$$

3.6 Logistic Regression.

The Logistic Regression model we use the logistic function that is given by the equation [36].

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \tag{3.9}$$

3.6.1 Multiple Logistic Regression.

In this work it is intended to make a binary classification using multiple predictions, taking this into account, then we can define a multiple regression by the equation [36]

$$log(\frac{p(X)}{1-p(X)}) = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p \tag{3.10}$$

where $X = (X_1, ..., X_p)$ are p predictors, then the last equation can be written as

$$p(X) = \frac{e^{\beta_0+\beta_1 X_1+....+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+....+\beta_p X_p}} \tag{3.11}$$

3.7 Maximum likelihood.

The approach to data analysis is based on the mean of the distribution represented in equation 3.12, which is the standard form of the Gaussian distribution [37].

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.12}$$

Where $P(x)$ is the probability of occurrence of the variable n, σ is the standard deviation σ^2 is the variance, μ is the mean.

The method is based on the postulate that the values of the unknown parameters are those that produce a maximum probability of observing the measured data. Assuming the measurements are independent of each other, the quantity:

$$P = \prod_{i=1}^{N} P(x_i) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x_i-\mu^2)}{2\sigma^2}} \tag{3.13}$$

The equation 3.13 can also be expressed as follows:

$$L(\mu, \sigma | x_i, x_2, ..., x_n) = L(\mu, \sigma | x_i) x, ...x L(\mu, \sigma | x_n) \tag{3.14}$$

Substituting the equation 3.12 we obtain:

$$\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{(x-\mu)^2}{2\sigma^2}}x, ..., x\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.15}$$

Applying ln to the equation 3.14 and 3.15

$$ln(L(\mu,\sigma|x_i)x, ...xL(\mu,\sigma|x_n)) = ln[\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{(x-\mu)^2}{2\sigma^2}}x, ..., x\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{(x-\mu)^2}{2\sigma^2}}] \tag{3.16}$$

Once ln is applied, we can express the equation 3.16, as follows.

$$-\frac{1}{2}ln(2\pi) - ln(\sigma) - \frac{(x_i-\mu)^2}{2\sigma^2} \tag{3.17}$$

then

$$-\frac{1}{2}ln(2\pi) - ln(\sigma) - \frac{(x_i-\mu)^2}{2\sigma^2} - .... - \frac{1}{2}ln(2\pi) - ln(\sigma) - \frac{(x_i-\mu)^2}{2\sigma^2} \tag{3.18}$$

simplifying we obtain.

$$-\frac{n}{2}ln(2\pi) - nln(\sigma) - \frac{(x_1-\mu)^2}{2\sigma^2} - .... - \frac{(x_n-\mu)^2}{2\sigma^2} \tag{3.19}$$

To find the maximum Likelihood it is necessary to find its derivative where the slope is zero.

$$-\frac{\partial}{\partial\mu}ln[L(\mu,\sigma|x_1, ..., x_n)] \tag{3.20}$$

Then applying the derivative to the equation 3.20 we obtain:

$$\frac{1}{\sigma^2}[(x_1+, ..., +x_n) - n\mu] \tag{3.21}$$

Now it is necessary to obtain the derivative of Likelihood with respect to sigma.

$$-\frac{\partial}{\partial\sigma} ln[L(\mu, \sigma|x_1, ..., x_n)] \tag{3.22}$$

Applying the derivative and simplifying we obtain.

$$\frac{n}{\sigma} + \frac{1}{\sigma^3}[(x_1 - \mu)+, ..., +(x_n - \mu)^2] \tag{3.23}$$

So knowing that the maximum Likelihood is when the slope is zero, we set the equations 3.21, 3.23 equal:

$$0 = \frac{1}{\sigma^2}[(x_1+, ..., +x_n) - n\mu] \tag{3.24}$$

$$0 = \frac{n}{\sigma} + \frac{1}{\sigma^3}[(x_1 - \mu)+, ..., +(x_n - \mu)^2] \tag{3.25}$$

So by solving the equation, we have that the maximum Likelihood estimated by μ is the mean of the data.

$$\mu = \frac{x_1 + ... + x_n}{n} \tag{3.26}$$

Solving for the equation, we have the maximum Likelihood estimate given by σ is the standard derivation of the data.

$$\sigma = \sqrt{(}\frac{(x_1 - \mu)^2 + ... + (x_n - \mu)^2}{n}) \tag{3.27}$$

Algorithm 2: Maximum Likelihood.

input : train-test Data

output: probability of occurrence

$maxlikehood \leftarrow function(train, test)\{$

$\quad x_i \leftarrow train$

$\quad y_i \leftarrow test$

$\quad n \leftarrow length(x)$

$\quad \mu \leftarrow \frac{1}{n}\sum_{i=1}^{n} x_i$

$\quad \sigma \leftarrow \sqrt{\sum_{i=1}^{n}(x_i - m)/n}$

$\quad p(y_i) \leftarrow \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{y_i-\mu^2}{2\sigma^2}}$

$\}$

$prob\_oc \leftarrow maxlikelihood(train, test)$

CHAPTER 4

Results

4.1 Results ResNet-152

The first model to be trained, evaluated and tested, is the ResNet-152 model, in the table 4.1, the values obtained by different types of tests, both Control and ADHD, are observed.

Test	Label	tp	fp	tn	fn
test	ADHD	17	1	–	–
test1	ADHD	6	1	–	–
test2	ADHD	4	1	–	–
test3	ADHD	8	3	–	–
test4	ADHD	15	0	–	–
test	Control	–	–	9	0
test1	Control	–	–	12	1
test2	Control	–	–	8	4
test3	Control	–	–	9	0
test4	Control	–	–	9	4

Table 4.1: Results of test ResNet-152

The figure 4.1 shows the confusion matrix of the results obtained, which can be seen in the table 4.1, in this way it is clearer to have a visualization of the results.

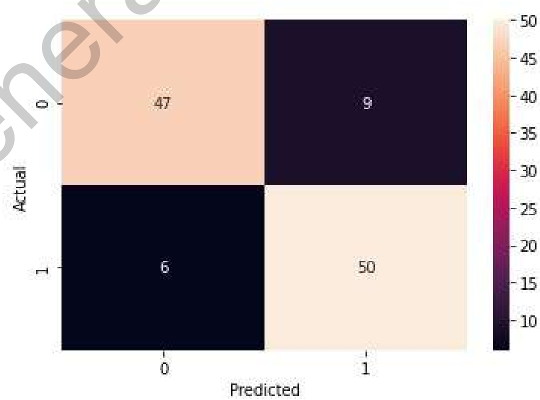


Figure 4.1: Confusion Matrix of ResNet-152

4.2 Results ResNeXt-101

The second model to be tested is the ResNeXt-101 model, the same as the first model, the results are shown in a table, in this case, the table 4.2.

Test	Label	tp	fp	tn	fn
test	ADHD	13	5	–	–
test1	ADHD	6	1	–	–
test2	ADHD	5	0	–	–
test3	ADHD	7	4	–	–
test4	ADHD	9	6	–	–
test	Control	–	–	9	0
test1	Control	–	–	12	1
test2	Control	–	–	11	1
test3	Control	–	–	9	0
test4	Control	–	–	13	0

Table 4.2: Results of test ResNeXt-101

We can see in figure 4.2, the confusion matrix of the data exposed in the table 4.2.

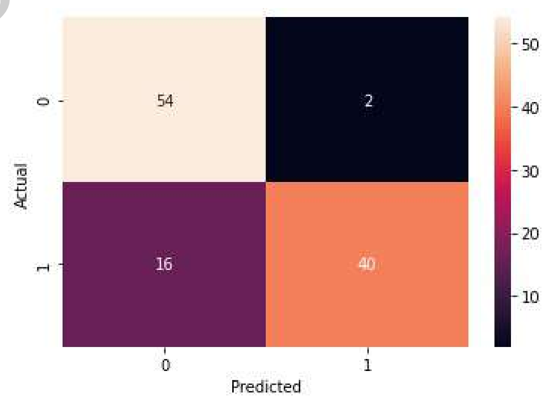


Figure 4.2: Confusion Matrix of ResNeXt-101

4.3 Results GoogLeNet

The last model is GoogLeNet, following the same procedures we can observe the results obtained in the table 4.3.

Test	Label	tp	fp	tn	fn
test	ADHD	15	3	–	–
test1	ADHD	7	0	–	–
test2	ADHD	5	0	–	–
test3	ADHD	7	4	–	–
test4	ADHD	12	3	–	–
test	Control	–	–	9	0
test1	Control	–	–	12	1
test2	Control	–	–	8	4
test3	Control	–	–	9	0
test4	Control	–	–	12	1

Table 4.3: Results of test GoogLeNet

The figure 4.3 shows the confusion matrix of the results obtained in this last test.

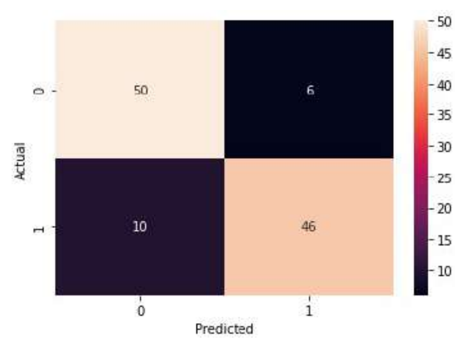


Figure 4.3: Confusion Matrix of GoogLeNet

The figure 4.4 shows the different graphs during the training and validation of the three proposed models, we observe in section a) the accuracy by epoch during training, b) it shows the same but in the validation stage, c) is the loss by epoch during training and d) during validation.

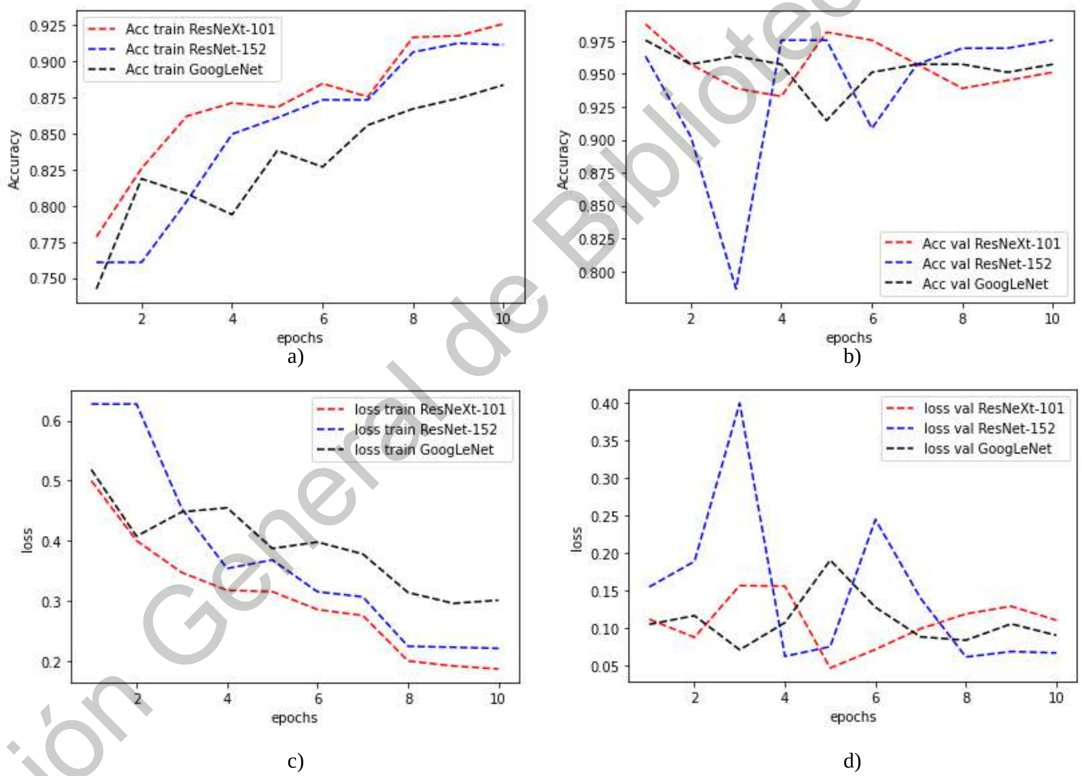


Figure 4.4: a) Train Accuracy of ResNeXt-101, ResNet-152 and GoogLeNet. b)Validation Accuracy of ResNeXt-101, ResNet-152 and GoogLeNet. c) Train Loss of ResNeXt-101, ResNet-152 and GoogLeNet d)Validation Loss ResNeXt-101, ResNet-152 and GoogLeNet.

The models were evaluated with the error metrics technique that was mentioned earlier. The results obtained by this, are observed in the table 4.4, the results are shown by model.

Model	Precision	Recall	F1	F2	Accuracy
ResNet-152	0.8928	0.8475	0.8695	0.721	0.8660
ResNeXt-101	0.7142	0.9523	0.8162	0.7515	0.8392
GoogLeNet	0.8214	0.8846	0.8518	0.7275	0.8574

Table 4.4: Error metrics ResNet-152, ResNeXt-101 and GoogLeNet.

4.4 Results Logistic Regression.

In the implementation of the Logistic Regression algorithm, the database containing the measures of standard deviation, variation, entropy, absolute power, relative power was used, in the table 4.5, different values of these characteristics are observed, among these, we have p-value, which mentions that entropy is the measure with the highest statistical significance.

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
(Intercept)	-32.49374	5.73896	-5.662	1.50e-08 ***
entropy	2.84579	0.56312	5.054	4.33e-07 ***
sd	0.18525	0.06763	2.739	0.00616 **
var	0.40138	2.05714	0.195	0.84530
AbsolutePower	-0.40316	2.05737	-0.196	0.84464
RealtivePower	2.91490	3.81132	0.765	0.44439

Figure 4.5: Output values of Logistic Regression algorithm.

In the figure 4.6, the prediction relationship concerning entropy is observed

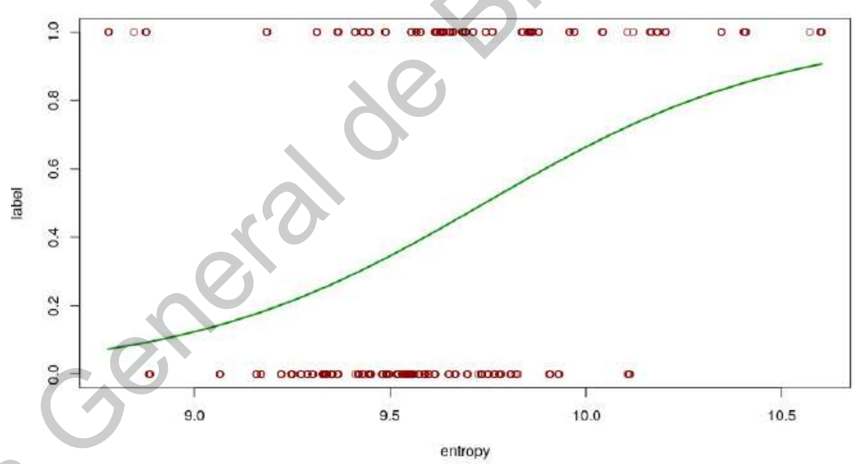


Figure 4.6: Predicted relationship of entropy.

The model implemented in this research was validated with the metrics mentioned above, using a 70-30 method (this is 70% of the randome data is used for training and the remaining 30% is used for validation). The algorithm was run a certain number of times and the best result is the which is presented in table 4.5.

Model	Precision	Recall	F-score	AUC	Accuracy
Logistic Regression	0.7169	0.7755	0.7450	0.74	0.8030

Table 4.5: Error metrics Logistic Regression.

In figure 4.7, the results obtained are observed, we can observe a comparison between false negatives, false positives, true negatives, and true positives, between the real class and the prediction.

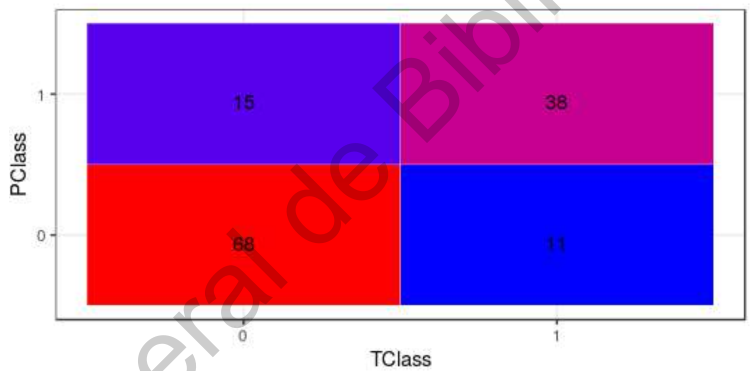


Figure 4.7: Confusion Matrix of Logistic Regression.

In the figure 4.8 we can see the ROC curve resulting from our model, the sensitivity of the model is observed on the y-axis and the specificity on the x-axis.

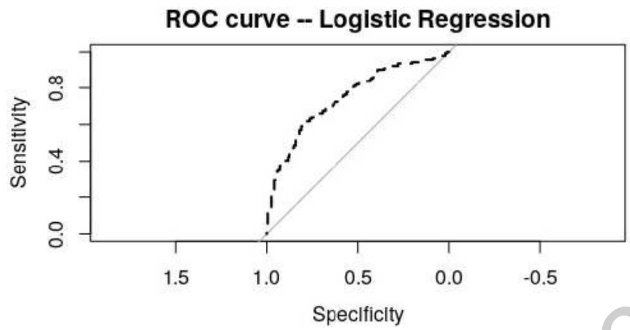


Figure 4.8: ROC curve Logistic Regression.

4.5 Results Maximum Likelihood.

The results obtained with the Maximum Likelihood model are observed in the table 4.9, where it was evaluated with the metrics described in previous sections.

Model	Precision	Recall	F-score	Accuracy
Maximum Likelihood	0.8230	0.9190	0.8680	0.8440

Figure 4.9: Error metrics Maximum Likelihood.

In the figure 4.10, the confusion matrix of the different values obtained by the Maximum Likelihood model is observed.

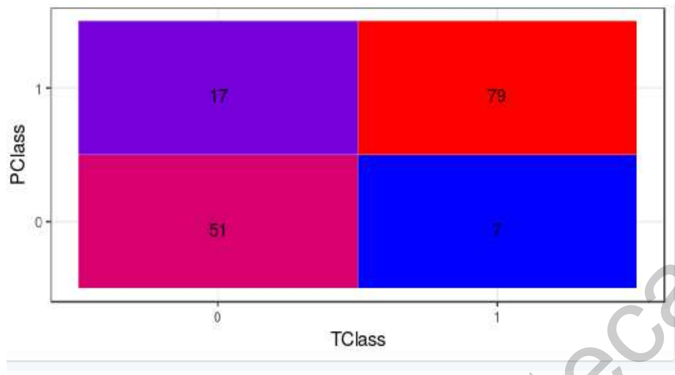


Figure 4.10: Confusion Matrix of Maximum Likelihood.

In the figure 4.11 the comparison of the different metrics used as evaluation methods for the two models implemented in this research work is observed, in the x-axis, two groups, which are: 0 and 1, where 0 belongs to the Maximum Likelihood model and 1 belongs to the Logistic Regression model, a trend of higher values are observed in the Maximum Likelihood model.

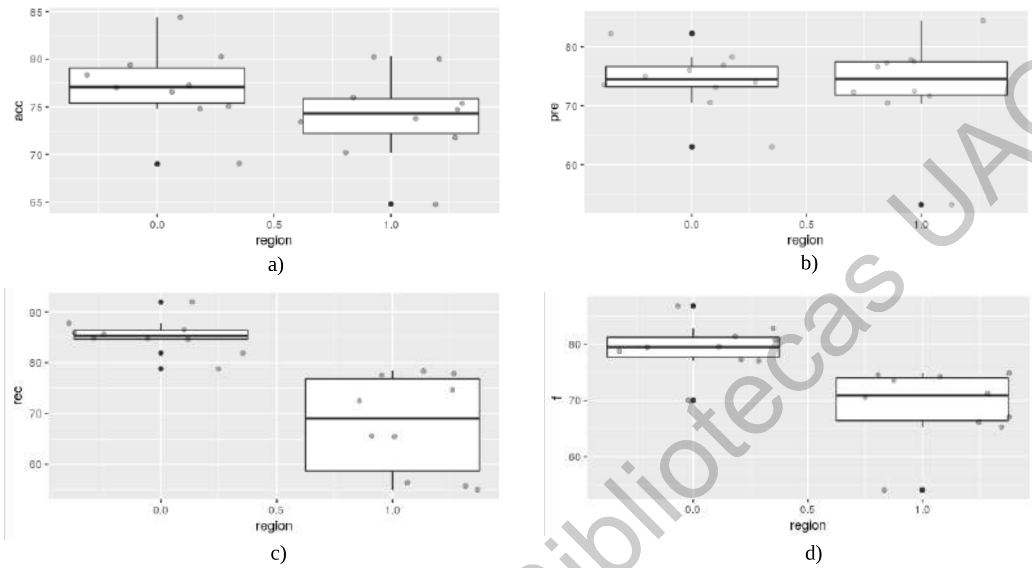


Figure 4.11: a) Accuracy of the Maximum Likelihood and Logistic Regression models. b) Precision of the Maximum Likelihood and Logistic Regression models. c) Recall of the Maximum Likelihood and Logistic Regression models. d) F1 of the Maximum Likelihood and Logistic Regression models.

4.6 Discussion of results.

In this work, three different models of deep learning are proposed, in this case, ResNet-152, ResNeXt-101, and GoogLeNet, each of these models have individual characteristics, the network with greater depth is ResNet-152, as its name indicates. It is composed of 152 layers of deep, followed by ResNeXt-101 with 101 layers of deep and finally GoogLeNet with only 22 layers of deep, the first model obtains excellent results for the deep of its architecture, while the second proposes a lower depth but compensated for parallelism and the third one that has the least amount of layers, but the strength of this model lies in the use of the Inception module.

We observe in the table 4.4 that the results obtained by these three models are very similar, so we can reiterate that the depth of the first one is compensated by parallelism in the second and third by the use of the Inception module.

In the figure 4.4 we observe the performance of the different models, even though the performance of the three is very similar, we can observe a greater speed in the performance of the ResNeXt-101 model, followed by the ResNet-152 model, and finally GoogLeNet.

Taking into account these assertions we can see that the model with the highest accuracy is ResNet-152 followed by the GoogLeNet model and finally the ResNeXt-101 model, giving a comparison between accuracy and computational cost.

The results obtained between the models implemented in this research work have similar values to each other where the greatest difference is found in the implemented methodologies, which are separated into 2 different ones, the first one implemented for the Deep Learning models (ResNet-152, ResNeXt-101, and GoogLeNet) and the second one implemented for the Maximum Likelihood and Logistic Regression models. Both methodologies are the same up to the point of extraction of the β band, but it is observed that for the Deep Learning methodology only the feature of Absolute Power is obtained, then it is necessary to extract the most relative attributes through GA, once these attributes have been obtained, it is necessary to convert them to images using the CWT technique in order to use these images as input for training, validation and testing of the three models. The second methodology makes use of a greater number of features which are; Standard Deviation, Variance, Entropy, Absolute Power and Relative Power, these are the characteristics that are used as training and testing for the

Maximum Likelihood and Logistic Regression models.

The table 4.6 shows the results obtained and the characteristics implemented for each model, we can see that the results obtained have a certain similarity with each other.

Model	Measures	Precision	Recall	F1	Accuracy
ResNet-152	Absolute β band Power	0.8928	0.8475	0.8695	0.8660
ResNeXt-101	Absolute β band Power	0.7142	0.9523	0.8162	0.8392
GoogLeNet	Absolute β band Power	0.8214	0.8846	0.8518	0.8574
Regression Logistic	Absolute Power, Relative Power, Entropy, Standard Deviation and Variance of β band	0.7169	0.7755	0.7450	0.8030
Maximum Likelihood	Absolute Power, Relative Power, Entropy, Standard Deviation and Variance of β band	0.8230	0.9190	0.8680	0.8440

Table 4.6: Error metrics and measures.

The extraction of a greater amount of features means a greater amount of information, this allows the implementation of the Maximum Likelihood model, which in comparison with the three Deep Learning models, can be considered a model of less complexity and faster execution. Not only does the methodology implemented in the Maximum Likelihood model make use of a greater amount of information, but also during the process a lower loss of it is obtained, in the other hand, the methodology implemented in the Deep Learning models a loss of relevant information is observed, both in the section for a selection of attributes by the GA and in the conversion to 2D images by the CWT technique.

In the figure 2.6, the distribution of the different features obtained from the EEG signals of the ADHD and Control groups is observed, it is showed that the groups share a considerable number of values, but there is also a difference between both groups, each unique value has a greater probability of occurrence in one group than in the other, it is the principle that the Maximum Likelihood model is based on. Each value that is used as a test was evaluated by the model Maximum Likelihood in the five different features. In each feature makes a decision depending on the probability of occurrence after the mode of the five decisions is obtained and this is the final result of the evaluation.

4.7 Conclusion.

Taking into account the results and their discussion, we can conclude that the technique of using inception models is one of the most effective ways to address this problem, that is, according to the tests carried out, we observe that it is not the model with greater accuracy, nor is it the least accurate, also in learning performance it shows an intermediate place, and it is the one that takes the least time in execution, analyzing the last and most important point which is the enormous difference that exists between the number of deep layers and parallel layers, it shows better performance.

As observed throughout this research work, a greater number of features are converted into a greater amount of information, which are: the Standard Deviation, the Variance, the Entropy, the Absolute Power and the Relative Power which provide sufficient information to be able to make a classification between the groups of ADHD and Control. In this research work complex models such as Deep Learning and not as complex as Maximum Likelihood are observed, but if the problem is analyzed and the results obtained, there is no high relationship between complexity and results, that is, it is necessary to know the problem and search for the best way to approach it, this research work concludes that, between the two methodologies proposed for the classification of people with ADHD and Control, the methodology implemented in the Maximum Likelihood model is more efficient, due to the fact that the features provide sufficient information to achieve classification with high accuracy.

4.8 Future work

Although some authors mention that the use of Principal Components Analysis (PCA) and Independent Component Analysis (ICA) are meaningful to solve this problem, it may be noteworthy to explore these techniques looking for new solutions.

Bibliography

[1] R. Yaghoobi Karimui, S. Azadi, and P. Keshavarzi, "The adhd effect on the high-dimensional phase space trajectories of eeg signals," *Chaos Solitons Fractals*, vol. 121, pp. 39–49, 02 2019.

[2] H. Helgadóttir, O. Gudmundsson, G. Baldursson, P. Magnusson, and Blin, "Electroencephalography as a clinical tool for diagnosing and monitoring attention deficit hyperactivity disorder: A cross-sectional study," *BMJ Open*, vol. 5, p. e005500, 01 2015.

[3] A. Tenev, S. Markovska-Simoska, L. Kocarev, J. Pop-Jordanov, A. Mueller, and G. Candrian, "Machine learning approach for classification of adhd adults.," *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, vol. 93, 01 2013.

[4] R. Yaghoobi Karimui and S. Azadi, "Diagnosing the adhd using a mixture of expert fuzzy models," *International Journal of Fuzzy Systems*, vol. 20, 01 2017.

[5] G. Ogrim, J. Kropotov, and K. Hestad, "The quantitative eeg theta/beta ratio in attention deficit/hyperactivity disorder and normal controls: Sensitivity, specificity, and behavioral correlates," *Psychiatry Research*, vol. 198, no. 3, pp. 482 – 488, 2012.

[6] S. M. Snyder, H. Quintana, S. B. Sexson, P. Knott, A. Haque, and D. A. Reynolds, "Blinded, multi-center validation of eeg and rating scales in identifying adhd within a clinical sample," *Psychiatry Research*, vol. 159, no. 3, pp. 346 – 358, 2008.

[7] V. Monastra, J. Lubar, and M. Linden, "The development of a quantitative electroencephalographic scanning process for attention deficit-hyperactivity disorder: Reliability and validity studies," *Neuropsychology*, vol. 15, pp. 136–44, 02 2001.

[8] K. Sadatnezhad, R. Boostani, and A. Ghanizadeh, "Classification of bmd and adhd patients using their eeg signals," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1956 – 1963, 2011.

[9] M. Ahmadlou, R. Rostami, and V. Sadeghi, "Which attention-deficit/hyperactivity disorder children will be improved through neurofeedback therapy? a graph theoretical approach to neocortex neuronal network of adhd," *Neuroscience Letters*, vol. 516, no. 1, pp. 156 – 160, 2012.

[10] C. A. Magee, A. R. Clarke, R. J. Barry, R. McCarthy, and M. Selikowitz, "Examining the diagnostic utility of eeg power measures in children with attention deficit/hyperactivity disorder," *Clinical Neurophysiology*, vol. 116, no. 5, pp. 1033 – 1040, 2005.

[11] M. R. Mohammadi, "Eeg classification of adhd and normal children using nonlinear features and neural network," *Biomedical Engineering Letters*, 2016.

[12] P. Bizopoulos, G. I. Lambrou, and D. Koutsouris, "Signal2image modules in deep neural networks for eeg classification," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 702–705, 2019.

[13] M. Hasan, D. Shon, H. Im, K.and Choi, D. Yoo, and J. Kim, "Sleep state classification using power spectral density and residual neural network with multichannel eeg signals," *Applied Sciences.*, vol. 10, no. 21, 2020.

[14] A. Shalbaf, S. Bagherzadeh, and A. Maghsoudi, "Transfer learning with deep convolutional neural network for automated detection of schizophrenia from eeg signals," *Physical and Engineering Sciences in Medicine*, 2020.

[15] Y. Gao, B. Gao, Q. Chen, J. Liu, and Y. Zhang, "Deep convolutional neural network-based epileptic electroencephalogram (eeg) signal classification," *Frontiers in Neurology*, vol. 11, p. 375, 2020.

[16] F. Duan, Z. Huang, Z. Sun, Y. Zhang, Q. Zhao, A. Cichocki, Z. Yang, and J. Solé-Casals, "Topological network analysis of early alzheimer's disease based on resting-state eeg," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 10, pp. 2164–2172, 2020.

[17] N. Hatami, Y. Gavet, and J. Debayle, "Classification of time-series images using deep convolutional neural networks," 2017.

[18] s. Wang, Y. Wang, and S. Wang, "Time-resnext for epilepsy recognition based on eeg signals," *bioRxiv*, 2019.

[19] Q. Chen, H. Yue, X. Pang, W. Lei, G. Zhao, E. Liao, and Y. Wang, *Mr-ResNeXt:*

A Multi-resolution Network Architecture for Detection of Obstructive Sleep Apnea, pp. 420–432. 08 2020.

[20] A. M. N. A. A. M. S. M. R. Mohammadi, "Eeg data for adhd / control children," 2020.

[21] K. Sadatnezhad, R. Boostani, and A. Ghanizadeh, "Classification of bmd and adhd patients using their eeg signals," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1956 – 1963, 2011.

[22] Z. Z. Li Hu, *EEG Signal Processing and Feature Extraction.* Springer Nature Singapore Pte Ltd., 2019.

[23] M. H. Hayes, *Statistical digital signal processing and modeling.* John Wiley & Sons, Inc, 1996.

[24] Z. Bian, Q. Li, L. Wang, C. Lu, S. Yin, and X. Li, "Relative power and coherence of eeg series are related to amnestic mild cognitive impairment in diabetes," *Frontiers in aging neuroscience*, vol. 6, p. 11, 02 2014.

[25] R. Q. dao-er ji and Y. Wang, "A new hybrid genetic algorithm for job shop scheduling problem," *Computers and Operations Research*, vol. 39, no. 10, pp. 2291–2299, 2012.

[26] A. Khan, I. A. Chaudhry, and A. Khan, "A genetic algorithm for flexible job shop scheduling," vol. 1, 07 2013.

[27] P. Ghorbanian, D. Devilbiss, A. Simon, A. Bernstein, T. Hess, and H. Ashrafiuon, "Continuous wavelet transform eeg features of alzheimer's disease," 10 2012.

[28] E. Stevens, L. Antiga, and T. Viehmann, *Deep Learning with PyTorch*. Manning Publications Co, 2020.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[30] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," 2017.

[31] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," 2016.

[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014.

[33] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015.

[34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[35] K. P. Murphy, *Machine Learning a Probabilistic Perspective*. The MIT Press, 2012.

[36] J. Gareth, W. Daniela, H. Trevor, and T. Robert, *An introduction to Statistical Learning*. Springer New York Heidelberg Dordrecht London, 2014.

[37] M. Bonamente, *Statistics and Analysis of Scientific Data*. Springer Science Business Media New York 2013, 2013.