REM SLEEP STAGE IDENTIFICATION WITH WAVELET DECOMPOSITION AND ARTIFICIAL NEURAL NETWORK USING A SINGLE CHANNEL EEG

By

Gabriel Toban

A qualifying paper submitted in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Computational Science

Middle Tennessee State University

May 2020

Project Supervisor:

Dr. Hong

ACKNOWLEDGEMENTS

ABSTRACT

The abstract can be no more than 150 words. It goes on this page.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I.

INTRODUCTION

Sleep stages are the most precise way to separate wakefulness from sleep state. Physiologic and pathologic events can be identified and studied through examining stages of sleep or sleep stages. Certain health insurances require that wakefulness is separated from sleep state using an electroencephalogram (EEG) in order for a sleep study to be conducted.

An EEG uses electrodes placed on the head to detect electric potentials in the brain. The difference between the electric potentials of two electrode positions, called an EEG channel or channel, are recorded in regular intervals to produce brain waves. One electrode is the active electrode while the other is the reference electrode [11]. The electrode positions are well-defined by the International 10/20 system [11] [2] [1]. The EEG channels that are best to use were standardized originally by Rechtscheffen and Kales [26] commonly known as the R&K standard. The American Academy of Sleep Medicine (AASM) maintains and updates the R&K standard today.

Sleep stages are typically identified with a combination of parameters known as a polysomnographic record (PSG). These parameters include, but are not limited to, electroencephalogram (EEG) derivations, electrooculogram (EOG) derivations, and a chin electromyogram (EMG) [11]. Elements from all 3 are used to identify sleep stages [11]. An EEG monitors brainwaves. An EOG monitors eye movements. An EMG monitors muscle movements. EEGs, EOGs, and EMGs all can use electrodes placed on the head and face.

An alternative to a PSG is an at home study. There can be many electrodes and channels used in a PSG. The AASM recommends that 8 electrode positions creating 6 channels are used for an EEG alone [11]. Using this many channels and the variety of parameters for the PSG requires special training to use, special training to interpret, expensive equipment, and likely a lab. Some at home studies use a single channel EEG requiring very little training to use, cheaper equipment, and no lab. The at home study still requires special training to

interpret.

Automated sleep staging can be accomplished with a combination of parameters from a PSG [28] [17] [25] [20] [27] or a single channel EEG [29] [16] [13] [30]. Ebrahimi et al. achieved 93% accuracy using the wavelet decomposition of a single channel EEG and a neural network [16]. Kurt et al. achieved 97-98% accuracy using the wavelet decomposition of EEG, EOG, and chin-EMG [20].

In order to automate sleep staging with an at home study, wavelet decomposition of a single channel EEG was used identify REM sleep using a feed-forward back-propagation (FFBB) artificial neural network (ANN). Chapter 2 explores the factors of picking a single EEG channel, the reasons to use wavelet decomposition for signal processing, and the reasons behind the feed-forward back-propagation ANN. Chapter 3 explains the exact methods and results. Chapter 4 discusses the conclusions and future studies.

CHAPTER II.

BACKGROUND

The Single Channel EEG

An EEG uses electrodes placed on the head to detect electric potentials in the brain. The difference between the electric potentials of two electrode positions, called an EEG channel or channel, are recorded in regular intervals to produce brain waves. The electrode positions are well-defined by the International 10/20 system [11] [2] [1]. The EEG channels that are best to use were standardized originally by Rechtscheffen and Kales [26] commonly known as the R&K standard. The American Academy of Sleep Medicine (AASM) maintains and updates the R&K standard today.

Brain waves, activities, and events are defined by the amplitude, frequency, and identifying factors of a single wave or grouping of waves. There are 5 EEG frequencies defined in table 1. The identifying factors include things such as what happens before/after the wave, the difference in peeks of wave, and how much a wave stands out. [11]. Waves, activities, or events, such as spindles or arousals, can be identified by a combination of these identifying factors, the frequency, and the amplitude [11].

Table 1: The frequency and peak-to-peak amplitude of different brain waves as identified by the AASM [11].

Name	Frequency (Hz)	Peak-to-Peak Amplitude (uV)
Slow Wave Activity	0.5-2.0	>75
Delta Waves	0-3.99	N/A
Theta Waves	4-7.99	N/A
Alpha Waves	8-13	N/A
Beta Waves	>13	N/A

The international 10/20 system uses labels with letters and numbers to identify the different electrode positions on the head referenced in figure 1. The letters identify the brain lobe that the electrode is over identified in table 2. The numbers identify the position in the direction going from one ear to the other. To find each position, a technician begins at a starting position and moves around the head in percentages. [1] [2]
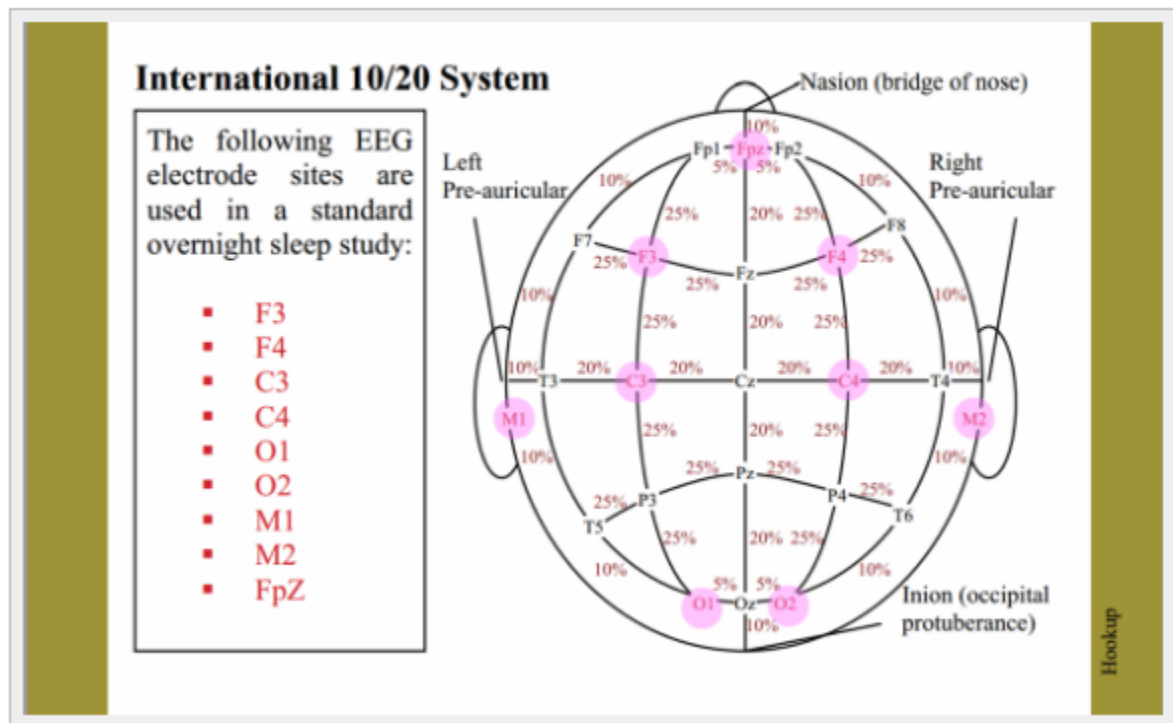
Figure 1: A representation of the electrode positions identified by the 10/20 system. It originated from https://sleeptechstudy.wordpress.com/category/1020-system/ [1]. Each position is labeled by 1 to 3 letters and/or a number which are listed on the left side of this image.

Table 2: The brain region that relates to every electrode location. It originated from https://sleeptechstudy.wordpress.com/category/1020-system/ [1].

Abbreviation	Brain Lobe/Region
Fp	Frontal pole
F	Frontal
T	Temporal
C	Central
P	Parietal
O	Occipital
M	Mastoid

EEG channels are defined by two electrode positions. For an adult EEG, the AASM recommends the channels listed in table 3. The data used for this study does not use the channels recommended by the AASM [5], but the efficacy of the alternative electrode positions are described by van Sweden et al. [9].

Table 3: The channels for an EEG to identify sleep stages as recommended by the AASM [11]. Each electrode has backup electrode. Some important backups are Fpz for Fz and Cz for C3.

Recommended	Acceptable
F4-M1	Fz-Cz
C3-M1	Cz-Oz
O2-M1	C4-M1

Different electrode positions and channels detect different brain waves, activities, and events. Table 4 list some of the brain waves, activities, and events and which lobe position or channel detects them. This is one the reasons to use multiple channels for an EEG.

Table 4: The brain regions/electrode channel that brain waves, activities, and events (marker) are most likely to appear. To understand what stages are identified with each marker, reference the AASM scoring guide [11]. Blanks space mean there was not any specifications for this region/channel and marker combination. The Yes means that this channel does identify Slow Waves while the No means that the channel does not identify Slow Waves. Maximal implies the maximal wave amplitude will be recognized. Adequate means adequate for identification. All the information except Adequate was gathered from the AASM scoring guide [11]. The Adequate comes from the R&K standard [26]. E1-Fpz is the frontal regions referenced to the contralateral ear or mastoid.

	Slow Wave	V-Waves	K-complexes	Spindles	Sawtooth	Alpha
Frontal			Maximal			
Central		Maximal		Maximal	Maximal	Adequate
Occipital						Maximal
Fz-Cz	No					
E1-Fpz	Yes					

The definition and number of sleep stages has evolved over time. The R&K standard originally defined 7 stages to identify sleep excluding non rapid eye movement (NREM) which is a combination of stage 1, 2, 3, and 4 [26]. The AASM defines 5 stages to identify sleep [11]. The AASM definition excludes Movement Time and combines stage 3 and 4. Table 5 list all the stages from R&K. The data used for this study follows the R&K standard stages with an extra identifier for unknown or ungraded stages [5].

Table 5: The different sleep stages as identified by the R&K [26] and the AASM [11]. It is important to note that Movement Time (MT) and Non-REM (NREM) are still used. Also, NREM 3 from the AASM is a combination of Stage 3 and Stage 4 from R&K.

R&K	AASM
Wakefulness (W)	Wakefulness (W)
Movement Time (MT)	
Stage 1	NREM 1 (N1)
Stage 2	NREM 2 (N2)
Stage 3	NREM 3 (N3)
Stage 4	
NREM (1,2,3,4)	
REM (R)	REM (R)

Different brain waves, activities, and events (EEG markers) help define different sleep stages. Every sleep stage is completely defined according to the AASM using many factors of a PSG (polysomnogram) including the EEG, EOG, and EMG [11]. For example, REM stage sleep uses an EOG to identify the rapid eye movement the stage is named after, but it may be identified with sawtooth waves detected by an EEG. Automated sleep staging with an EEG does not consider all the additional defining parameters from the PSG. Table 6 gives an idea of the relationship between EEG markers and sleep stages.

Table 6: Which sleep stages are identified by certain EEG markers as recommended by the AASM [11]. It does not identify every possible marker. For example, stage N1 can follow stage W, stage N2, and stage R under different conditions. An arousal is the cause of many significant events including k-complexes.

NR: Not Required; R: Required; P: Possible; {Blank}: Not mentioned or not possible

	W	N1	N2	N3	REM
Slow Wave				R	
V-Waves		NR			
K-complexes		P	R		P
Spindles		P	R		P
Sawtooth					NR
Alpha	NR	P			P
Beta	NR				
LAMF {theta}		R	P		R
Theta		P			
No Arousal			R		P

There is a current push to diagnose sleep associated diseases with a cheaper at home study. PSG requires expensive equipment and a technician present. The AASM [11] recommends that at least 3 EEG channels, 2 EOG channels, and a chin EMG among other things for a PSG. Ventouras et al. [29], Ebrahimi et al. [16], and Crasto et al. [13] all used a

single EEG channel with accuracies between 85% and 95%. Shambroom et al. [27] tested current equipment that uses a single channel EEG to do sleep staging. Sleep staging using a single EEG channel would be cheaper than a PSG. This study will use a single EEG channel.

Aside from slow waves, any single brain wave, activity, or event (EEG marker) from an EEG is not uniquely required in only 1 sleep stage. Slow waves are identified by E1-Fpz channel, table 4, and identify stage N3, table 6. Although, Crasto et al. [13] was able to break down stage N3 into stage 3 and 4 using channel Fpz-Cz. The dataset used in this study has channels Fpz-Cz and Pz-Oz available. The channel to identify slow waves is not available. Since all other waves in table 4 are adequately identified in the central and frontal lobes, Fpz-Cz channel was used to for this study. Future studies may consider using other frontal lobe channels.

The AASM and R&K recommend that sleep staging is completed in 30 second epochs [11] [26]. Each epoch is assigned a single sleep stage [26] [11]. If an epoch exhibits features of more than one stage, the R&K standard [26] strictly assigns each epoch the stage that takes the most of the epoch . The AASM [11] has additional rules when the features of more than one stage are exhibited in any epoch. The R&K standard [26] states 3 reason for 30 second epochs: too small epoch will be too cumbersome, too large epochs will miss out on epoch details, and paper sizes for EEG machines typically fit 30 seconds of data. The Physionet data set [5] used in this study has 30 second epochs marked with their sleep stage. Current study uses 30 second epochs. The AASM and R&K both mention the necessity to consider the stage before and after the epoch currently being scored. This study will use 3 epochs to stage a single epoch. In future studies, it may be best to identify each individual wave. Once each wave is identified, epochs will not be required. Each stage could be uniquely identified rather than being a potential group of stages identified by 1 stage.

REM sleep has not been the specific goal of any study found. REM sleep is the only stage that can be easily found by rapid eye movements. Rapid eye movements are not

detected with an EEG. There is not a required unique identifier for REM sleep in the EEG. REM sleep is almost indistinguishable from N1 sleep in an EEG. Stages W, N1, N2, and N3 are easily distinguishable from each other with an EEG (see table 6). Stages N1, N2, and N3 all have a unique EEG marker to identify them. Current study will concentrate on REM sleep.

Signal Processing and Wavelet Decomposition

A brainwave or a single channel from an EEG is a signal. It is a record of the difference in electrode potentials over time. Byrne [12] states that the Fourier series is the most important tool in signal processing. Achermann [8] stated that Dietsch [15] performed the first Fourier analysis on an EEG in 1932.

The Fourier series of a function is [14]

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{-\frac{in\pi x}{l}}$$

where

$$c_n = \frac{1}{2l} \int_{-l}^{l} f(t) e^{-\frac{in\pi x}{l}}.$$

The Fourier integral theorem takes the limit as l approaches infinite of the Fourier series. [14]

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iwx} dw \int_{-\infty}^{\infty} e^{-iwt} f(t) dt$$

The Fourier transform of a signal is derived from the Fourier integral [14].

$$Ff(t) = \hat{f}(w) = \int_{-\infty}^{\infty} e^{-iwt} f(t) dt = \langle f, e^{iwt} \rangle$$

All of these represent harmonic or sinusoidal waves. The transform and series go on forever. The frequency of the waves of a Fourier Transform represent the Spectral Density of

the original wave [6]. The spectral density is a representation of the amplitude of a wave[7], but the sinusoidal component of the transform does not allow if to capture local information [14]. In other words, abrupt changes in amplitude may not be captured. It is well-known that Fourier Transforms cannot investigate time and frequency simultaneously [28] [16] [20] [13] [14].

Some of the different brain waves, activities, and events in the EEG signal are defined by abrupt changes in amplitude over time. V-Waves are defined as "sharply contoured waves with a duration less than 0.5 seconds" [11] and literally look like a "V". K-complexes are defined by "a well-delineated, negative, sharp wave immediately followed by a positive component standing out from the background EEG, with total duration greater or equal to 0.5 seconds" [11]. Figure 2 shows the V-Wave and K-complex.
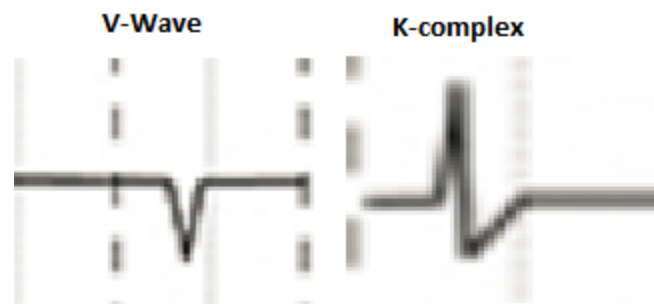
Figure 2: The V wave and K-complex used in the sleep stage scoring manual from the AASM [11].

In order to provide descriptions of abrupt changes in amplitude over time like those found in the EEG signal, time-frequency signal analysis representations were created. Time-frequency representations can be classified as linear, quadratic, or higher-order representations. Hlawatsch and Boudreaux-Bartels listed at least 25 different time-frequency representations including 3 linear representations, Gabor, Short-time Fourier transform (STFT), and Wavelet transform (WT), and 2 quadratic representations, Choi-Williams Distribution (CWD) and Wigner Distribution (WD) [19]. Several of the other representations were modifications of WD [19]. The Wigner Distribution (WD), also known as Wigner-Ville

Distribution (WVD), and the Choi-Williams Distribution are both generated from the Cohen
"general class of bilinear shift-invariant, quadratic time-frequency distributions" [14]. The
WVD has been modified many times to create many time-frequency distributions [14] [6].

A general review of the methods used shows that Fourier and Wavelets are the primary
representations used for analyzing an EEG in sleep staging [30]. Among those that used
Fourier and Wavelets, the Wavelet representations resulted in the highest accuracies [28]
[16] [20] [13] [30]. Fraiwan et al. [17] used Choi-Williams Distribution (CWD), continuous
wavelet transform (CWT), and Hilbert-Huang Transform (HHT). Fraiwan et al. [17] found
that CWT had the highest accuracy. A Discrete Wavelet Transform (DWT) can decrease the
processing time of a CWT [14]. Ebrahimi et al. [16] and Crasto and Upadhyay [13] used
Daubechies wavelet order 2 filter with DWT. Crasto and Upadhyay [13] tested DWTs with
Daubechies wavelet order 2 to 6 to decompose a signal for input into an artificial neural
network (ANN) finding Daubechies wavelet order 2 to produce the most accurate results.

There are many time-frequency signal analysis representations that could be tested. The
Gabor Transform, Short-time Fourier Transform, the Wigner Distributions, and many more
could be tested with future studies. Since the Discrete Wavelet Transform has been tested
with the highest accuracy, this study used DWT to decompose the signal.

The discrete wavelet transform is based on the concept of a function called the "mother
wavelet" [14].

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Psi(\frac{t-b}{a}), \ a,b \in \mathbb{R}, a \neq 0$$

This mother wavelet can be translated with b and dilated with a. It is part of the wavelet
transform of a function f [14].

$$W_{\Psi}[f](a,b) = \langle f, \Psi_{a,b} \rangle = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \overline{\Psi(\frac{t-b}{a})} dt$$

The function $\Psi_{a,b}(t)$ replaces e^{iwt} in the Fourier integral [14]. $\Psi_{a,b}(t)$ can be one of many wavelets. In the Fourier series, the sinusoidal waves produced from e^{iwt} are the problem. Wavelet bases produced were orthonormal or orthogonal still lacked in the ability to decompose non-stationary waves [14]. Daubechies was the first nonorthogonal wavelet produced that decomposes a non-stationary wave [14]. Since Daubechies wavelet order 2 produced high accuracy results in multiple papers [13] [16], this study chose to use Daubechies wavelet order 2. Figure 3 shows the Morlet and Daubechies 2 wavelet.
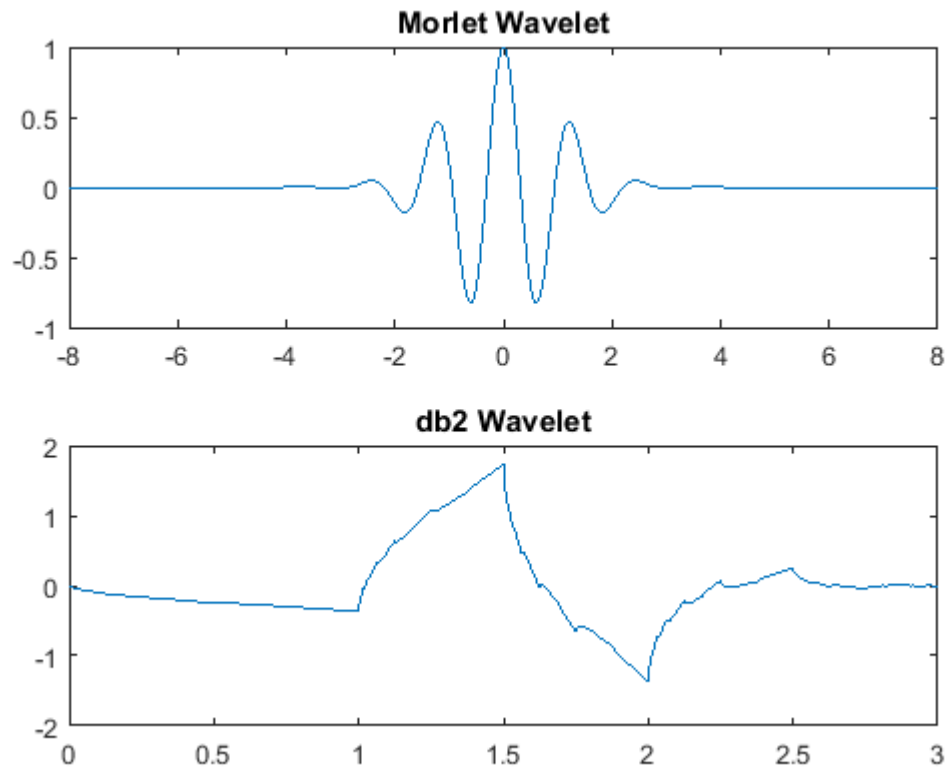
Figure 3: A plot of the Morlet wavlet and the Daubechies order 2 wavelet produced by Matlab.

The discrete wavelet transform has many modifications to help it improve. Ebrahimi et al. [16] used a wavelet packet tree (WPT), but there are many more such as the stationary wavelet transform (SWT), complex-valued wavelet transform (CWT), dual-tree complex wavelet transform (DTCWT), etc. Testing the differences in these modifications can be

completed in a future study.

This study used the multilevel discrete wavelet transform. Each level is the transform of the previous level. Each level is not as close to the originial wave as the previous level, but it greatly reduces the number of coefficients (a and b described above).

Categorization and Artificial Neural Networks

The purpose of this article is to classify REM sleep using the output of multilevel discrete wavelet transform. This paper was completed to use an artificial neural network (ANN), but there are some alternative classifcation algorithms. Alvarez et al. [30] listed the following algorithms, among others, in order of highest accuracy for sleep staging: ANN, support vector machine (SVM), hidden Markov models, and discriminant analysis. Lotte et al. [24] scored many algorithms for classification in EEG-based brain-computer interface including linear discriminant analysis, support vector machine (SVM), ANN multilayer perceptron, others ANNs, Bayes quadratic, hidden Markov model, k nearest neighbors, Mahalanobis distance, and combinations. Even though Lotte et al. [24] concluded that the SVM would be best for EEG-based brain-computer interface, they also stated neural networks are the most used category of classifier. Seeing that both Alvarez et al. and Lotte et al. put significance on SVMs and hidden Markov models, these two algorithms may be tested in the future.

There are many types of ANNs. Heaton [18] listed feedforward networks, deep belief networks, deep feedforward networks, and convolution networks as the best ANNs for classification. Other studies have had high accuracy with feed forward back propagation neural networks [13] [28] and multilayer perceptron neural networks [20] [29]. All of these algorithms could be tested. This paper used a feed forward backward propagation neural network.

A neural network produces a non-linear decision boundary [24]. It is easy to understand a neural network when it is described from regression which produces a linear decision boundary.

Linear regression is based on the equation of a line. The equation of a line is $y = mx + b$. Given any set of x and y coordinates, there exist a line that minimizes the distance between the line and the coordinates [22]. In other words, there exist a red line that minimizes the sum of the lengths of the blue dashed lines in 4.
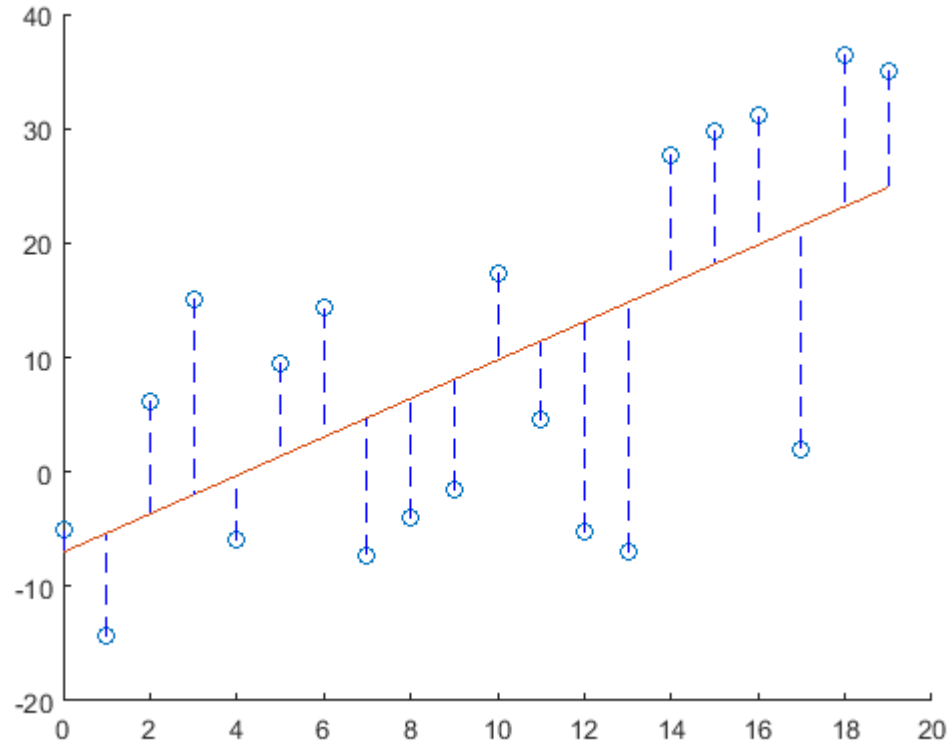
Figure 4: The plot of line that was produced via linear regression on a set of coordinates.

If the coordinates of are represented by set $x = [x_i] and y = [y_i]$ for i in range 1 to the number of points n and $yhat = m * x + b$, where $yhat = [yhat_i]$ for i in range 1 to n, is the equation of a line with m and b unknown, then the sum of the distances between the line and the set is $SUM_{i=1}^{n} |y_i - yhat_i|$ where $|y - yhat|$ is the absolute value of $y_i - yhat_i$. In linear regression, the the square of $y_i - yhat_i$ is used instead of absolute value. This makes the equation for the sum of the distances from the line to the set of coordinates the least squares equation [22]

$$SUM_{i=1}^{n}(y_i - yhat_i)^2.$$

If yhat is replaced with the equation of a line, the least squares equation becomes the cost function for linear regression [22]

$$cost = C = SUM_{i=1}^{n}(y_i - (m*x_i + b))^2.$$

In order to minimize this equation, the derivative with respect to m and b is taken and set equal to 0. This results in a system of two equations and two unknowns that can be solved. When x is multidimensional, the solution is still the same in the sense that the least squares equation is minimized. [22]

When there is not a solution to the multidimensional derivative of least squares, gradient descent can be used to find m and b. Gradient descent is stepping along the gradient of a function. Instead of solving for where the gradient equals zero by setting it to zero, gradient descent takes small steps towards gradient value zero until it reaches it. Mathematically, it looks like

$$w_{i+1} = w_i + \lambda * \nabla f(w_i)$$

where f is the least squares equation and lamda is the step size or the learning rate. The following image can help you visualize gradient descent. [22]
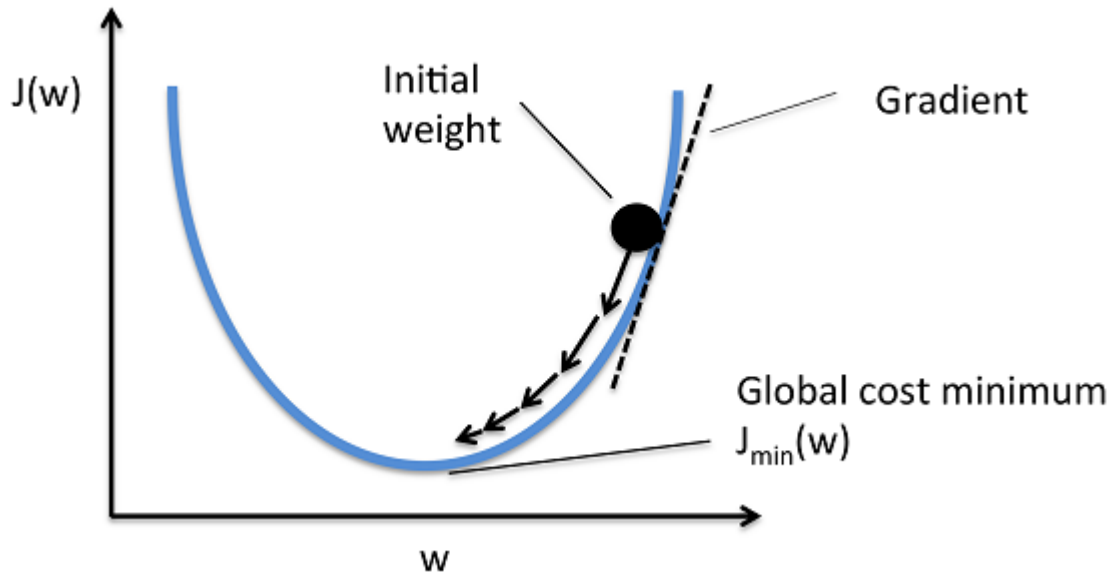
Figure 5: Gradient Descent image from https://sebastianraschka.com/faq/docs/closed-form-vs-gd.html [4].

Let x and y be nodes. Let those nodes be represented by circles. The picture below is a picture of 1 sample of a function in 2 dimensions. Take note that this is a picture of one x_i and one y_i value or one sample.
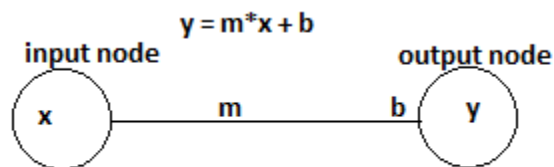
Figure 6: Function in node picture.

When this image is scaled up to multiple dimensions, it looks like.

$$y = m1*x1 + m2*x2 + b$$
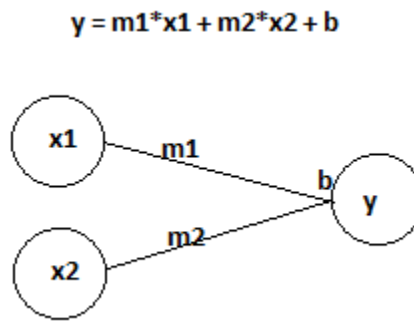
Figure 7: Function with multiple inputs in node picture.

In order to turn this picture from a linear regression picture to a logistic regression picture, the sigmoid function must be used. The sigmoid function returns a value between 0 and 1 always. [23]

$$s(x) = \frac{1}{1 + e^{-x}}$$
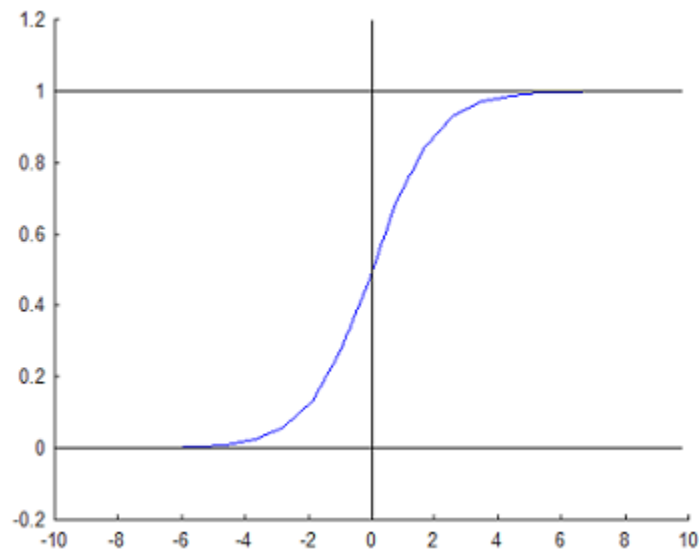
Figure 8: Plot of the sigmoid function.

Once the sigmoid is applied, the nodes picture looks like the following. [23]

$$y = s(\ m1x1 + m2x2 + b\ )$$
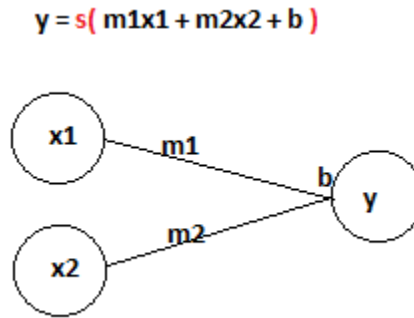
Figure 9: Function with multiple inputs in sigmoid node picture.

Now that this picture is logistic regression, the output is always between 0 and 1. Instead of using least squares as the cost function, the cross-entropy is used as the cost function. [23]

$$C = -\sum_{n=1}^{} N y_n log(yhat_n) + (1 - y_n)log(1 - yhat_n)$$

y is the real solution, yhat is the solution from logistic regression, and N is the number of samples. Just like with multiple linear regression, the derivative of the cross-entropy function C with respect to m and b is taken. When this derivative is set to 0, it can be solved in certain situations. In the situations where it can not be solved, the gradient descent method can be used to find the solution. [23]

The next step to reaching a neural network is to assume that there are multiple outputs per sample. The nodes picture becomes the following:

$$y1 = s(m11x1 + m12x2 + b1)$$
$$y2 = s(m21x1 + m22x2 + b2)$$
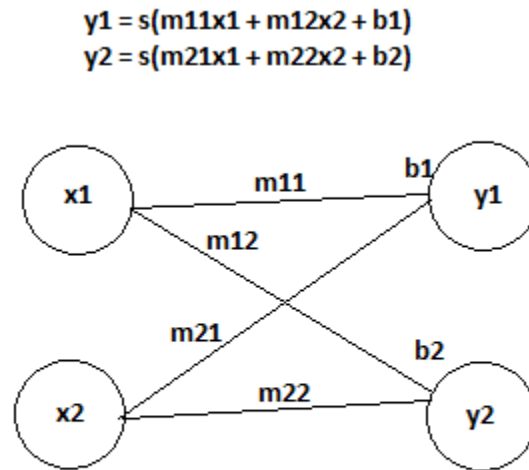
Figure 10: Function with multiple inputs and multiple outputs in a sigmoid node picture.

The final step in reaching a neural network is adding one more layer.

$$z1 = s(m11x1 + m12x2 + b1)$$
$$z2 = s(m21x1 + m22x2 + b2)$$

$$y=s(z1w1 + z2w2 + g)$$
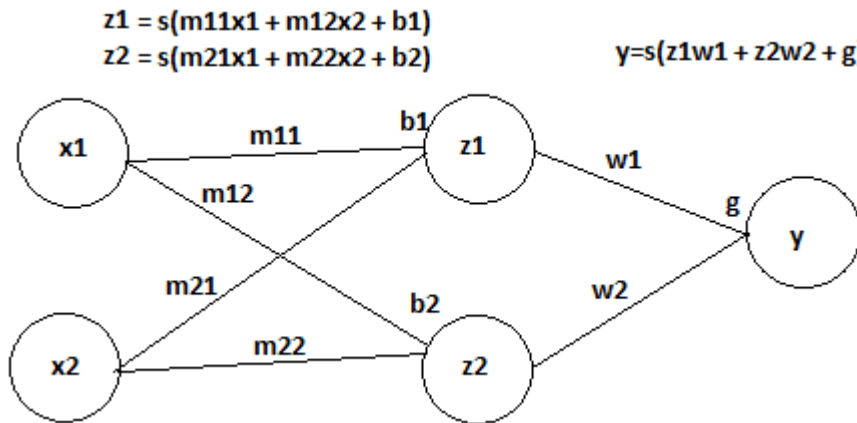
Figure 11: Function with multiple inputs and layers and a single in a sigmoid node picture.

Notice that the function for y, z1, and z2 is still sigmoid. For the middle layers, the sigmoid function can be replaced with 1 of many nonlinearity functions including hyperbolic tangent, linear, step, and rectified linear unites (ReLU) [18]. For the sigmoid function for y, there is a special case. While the function for y is sigmoid, the cost function can remain the cross-entropy function. When there are mutliple output nodes, there is a special function

called the softmax function which should be used. [21]

$$softmax = sf = \frac{a_k}{\sum_{j=1}^{K} a_j}$$

where K is the number of nodes in the layer before y and

$$a_k = \sum_{i=1}^{K} w_{ik} * z_i$$

where z_i is the ith node of the previous layer. [21] The nodes picture become the following:

z1 = s(m11x1 + m12x2 + b1)   y1 = sf(a1 / SUM(aj))
z2 = s(m21x1 + m22x2 + b2)   y2 = sf(a2 / SUM(aj))
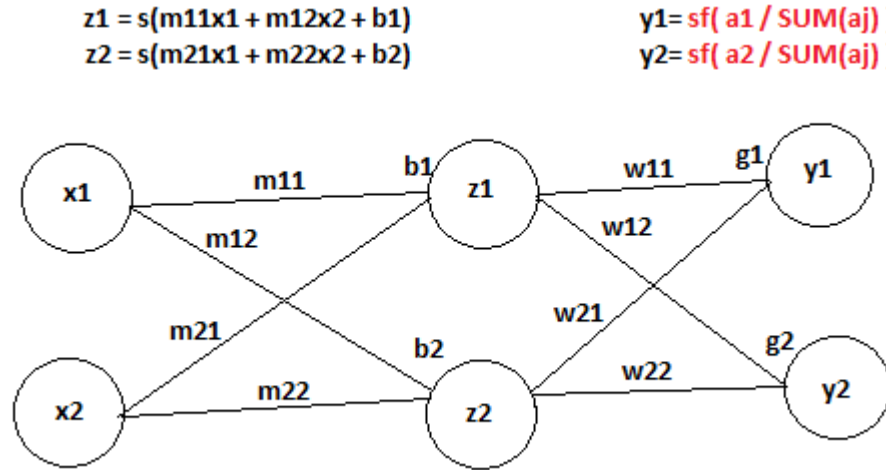
Figure 12: Function with multiple inputs, layers, and outputs in a ANN node picture.

When the softmax function is used, the cost function becomes the multiclass log likelihood (ll) function

$$C = ll = \sum_{n=1}^{N} \sum_{k=1}^{K} y_k^n log(yhat_k^n)$$

where N is the number of samples and K is the number of outputs. Once again, the derivative of this function is taken. The variables w, g, m, and b, or the weights and biases, are solved for using gradient descent. This is the feed forward backward propagation artificial neural network used for this study. [21]

CHAPTER III.

METHODS AND RESULTS

Methods

This study used the "Sleep Recording and Hypnograms in European Data Format (EDF)" data set or "The Sleep-EDF Database [Expanded]" from physionet.org [5]. The portion of the database used is from a study on healthy patients from 1987-1991. There are 20 patients avaialable. There are 10 males and 10 females ranging from age 25-34 years old [5].

Each patient had two relevant files: a polysomnogram (PSG) recording in EDF format and a hypnogram of annotations in EDF+ [5]. EDF is a standard format for exchanging EEG recordings [3]. EDF+ has all the capabilities of EDF and the ability to contain annotations [3]. The PSG recording included an EEG from Fpz-Cz and Pz-Oz electrode locations, an EOG (horizontal), a submental chine EMG, an event marker, an oral-nasal respiration, and rectal body temperature. The hypnogram is an annotation of sleep patterns. The annotations included W for wakefulness, R for REM, 1 for stage 1, 2 for stage 2, 3 for stage 3, 4 for stage 4, M for movement time and ? for not scored. Once the first annotation was reached, the rest of the annotations were 30 seconds apart creating the 30 second epochs.

Special software is required to read the EDF and EDF+ files. EDFbrowser is a free EDF and EDF+ reader that can export the contents to a text file. All the contents of the PSG and hypnogram file were exported to a text file with the option to have time in seconds from the first recording.

The PSG and hypnogram text files were imported into a custom python 3 script. The python 3 script, epochs.py, used the pywavelets package to complete a wavelet decomposition of any suggested level. The output file contains the annotation and decomposition of each epoch per line unless the annotation is for movement time or not scored. Movement time and not scored epochs are skipped in a majority of the tests. The count of each annotation per file is printed to standard out.

The EEG has a recording every 10 milliseconds. There are 3000 recordings per epoch. This implies that the maximum decomposition level is 8. Each decomposition level produced a different number of coefficients listed in table 7.

Table 7: The number of coefficients produced for a 30 second epoch at each available level of decomposition.

Level	# of Coefficients
8	14
7	26
6	49
5	96
4	190
3	377
2	752
1	1501

The python 3 script for the feed forward backward propogation artificial neural network, softANN.py, accepted the output of epochs.py as input. The input file is put into a special format for the input to the neural network. The real value of the output of the neural network is y = [1,0] if the annotation is R and [0,1] if the annotation is not REM. The input to the neural network consist of the coefficients from 3 epochs: the prior epoch, the current epoch, and the post epoch. If the current epoch is the first epoch in the file then all the coefficients

to the prior epoch are set to zero. If the current epoch is the last epoch in the file then all the coefficients of the post epoch are set to 0.

softANN.py does the training and prediction testing of the artificial neural network. The outputs for training are 4 CSV files and current statistics to standard out at specified intervals. The 4 CSV files are the weights and biases between the input and hidden layers and the hidden and output layers. The current statistics include the cost and the accuracy of the weights and biases. The stopping condition was split between 100% accuracy and cost less than $10e^{-1}$. The output for prediction testing includes the accuracy, sensitivity, and specificity [10].

Results

Because each patient has 2 nights of recordings, there are 40 nights available from the sleep EDF database. Of the 40 nights available, 24 were used. Excluding movement time and not scored, there were an average of 1589 epochs per night. Of those epochs, 89% were not REM and 10% were REM.

Using one of the nights, some preliminary test were run on the softANN.py to determine the optimal wavelet decomposition level, the optimal number of hidden nodes, and the optimal learning rate. It appears any level is capable of achieving 100% accuracy on the test night. Lower levels achieve 100% accuracy while reducing the Cost more than higher levels. The difference in time to 100% between levels is negligible. 10e-5 is the optimal learning rate for all levels. It was found that learning rates 10e-3 and larger caused an overflow. Crasto et al. [13] used principal component analysis to reduce the number of coefficients from approx 3000 to 400. These wavelet levels can reduce the number of components to 14, but level 3 reduces the number of coefficients to 377 which is closest to 400. Since 3 epochs are used, there are 1131 input nodes with 3 levels of decomposition. Various numbers of hidden nodes were tested, but the results were not stored. The number of hidden nodes were set to half the number of input nodes, equal the number of input nodes, and one and a half

the number of input nodes with no difference noticed. The test for training and predictions all used the same number of hidden nodes as input nodes.

The training data sets were all used 1, 2, or 4 nights of data. Of the 8 trainings, 1 was stopped by the user, 1 resulted in overflow, and 1 had the weights lost as a result of a copying error. The training stopped by the user was stopped after 3 days. No other training took longer than 12 hours. All of the training was stopped in the program when the accuracy of the current data set reached 100% except 1 which was stopped after the cost reached 10e-1. The results of the trainings are listed in table 8. Each training set was given an id based on the level of decomposition and the number of iterations of training. The level of decomposition is the first number and the number of iterations is the second number in the id.

Table 8: The results of the training sessions.

ID	DATA	Learning Rate	Result	Final Cost	STOP
3.15000	Patient 001 + 031 + 101 + 161	10e-5	Copy Error	-4.45	Accuracy = 100
3.3000	Patient 001	10e-5	Complete	-16	Accuracy = 100
3.36000	Patient 001	10e-5	Complete	-0.99	Cost <10e-1
3.6000	Patient 001 - >101	10e-5	Complete	-26.9	Accuracy = 100
3.85000	Patient 001 + 101	10e-5	Complete	-16.16	Accuracy = 100
2.5500	Patient 001 + 101	10e-5	Complete	-6.07	Accuracy = 100
2.?1	Paitent 001+031+101+161	10e-5	OVERFLOW		Accuracy = 100
2.?2	Paitent 001+031+101+161	10e-7	Stopped		Accuracy = 100

Each training session produced weights and biases that were used to make test predictions. the test predictions were completed with all of the remaining nights. The average accuracy,

sensitivity, and specificity are listed in table 9.

Table 9: The results of the training sessions.

ID	Accuracy	Sensitivity	Specificity
3.15000	89.98%	38.07%	96.38%
3.3000	79.67%	36.01%	89.82%
3.36000	84.40%	29.89%	91.83%
3.6000	85.82%	17.54%	93.30%
3.85000	81.12%	42.79%	88.20%
2.5500	83.49%	39.14%	88.63%

CHAPTER IV.

CONCLUSIONS AND FUTURE

Conclusions

There are many factors left formerly untested: Over-fitting as a stopping condition in training, the number of nodes in the hidden layers, the number of hidden layers, more than 2 nights as input, level 2 decomposition, etc.

The best result is that the ability to pick REM sleep is not good. The percentage of correctly chosen REM was less than 50% a majority of the time. Sensitivity increases when the number of nights included in the input increased. There are not enough test to know if lowering the decomposition level had an effect.

A positive result is the specificity which was consistently near or above 90%.

The results support the idea that all sleep stages besides REM are easier to find in a single channel EEG.

Future Work

In future studies, every aspect of this study could be changed to improve necessary results. The choice of EEG channel. The choice of sleep channel. The choice of sleep stage could be changed. The choice of epoch size could be changed. Of time-frequency techniques, the wavelet transform is the best choice, but the modification of wavelet transform should be changed to possibly a continuous wavelet transform or a new modification of the discrete wavelet transform. The support vector machine, multilayer perceptron, convolution neural networks, deep feed forward neural networks, and the hidden Markov model should all be tested for classification.

This study was limited to a choice between two sleep channels. For at home studies, there are different devices that use different channels. One device uses a two electrodes around the Fpz electrode position. This position is basically on the forehead of a person and would be one of the easiest positions to locate for an untrained person.

The goal of this paper is to support an at home sleep study. A sleep study needs to be able to detect all stages, but it is very important for a sleep study to identify sleep from wakefulness. If REM can not be identified, all stages can not be identified with a single channel EEG.

The epoch was originally defined based on the ease of use. The complication of paper size and time required for an analysis can be eliminated in choice of epoch size. A training dataset may be difficult to find for a small epoch size, but each individual wave can be classified based on the definitions provided by the AASM. This could result in epochs being scored literally by which waves make up the majority of the epoch.

Time-frequency analysis has many new modifications for discrete wavelet transforms. These modifications should be tested for improvements to the current discrete wavelet transforms. Also, the continuous wavelet transform (CWT)can contain more detail than the discrete wavelet transform (DWT). The CWT should be test against the new modifications of the DWT.

BIBLIOGRAPHY

[1] 10/20 system | Polysomnography Study Guide.

[2] *10/20 System Positioning Manual*. Trans Cranial Technologies.

[3] European Data Format (EDF).

[4] Machine Learning FAQ.

[5] The Sleep-EDF Database [Expanded].

[6] Time-frequency Signal Analysis with Applications.

[7] Spectral density, July 2017. Page Version ID: 793299266.

[8] Achermann, P. EEG analysis applied to sleep. *Epileptologie*, 26:28–33, 2009.

[9] B, v. S., B, K., Ha, K., and Ea, V. d. V. Alternative electrode placement in (automatic) sleep scoring (Fpz-Cz/Pz-Oz versus C4-A1). *Sleep*, 13(3):279–283, June 1990.

[10] Baratloo, A., Hosseini, M., Negida, A., and El Ashal, G. Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. *Emergency*, 3(2):48–49, 2015.

[11] Berry, MD (Chair), R. B., Brooks, MEd, R. R. R., Gamaldo, MD, C. E., Harding, MD, S. M., Lloyd, MD, R. M., Quan, MD, S. F., Troester, DO, M. M., and Vaughn, MD, B. V. *AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, volume 2.4. American Academy of Sleep Medicine.

[12] Byrne, C. L. *Signal Processing, 2nd Edition*. CRC Press, 2 edition, November 2014.

[13] Crasto, N. and Upadhyay, R. Wavelet Decomposition Based Automatic Sleep Stage Classification Using EEG. In *Bioinformatics and Biomedical Engineering*, pages 508–516. Springer, Cham, April 2017.

[14] Debnath, L. and Shah, F. *Wavelet Transforms and Their Applications*. Springer, November 2014. Google-Books-ID: qPuWBQAAQBAJ.

[15] Dietsch, G. Fourier-Analyse von Elektrencephalogrammen des Menschen. *Pflger's Archiv fr die gesamte Physiologie des Menschen und der Tiere*, 230(1):106–112, December 1932.

[16] Ebrahimi, F., Mikaeili, M., Estrada, E., and Nazeran, H. Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1151–1154, August 2008.

[17] Fraiwan, L., Lweesy, K., Khasawneh, N., Wenz, H., and Dickhaus, H. Automated sleep stage identification system based on timefrequency analysis of a single EEG channel and random forest classifier. *Computer Methods and Programs in Biomedicine*, 108(1):10–19, October 2012.

[18] Heaton, J. *Artificial Intelligence for Humans Volume 3: Deep Learning and Neural Networks*. December 2015.

[19] Hlawatsch, F. and Boudreaux-Bartels, G. F. Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine*, 9(2):21–67, April 1992.

[20] Kurt, M. B., Sezgin, N., Akin, M., Kirbas, G., and Bayram, M. The ANN-based computing of drowsy level. *Expert Systems with Applications*, 36(2):2534–2542, March 2009.

[21] LazyProgrammer.me,. Data Science: Deep Learning in Python.

[22] LazyProgrammer.me,. Data Science: Linear Regression in Python.

[23] LazyProgrammer.me,. Data Science: Logistic Regression in Python.

[24] Lotte, F., Congedo, M., Lcuyer, A., Lamarche, F., and Arnaldi, B. A review of classification algorithms for EEG-based braincomputer interfaces. *Journal of Neural Engineering*, 4(2):R1, 2007.

[25] Rao, T. and Vishwanath, D. D. Detecting sleep disorders based on EEG signals by using discrete wavelet transform. In *Green Computing Communication and Electrical Engineering (ICGCCEE), 2014 International Conference on*, pages 1–5. IEEE, 2014.

[26] Rechtschaffen, A. and Kales, A. *A manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*.

[27] Shambroom, J. R., Fbregas, S. E., and Johnstone, J. Validation of an automated wireless system to monitor sleep in healthy adults. *Journal of Sleep Research*, 21(2):221–230, April 2012.

[28] Sinha, R. K. Artificial Neural Network and Wavelet Based Automated Detection of Sleep Spindles, REM Sleep and Wake States. *Journal of Medical Systems*, 32(4):291–299, August 2008.

[29] Ventouras, E. M., Monoyiou, E. A., Ktonas, P. Y., Paparrigopoulos, T., Dikeos, D. G., Uzunoglu, N. K., and Soldatos, C. R. Sleep spindle detection using artificial neural networks trained with filtered time-domain EEG: A feasibility study. *Computer Methods and Programs in Biomedicine*, 78(3):191–207, June 2005.

[30] lvarez, D., Cerezo-Hernndez, A., Lpez-Muiz, G., Castro, T. .-D., Ruiz-Albi, T., Hornero, R., and Campo, F. d. Usefulness of Artificial Neural Networks in the Diagnosis and Treatment of Sleep Apnea-Hypopnea Syndrome. 2017.