PAP THERAPY AND SLEEP STAGE CLASSIFICATION WITH STATISTICS, SIGNAL PROCESSING, AND DEEP LEARNING

By

Gabriel Toban

A Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in

Computational & Data Science

Middle Tennessee State University

March 2023

Dissertation Committee:

Dr. Don Hong, Chair

Dr. Joshua L. Phillips

Dr. John Wallin

Dr. Qiang Wu

Dr. Lu Xiong

ABSTRACT

This paper focuses on PAP (Positive Airway Pressure) therapy compliance and automated sleep staging. These are related to sleep medicine, data science, and signal processing. During a sleep study, an EEG (electroencephalogram) recording is labeled with sleep stages by a technician. The sleep stage labels combined with other recordings, observations, and analysis are used to diagnose a patient with obstructive sleep apnea. PAP therapy is the primary treatment for obstructive sleep apnea. Compliant use of the PAP thereapy is considered 4 hours of use on 70% of nights. The PAP therapy focus is a retrospective study on patients from OSA In Home (formerly Sleep Centers of Middle Tennessee) to find the impacts of the dispenser of PAP therapy between durable medical equipment (DME) supplier (DME group) versus those provided directly by an integrated sleep practice (ISP group). The ISP group had a significantly higher rate of PAP adherence at 30 days (71% vs 66%; P = .004), 90 days (66% vs 56%; $P < .00001$), and 1 year (52% vs 33%; $P < .00001$) following initiation of PAP therapy, relative to the DME group. There was a significantly higher duration of PAP use among the ISP group at 30 days (357 vs 345 minutes; P = .002), 90 days (348 vs 319 minutes; $P < .00001$), and 1 year (312 vs 164 minutes; $P < .00001$). The automated sleep stageing focus is a study on REM sleep classification with interpretable models. The interpretability of the model is based on relating the filters of the CNN (Convolutional Neural Network) part of the model to clincal sleep markers used by technicians to grade EEGs during a sleep study. The model is created using CNNs, full connected neural networks, and GRUs (gated recurrent unit). The data is a unprocessed or raw single channel EEG from "The Sleep-EDF Database [Expanded]" [48]. The best results produced 97% accuracy, 93% precision, and 89% recall. The data is then preprocessed by DWT (multilevel discrete wavelet transform) with 45 different mother wavelets and reconstructed as input to the CNN. The best f1 score with DWT preprocessing and with raw data are 95% with coiflet 4 mother wavlet and 93% respectively. The filters of the CNNs are

compared with Spearman's rank coefficient to find filters that are most correlated in each EEG frequency band. The most correlated filter was preprocessed with Daubechies order 8 mother wavelet had an aboslute value of the Peasron's corelation coefficient of greater than 0.3415 with 15% of the other filters related to the same frequency band.

ACKNOWLEDGEMENTS

This paper would not have been possible without the assistance of many people that I owe many thanks. Dr. Josh Phillips' work in long short term memory and advice in his courses inspired me to use Recurrent Neural Networks in my research. Dr. Qiang Wu's and Dr. Lu Xiong's work in computational methods inspired me to look deeper into the mathematics of deep learning. Dr. Shuzhe Xu explained wavelet coefficients, coded the wavelet decomposition in MATLAB, and provided sound advice throughout the project. William H. Noah, MD and Moataz Toban, MD explained sleep staging and sleep medicine in general. Cynthia Chafin worked on grants, collaborations, opportunities, and much more to help much of my work. Dr. Noah and Cindy founded The Sleep Research Consortium which provided guidance and support on many papers. Todd Gary, PhD provided advice and connected me with so many opportunities. Dr. Don Hong steered the direction of this dissertation with a seemingly infinite depth of knowledge. Dr. John Wallin is the reason this paper is possible. There are many more who are not directly mentioned. Thank you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I.

Introduction and Background

1.1 Introduction

Data science is an area of study characterized by the use of computer science, mathematics, and subject matter expertise to gain knowledge from information or data. Sleep research advances topics like the algorithms, equations, and theories used to understand sleep, the impact of sleep on all aspects of health, and the practice of clinical sleep medicine. The following sections and chapters will cover the history, background, methods, data, and results of several projects that use data science in sleep research.

Several different topics in sleep research may be amenable to study using data science techniques. Some examples include PAP (positive airway pressure) therapy compliance, Rapid Eye Movement (REM) sleep classification on raw data (REM sleep classification), and REM sleep classification on data processed with wavelet decomposition (wavelet deep learner).

This chapter covers the background and history of each project. The next 3 chapters, Chapters 2, 3 and 4, each provide information specific to each project. This information includes any collaborations, data, considerations, similar projects, methods, and results. Each chapter also includes a discussion about the future of each project and next steps that could be taken.

1.2 PAP Therapy

Obstructive sleep apnea is a medical condition characterized by brief and repeated episodes of upper airway collapse during sleep. Untreated obstructive sleep apnea has been shown to have numerous adverse effects, including impaired daytime functioning, increased mortality, neurocognitive impairment, increased risk of traffic accidents, and higher chronic disease burden[27, 59, 61]. Despite the fact that positive airway pressure (PAP) therapy remains one of the most important treatment options for obstructive sleep apnea, long-term

adherence remains a major barrier to successful treatment [27, 74, 95].

Suboptimal adherence is not a problem unique to the use of PAP therapy. In developed countries worldwide, nonadherence rates for chronic disease care are estimated to be between 25% and 50% [42, 75, 35]. Under the World Health Organization framework [77], nonadherence to treatments for chronic disease is considered in a context broader than a patient's acceptance of recommended therapy. Rather, adherence is viewed as a multifaceted interplay between numerous factors, such as the patient, the health care delivery system, the medical condition, the nature of therapy, and the socioeconomic environment. The health care delivery system, in particular, can affect the accessibility of services, data availability, procedures around health information management, and training of clinicians. Regardless, programs that support adherence are characterized by robust lines of communication between the patient and his/her health care team [101] ideally in support of behavioral change.

The sleep medicine literature is rich with studies exploring demographic, clinical, and technological factors predictive of adherence to PAP therapy. Such studies have most consistently suggested that PAP adherence has been associated with consistent PAP usage early in the course of treatment [49, 24, 94, 14, 73, 60], apnea-hypopnea index [60, 47, 62], pretreatment sleepiness [49, 60, 47, 66], and perceived improvement in sleepiness after treatment [60, 62, 82]. The correlation with other seemingly important factors, such as patient age, sex, race, and presence of side effects, has been less consistent [95]. In addition to these factors, recent fascinating work has studied the impact that psychological constructs from the behavioral change literature [85, 68, 67, 69, 19, 20] seem to play on continuous PAP adherence [85, 15, 78].

1.3 Sleep Stage Classification

Sleep stages are the most precise way to separate wakefulness from sleep state. Physiological 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.

Sleep stages are typically identified with a combination of biological signals recorded during a polysomnography (PSG) or sleep study. These signals include, but are not limited to, EEG derivations, electrooculogram (EOG) derivations, and a chin electromyogram (EMG) [22]. Trained professionals and medical providers can score these signals by finding clinical markers in the signals. In this case, the clinical markers are time dependent and independent features of a signal with standards maintained by The American Academy of Sleep Medicine (AASM).

Algorithms used to extract features from and automatically score a signal are well established. The use of biological signals, such as the EEG, has been around for over 100 years. In 1968, the Rechtschaffen and Kales (R&K) sleep scoring standard [99], the first accepted attempt to standardize sleep scoring, was created. Automated analysis of EEG data has been around for over 50 years [84]. The AASM position states that home sleep apnea test (HSAT) with automatic scoring can be used by clinicians, but the raw data from HSAT recordings must be reviewed before decisions about a patient may be determined [8].

Many different algorithms can be used to categorize sleep stages. Successful methods include Random Forrest for spindle detection [97, 76], support vector machines (SVM) [12], wavelet decomposition [41], common frequency/spatial pattern [46, 16], autoencoders [55], recurrent neural networks (RNN) [64], graph theory [26], and entropy calculations [64]. By far, the most common method consists of using Convolutional Neural Networks (CNN) and RNNs [29, 93, 88, 87, 28, 56, 39]. Work has been completed to reduce the complexity of algorithms for sleep stage classification[100]. Our research contributes to the field by producing a model to increase the explainability of algorithms.

1.3.1 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 [22, 2, 1]. The EEG channels that are best to use were standardized originally by Rechtscheffen and Kales [71] 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. [22]. Waves, activities, or events, such as spindles or arousals, can be identified by a combination of these identifying factors, the frequency, and the amplitude [22].

Table 1: The frequency and peak-to-peak amplitude of different brain waves as identified by the AASM [22].

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 3. 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].

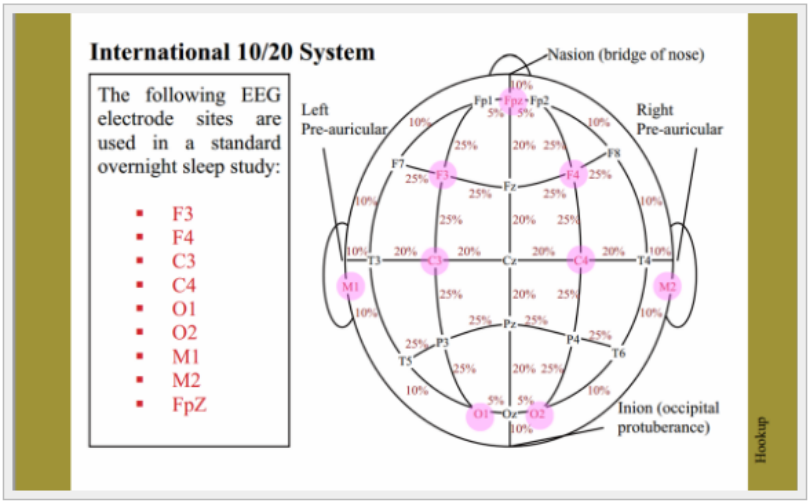


Figure 1: A representation of the electrode positions identified by the 10/20 system. Each position is labeled by 1 to 3 letters and/or a number which are listed on the left side of this image. Image courtesy of https://sleeptechstudy.wordpress.com/category/1020-system/ [1].

Table 2: The brain region that relates to every electrode location[1].

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

Table 3: The brain region that relates to every electrode location[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 4. The data used for this study does not use the channels recommended by the AASM [6], but the efficacy of the alternative electrode positions are described by van Sweden et al. [17].

Table 4: The channels for an EEG to identify sleep stages as recommended by the AASM [22]. 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 5 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 5: 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 [22]. 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 information except "Adequate" was gathered from the AASM scoring guide [22]. "Adequate" comes from the R&K standard [71]. E1-Fpz refers to 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 [71]. The AASM defines 5 stages to identify sleep [22]. The AASM definition excludes Movement Time and combines stage 3 and 4. Table 6 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 [6].

Table 6: The different sleep stages as identified by the R&K [71] and the AASM [22]. 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 [22]. 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 7 gives an idea of the relationship between EEG markers and sleep stages.

Table 7: Which sleep stages are identified by certain EEG markers as recommended by the AASM [22]. Not every possible marker is identified. 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 [22] recommends that at least 3 EEG channels, 2 EOG channels, and a chin EMG among other things for a PSG. Ventouras et al. [90], Ebrahimi et al. [36], and Crasto et al. [31], all used a single EEG channel with accuracies between 85% and 95%. Shambroom et al. [81] 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 5, and identify stage N3, Table 7. Although, Crasto et al. [31] 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 5 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 [22, 71]. Each epoch is assigned a single sleep stage [71] [22]. If an epoch exhibits features of more than one stage, the R&K standard [71] strictly assigns each epoch the stage that takes the most of the epoch . The AASM [22] has additional rules when the features of more than one stage are exhibited in any epoch. The R&K standard [71] 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 [6] 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 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 7). Stages N1, N2, and N3 all have a unique EEG marker to identify them. Current study will concentrate on REM sleep.

1.4 Signal Processing

A signal is a quantity that varies over space or time. There are numerous methods for algebraic and computational analysis of signals. A computational analysis of a signal starts

by taking samples of the signal. The following provides a brief description of computational analysis of signals in 1 dimension in time otherwise known as digital signal processing.

A signal can be defined by various features of the rising or falling of the quantity. A peak in a signal is any point that is higher than it's neighboring points. A trough in a signal is any point lower than it's neighboring points. A wavelength is the amount of time between two peaks or troughs. The frequency is the number of wavelengths that happen over a given time.

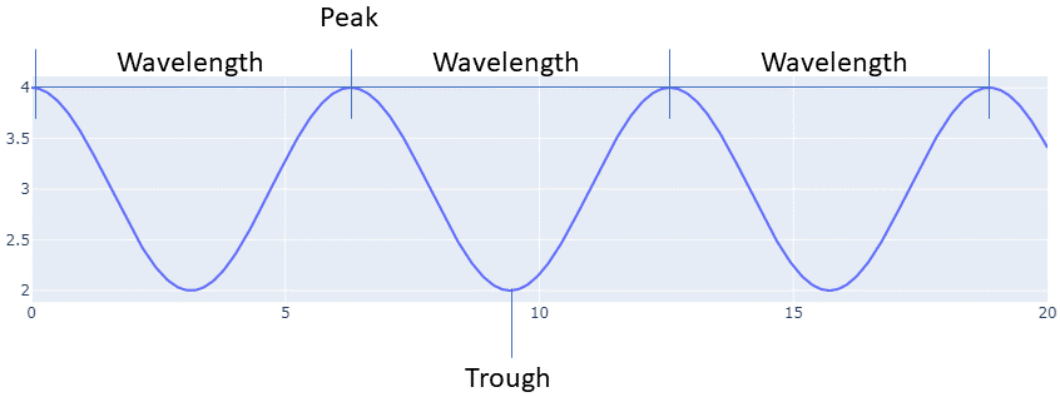


Figure 2: Signal Attributes

Part of signal processing involves finding and extracting frequencies from a signal. Two signals of the same number of samples can be combined by adding each sample in the same time step to create a signal with two frequencies. It requires computational methods to start with a signal with two frequencies and end with two signals with 1 frequency.

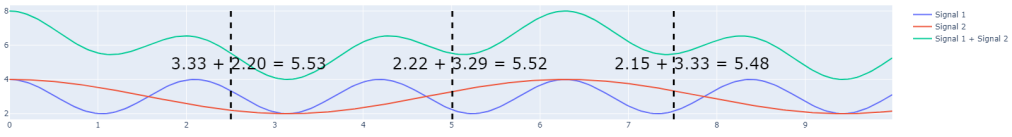


Figure 3: Combining Signals

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 [25] states that the Fourier series is the most

important tool in signal processing. Achermann [10] stated that Dietsch [34] performed the first Fourier analysis on an EEG in 1932.

The Fourier series of a function is [33]

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

where

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

The Fourier integral theorem takes the limit as ℓ approaches infinite of the Fourier series [33] to define a non-periodic function.

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

The Fourier transform of a signal is derived from the Fourier integral [33].

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

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 [7]. The spectral density is a representation of the amplitude of a wave[9], but the sinusoidal component of the transform does not allow if to capture local information [33]. In other words, abrupt changes in amplitude may not be captured. It is well-known that Fourier Transforms cannot investigate time and frequency simultaneously [83, 36, 50, 31, 33].

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” [22] 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” [22]. Figure 4 shows the V-Wave and K-complex.

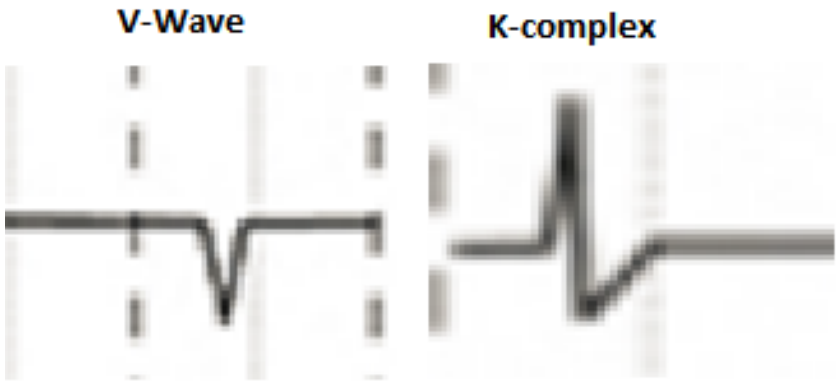


Figure 4: The V wave and K-complex used in the sleep stage scoring manual from the AASM [22].

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) [44]. Several of the other representations were modifications of WD [44]. 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” [33]. The WVD has been modified many times to create many time-frequency distributions [33] [7].

A general review of the methods used shows that Fourier and Wavelets are the primary representations used for analyzing an EEG in sleep staging [102]. The word “wavelets” means “small waves” (the sinusoids used in Fourier analysis are “big” waves), and in short, wavelet is an oscillation that decays quickly. Mathematically, wavelets usually are

basis functions of an L^2 space that satisfy so-called multi-resolution analysis requirements [32, 45, 58]. Wavelets, as building blocks of models, are well localized in both time and scale (frequency). Signals with rapid local changes can be precisely represented with just a few wavelet coefficients. Among those that used Fourier and Wavelets, the Wavelet representations resulted in the highest accuracy results [83] [36] [50] [31] [102]. Fraiwan et al. [38] used Choi-Williams Distribution (CWD), continuous wavelet transform (CWT), and Hilbert-Huang Transform (HHT). Fraiwan et al. [38] found that CWT had the highest accuracy. A Discrete Wavelet Transform (DWT) can decrease the processing time of a CWT [33]. Ebrahimi et al. [36] and Crasto and Upadhyay [31] used Daubechies wavelet order 2 filter with DWT. Crasto and Upadhyay [31] 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 uses the DWT to decompose the EEG signals.

The discrete wavelet transform is based on the concept of a function called the "mother wavelet" [33]

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

This mother wavelet can be translated with b and dilated with a. It is part of the wavelet transform of a function f [33].

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

The function $\Psi_{a,b}(t)$ replaces e^{iwt} in the Fourier integral [33]. $\Psi_{a,b}(t)$ can be one of many wavelets. In the Fourier series, the sinusoidal waves produced from e^{iwt} lacking com-

pact/local support property are the problem. Wavelet bases are orthonormal or orthogonal. The wavelet transform (WT) has been successfully applied over an extraordinary range of fields in order to decompose the non-stationary TS into time-frequency domain as well.

Since Daubechies wavelet order 2 produced high accuracy results in multiple papers [31] [36], this study will use Daubechies wavelet order 2. Figure 5 shows the Morlet (symmetric) and Daubechies-2 (non-symmetric) wavelets, respectively.

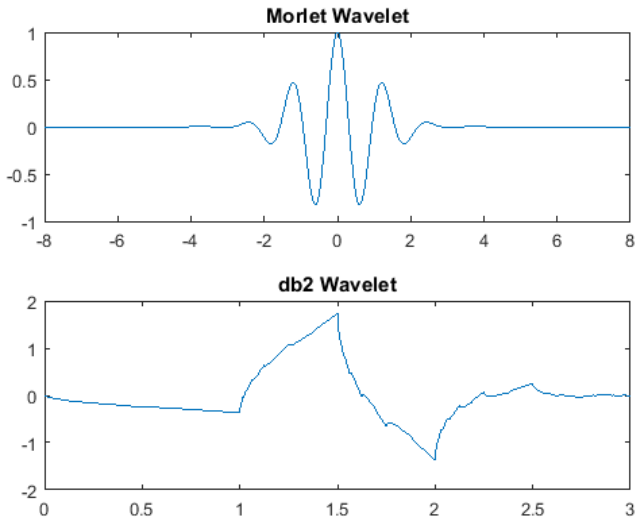


Figure 5: A plot of the Morlet wavlet and the Daubechies order 2 wavelet produced by Matlab.

The discrete wavelet transform has many modifications to help it improve. Ebrahimi et al. [36] 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 will use the multilevel discrete wavelet transform. Each level is the transform of the previous level. Each level is not as close to the original wave as the previous level, but it shows more details to original signal and greatly reduces the number of coefficients associated with the dilation parameter a and translation parameter b described above in the

mother wavelet.

1.5 Deep Learning

There are many types of ANNs. Heaton [43] 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 [31] [83] and multilayer perceptron neural networks [50] [90]. All of these algorithms could be tested.

A neural network produces a non-linear decision boundary [57]. 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 [52]. In other words, there exist a red line that minimizes the sum of the lengths of the blue dashed lines in Figure 6.

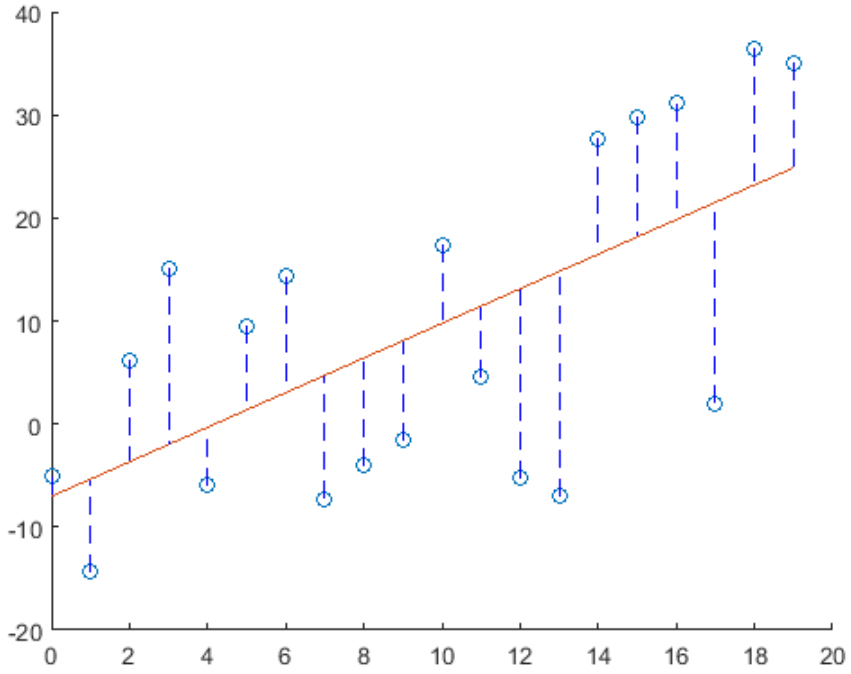


Figure 6: The plot of line that was produced via linear regression on a set of coordinates.

If the coordinates of the line are represented by set $x = [x_i]$ and $y = [y_i]$ for i in range 1 to the number of points n and $\hat{y} = m * x + b$, where $\hat{y} = [\hat{y}_i]$ for $i = 1, \cdots, n$, is the equation of a line with m and b unknown, then the sum of the distances between the line and the set is $\sum_{i=1}^{n} |y_i - \hat{y}_i|$, where $|y - \hat{y}|$ is the absolute value of $y_i - \hat{y}_i$. In linear regression, the the square of $y_i - \hat{y}_i$ 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 squares error equation [52]

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

If $\hat{y}$ is replaced with the equation of a line, the least squares equation becomes the cost function for linear regression [52]

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

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 [52].

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

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

where f is the least squares equation and λ is the step size or the learning rate. The following image can help you visualize gradient descent. [52]

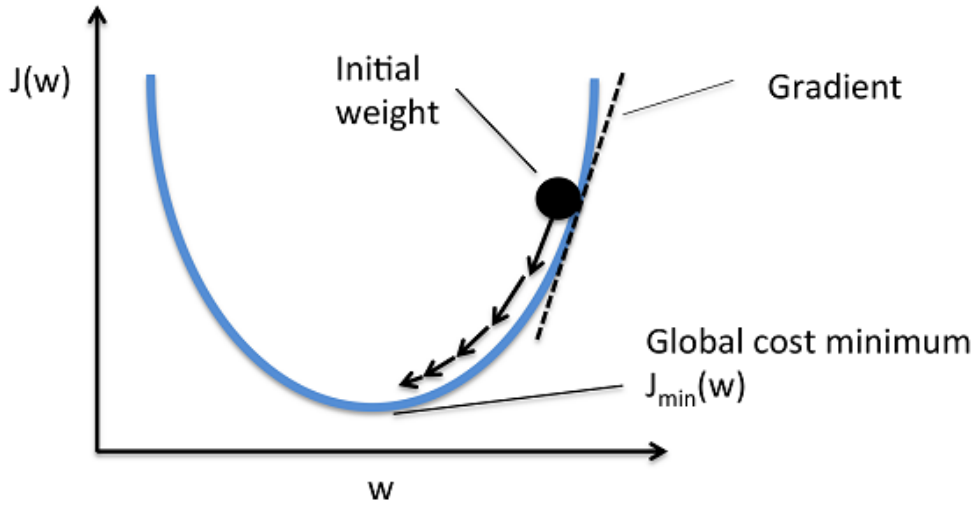


Figure 7: Gradient Descent image from https://sebastianraschka.com/faq/docs/closed-form-vs-gd.html [5].

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.

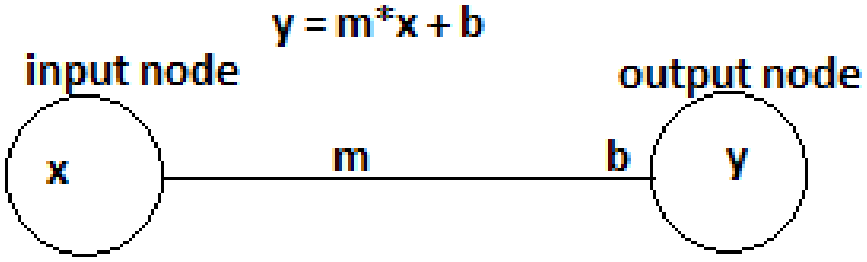


Figure 8: Function in node picture.

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

$$y = m1*x1 + m2*x2 + b$$

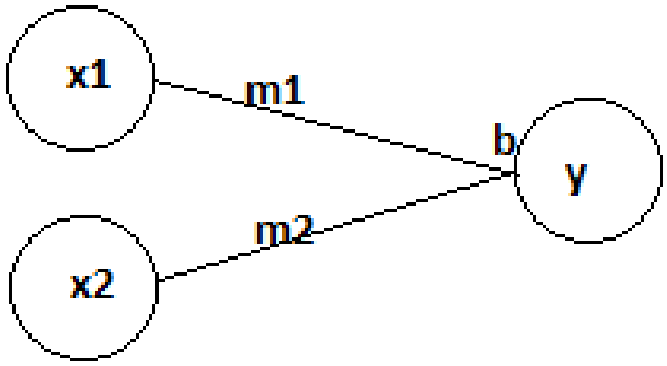


Figure 9: Function with multiple inputs in node picture.

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 [53]

$$s(x) = \frac{1}{1+e^{-x}}.$$

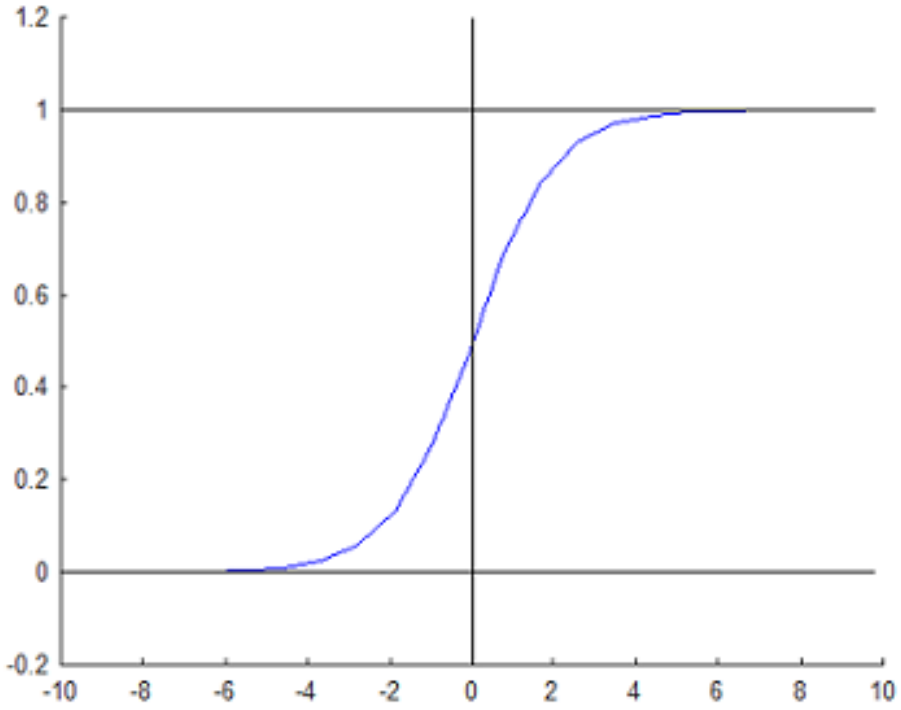


Figure 10: Plot of the sigmoid function.

Once the sigmoid is applied, the nodes picture looks like the following [53].

$$y = s(m1x1 + m2x2 + b)$$

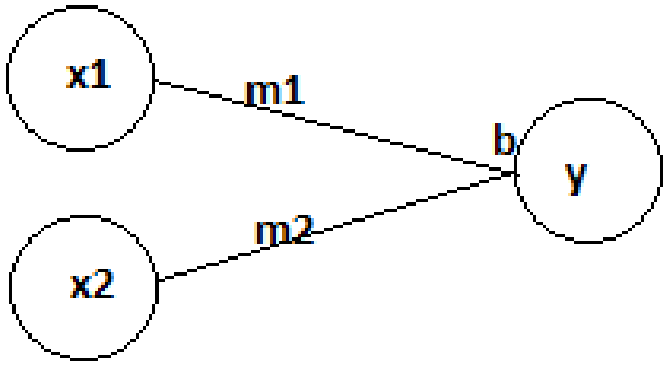


Figure 11: Function with multiple inputs in sigmoid node picture.

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 [53]

$$C = -\sum_{n=1}^{N} y_n log(\hat{y}_n) + (1 - y_n) log(1 - \hat{y}_n),$$

where y is the real solution, $\hat{y}$ 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 [53].

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

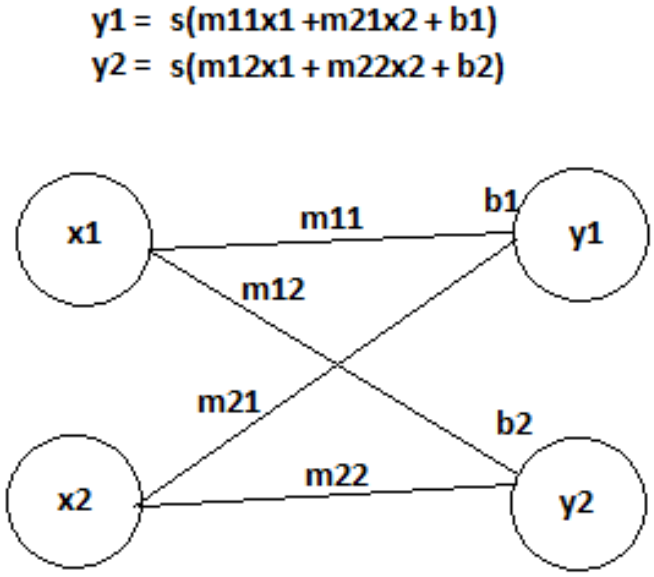


Figure 12: Function with multiple inputs and multiple outputs in a sigmoid node picture.

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

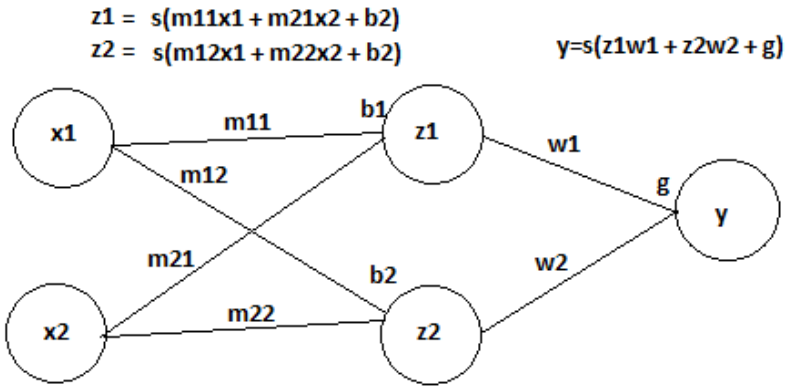


Figure 13: Function with multiple inputs and layers and a single in a sigmoid node picture.

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) [43]. 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 [51].

$$\text{softmax} = sf(a_i) = \frac{a_i}{\sum_{j=1}^{K} a_j},$$

where $i = 1, 2, \cdots, K$ with K the number of nodes in the output layer and

$$a_i = \exp(\sum_{j=1}^{Q} w_{ji} * z_j + b_i),$$

where Q is the number of nodes in the layer previous to the output layer [51]. The nodes picture become the following:

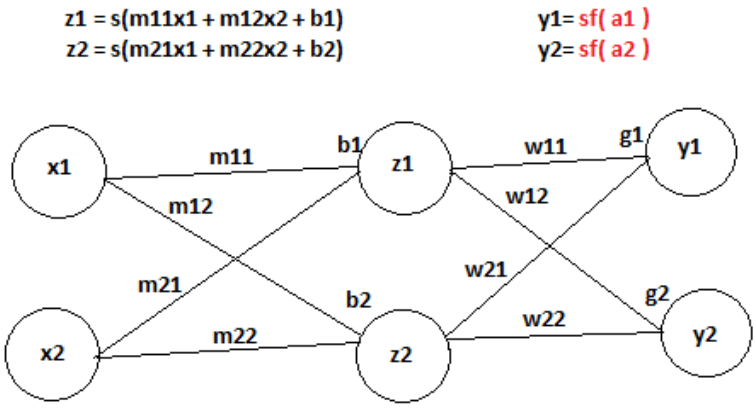


Figure 14: Function with multiple inputs, layers, and outputs in a ANN node picture.

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

$$C = ll = \sum_{s=1}^{S} \sum_{k=1}^{K} y_k^s log(\hat{y}_k^s),$$

where S 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 [51].

1.5.1 Convolutional Neural Networks (CNN)

In convolutional neural networks (CNN), the weights are the multipliers used during convolution.The basic form of convolution can be understood as a signal ($F = (f_1, \cdots, f_k)$) convolved by a filter ($G = (g_1, \cdots, g_n)$) to produce a convolved signal having the same frequency as the filter

$$\text{convolved signal} = F * G = \begin{pmatrix} f_1 * g_1 + f_2 * g_2 + \ldots + f_n * g_n \\ f_2 * g_1 + f_3 * g_2 + \ldots + f_{1+n} * g_n \\ \vdots \\ f_{k-n+1} * g_1 + f_{k-n+2} * g_2 + \ldots + f_k * g_n \end{pmatrix}.$$

A signal is a series of points over time. A filter, in the case of convolution, is a series of weights to act as multipliers on the signal. Each point in the convolved signal is a dot product of the filter and an equal number of points from the signal. This process creates a rolling window of points from the signal that are multiplied by the filter.

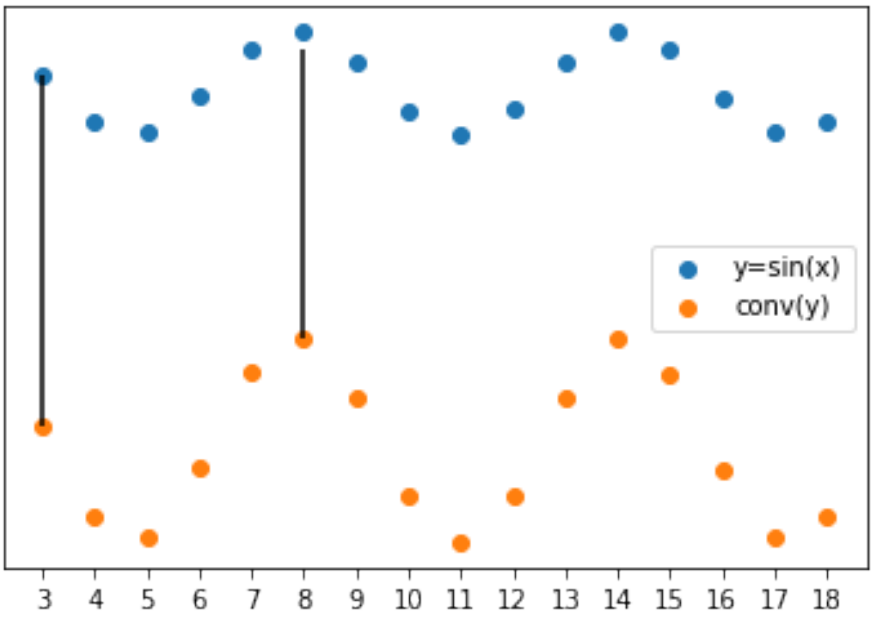


Figure 15: Convolution with Filter = [2]

In Figure 15, the filter has 1 point. Each point in the signal has a 1-1 mapping to the convolved signal by multiplying it by a number. The point in the signal maps to a point in the convolved signal at the same point in time.

$$\text{convolved signal}_0 = f1 * g1$$

$$conv(\sin(3)) = \sin(3) * 2$$

This is not generally true in convolution but illustrates a basic idea. When you choose a filter with more than 1 point, there is not a 1-1 mapping.

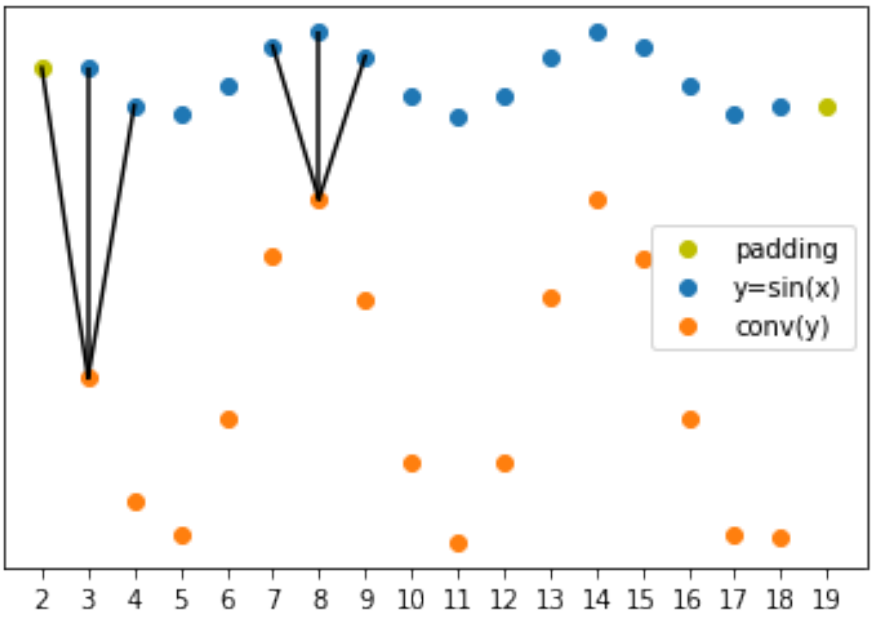


Figure 16: Convolution with Filter = [1,3,1]

In Figure 16, the filter has 3 points. Each point in the convolved signal is the dot product of 3 points in original signal and the 3 points in the filter. For each point time, the points taken from the original signal are the previous point, the current point, and the next point.

$$\text{convolved signal}_0 = f1 * g1 + f2 * g2 + f3 * g3$$

$$conv(\sin(3)) = \sin(2) * 1 + \sin(3) * 3 + \sin(4) * 1$$

In order for the convolved signal to begin and end at the same time as the original signal, an extra point is added to the beginning and to the end. This is known as padding the signal or simply padding.

There are many ways to implement padding and the number of points in the filter. Padding can result in a convolved signal that is bigger, smaller, or the same size as the original signal. The methods used in this algorithm always use padding that results in a convolved signal the same size as the original signal.

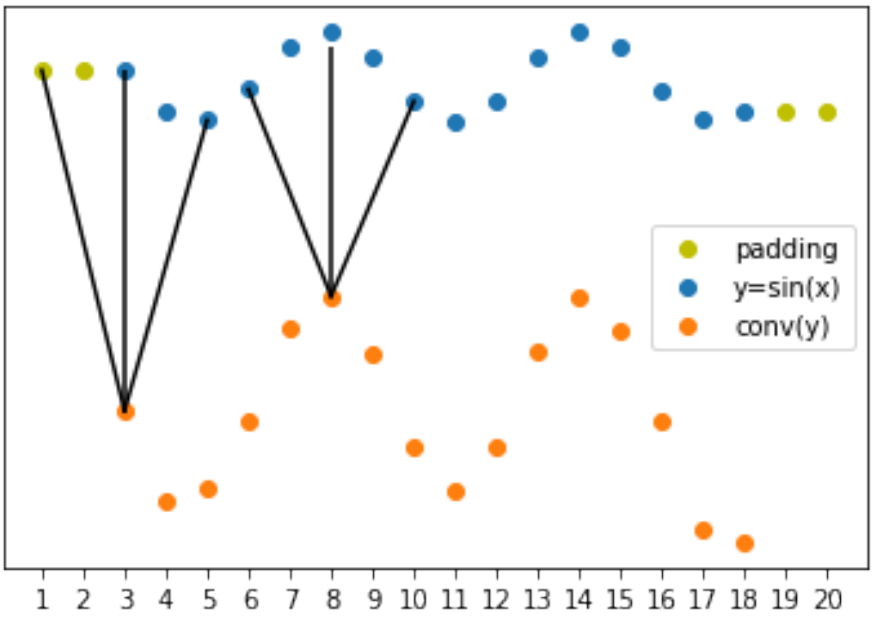


Figure 17: Convolution with Filter = [1,3,1] and Stride 2

The stride in convolution can be thought of as the number of time steps between each point in the original signal used to produce a point in the convolved signal. In figure 17, a stride size of 2 is used.

$$\text{convolved signal}_0 = f1 * g1 + f3 * g2 + f5 * g3,$$

$$\text{conv}(\sin(3)) = \sin(1) * 1 + \sin(3) * 3 + \sin(5) * 1.$$

The effect is using a limited number points to produce the convolved signal. When a stride of 2 is used, about half of the points are used from the original signal. Stride can help find trending information at various frequencies in the signal. In this algorithm, stride is always kept the same.

In convolutional neural networks, the elements of the filter (G) becomes the weights of the neural network layer (Wconv $= w1, w2, \cdots, wn$). These weights like all neural networks can have a bias.

$$\text{if ConvolvedSignal}_0 = f1 * g1 + f3 * g2 + f5 * g3,$$

$$\text{then output}_0 = f1 * w1 + f3 * w2 + f5 * w3 + b.$$

A convolutional unit can contain multiple filters of the same size. The convolutional unit is said to have a filter size consisting of the number of points in the filter or filters and the number of filters.

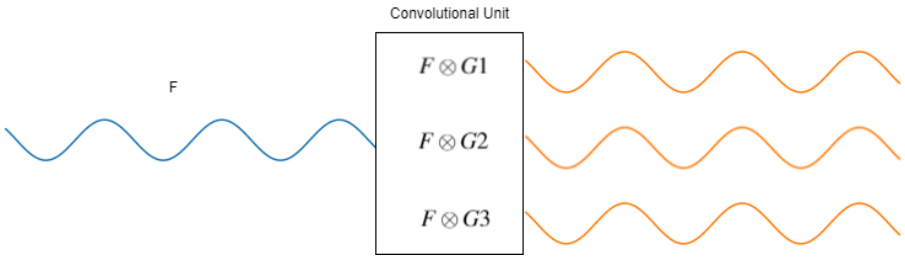


Figure 18: Convolution Unit with 3 Filters

In Figure 18, the convolutional unit has a filter size that includes 3 filters. Each filter convolves the entire signal producing 3 different convolved signals with the same number of points as the original signal. Then the example equation of Wconv becomes the following.

$$signal1_0 = f11 * w11 + f13 * w12 + f15 * w13 + b1,$$

$$signal2_0 = f21 * w21 + f23 * w22 + f25 * w23 + b2,$$

$$signal3_0 = f31 * w31 + f33 * w32 + f35 * w33 + b3.$$

Each filter in a convolutional neural network can be thought of as recognizing a different pattern in the signal. The output of the convolution is the parts of the original signal that match the pattern in the filter.

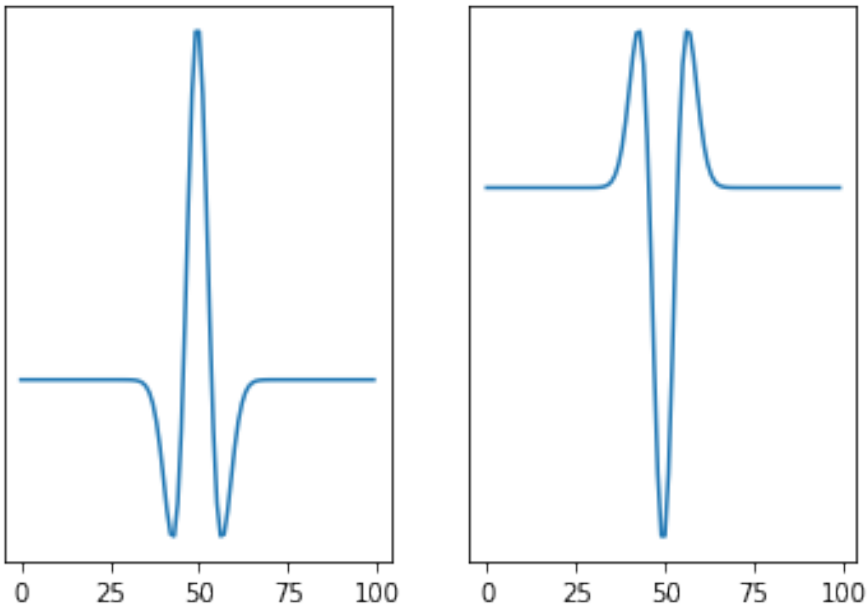


Figure 19: Two Signal Patterns

In Figure 19, there is 1 pattern plotted 2 times. One plot is an exact copy of the other except it is flipped vertically. In general situations, these two patterns could be considered the same pattern. In convolution, these are 2 different patterns. Each of these patterns would require their own filter to be recognized. If this is the exact pattern to find, allowing this to be the filter could extract them from the original signal.

When a convolutional neural network's weights are trained, they may produce patterns from the training data. This could make the convolution understandable. Also, these patterns could be reused in other signal processing analysis.

1.5.2 Recurrent Neural Network (RNN)

A Recurrent Neural Network (RNN), in it's simplest form, takes all or part of it's output and new data as input. In doing so, it uses information from previous outputs to make a decision.

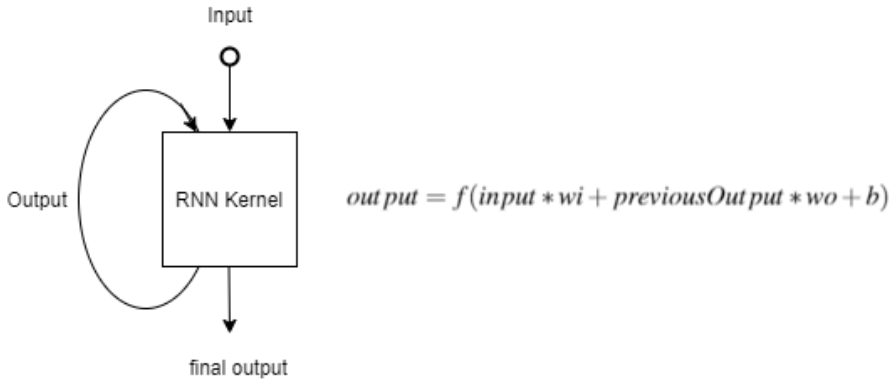


Figure 20: Simple RNN

In figure 20, the output is a linear combination of of the input and the previous output.

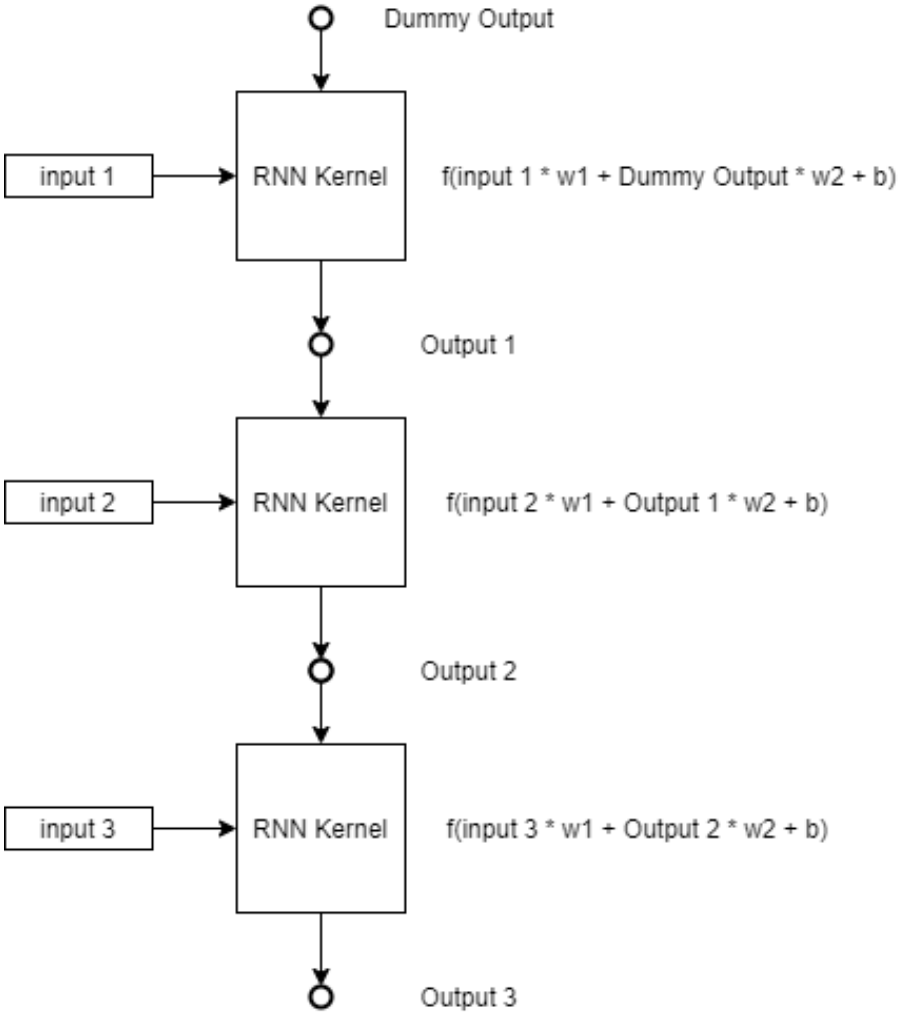


Figure 21: RNN Cell

Figure 21 is the exact same as figure 20 except each recurrent step is expressed explicitly. Each RNN Kernel is a linear function of the previous output and additional input. All of these RNN steps together make an RNN cell. Each input to an RNN Kernel may be a replicated version of the input to the RNN cell or a part of it. The outputs from all kernels can be combined to make the output from a single RNN cell.

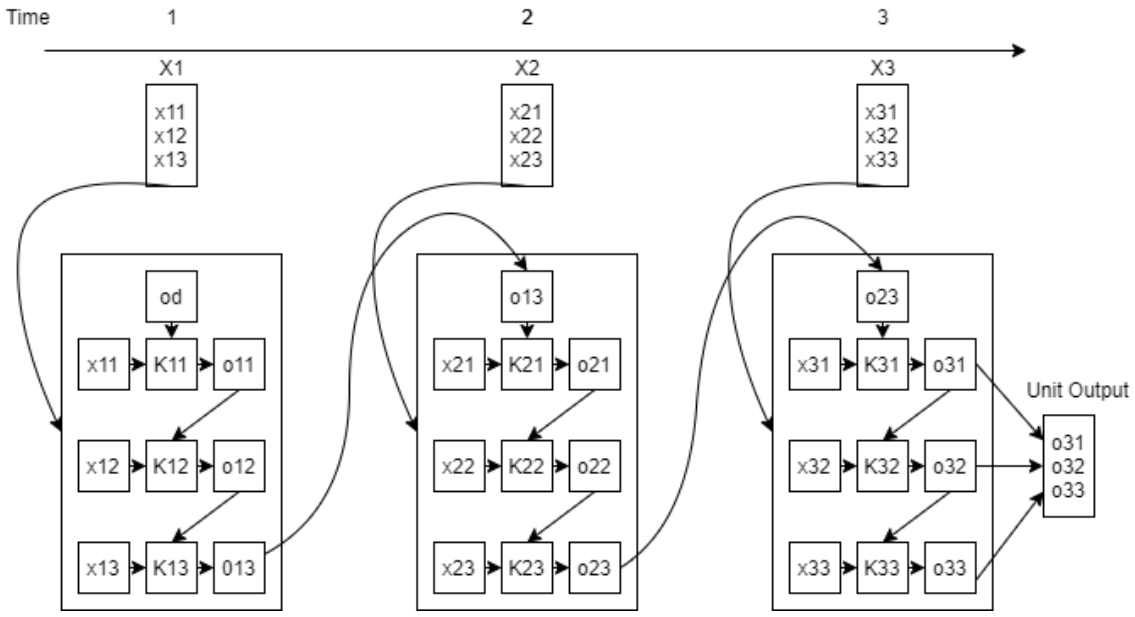


Figure 22: RNN Unit

Figure 22 illustrates how multiples cells can combine to make an RNN Unit or an RNN layer. An RNN cell can pass all the outputs from all of it's RNN kernels or just one of them. The output to the entire unit can be the output from all cells or just one of them. A bidirectional RNN can take the outputs from all cells and process the data backward in time.

One of the major downsides to an RNN is the number of parameters or weights that must be trained. An RNN is prone to over fitting and is notably impacted by the "curse of dimensionality". It can require more data to train an RNN than other algorithms.

The RNN kernel is typically not referred to specifically. An RNN layer can be said to use a different kernel or a different cell meaning the linear combination of inputs and outputs in the kernel is constructed differently. The most popular RNN cells at this time are

Long Short Term Memory (LSTM) and the Gated Recurrent Unit (GRU).

1.5.2.1 Long Short Term Memory (LSTM)

The LSTM is meant to help an RNN remember older information and forget information that's not useful. The LSTM works by using "gates" to add or remove weight for new (input gate i), old (forget gate f), and processed (output gate o) information.

It starts off similar to the equation as a regular RNN. The difference being the previous gated output. The following formula has a weight for the input *wi*, weight for the previous gated output *wo*, and bias *b*.

$$\text{output} = \tanh(\text{input} * wi + \text{previousGatedOutput} * wo + b).$$

This output is then weighted with the 3 gates to create a candidate gate (gatedOutput). The input gate (i), the forget gate (f), and the output gate (o). These gates are defined as an activated linear combination similar to the output each with their own weights (starting wit w) and biases (starting with b). The formula has weights for the input, wieghts for the previous gated output, and bias for each gate. Respectively, the input gate has *wigi*, *wogi*, and *bgi*. The forget gate has $wigf$, $wogf$, and bgf. The output gate has *wigo*, *wogo*, and *bgo*. The gates have to use sigmoid activation function while the output can use any activation function.

$$\begin{aligned}
\text{gatedOutput} &= o * \tanh(f * \text{previousOutput} + i * \text{output}),\\
i &= \text{activation}_g(\text{input} * wigi + \text{previousGatedOutput} * wogi + bgi),\\
f &= \text{activation}_g(input * wigf + \text{previousGatedOutput} * wogf + bgf),\\
o &= \text{activation}_g(input * wigo + \text{previousGatedOutput} * wogo + bgo).
\end{aligned}$$

The LSTM has more than 3 times as many weights as a simple RNN. The GRU algorithm attempts to reduce this problem by reducing the number of weights in a gated algorithm.

1.5.2.2 Gated Recurrent Unit (GRU)

The GRU attempts to accomplish the same task as the LSTM with fewer weights. It uses ”gates” to add or remove weight for old (reset gate f) and processed (update gate o) information. Each gate and output have thier own weights (starting with w) and outputs (starting with b).

$$f = \text{activation}_g(input * wigf + \text{previousGatedOutput} * wogf + bgf),$$
$$o = \text{activation}_g(input * wigo + \text{previousGatedOutput} * wogo + bgo),$$
$$\text{output} = \tanh(\text{input} * wi + \text{previousGatedOutput} * f * wo + b),$$
$$\text{gatedOutput} = (1 - o) * \text{output} + o * \text{previousGatedOutput}.$$

The GRU has fewer weights than the LSTM and more weights than a simple RNN. While it still has significantly more weights than a simple RNN, it has less weights than a LSTM.

CHAPTER II.

PAP Therapy Compliance

2.1 Introduction

One of the factors influencing health behaviors that has been underexplored in the PAP adherence literature are the structural health system impediments [20] that limit the effectiveness of individuals' attempts at self-management. Under the traditional model of PAP fulfillment, a clinical provider writes a prescription for PAP therapy, while a durable medical equipment (DME) company supplies the machine. Because of the numerous entities involved in PAP care delivery, this division of labor has the potential to contribute to unintended effects such as fragmented communication and heterogeneous troubleshooting procedures. This may lead not only to an additional barrier but also to lost opportunity to educate about the associated health risks and facilitate adherence. Furthermore, the cognitive link between PAP device usage and health outcomes may become less apparent to the patient when much of the PAP instruction is provided outside of a medical practice. Often, the prescribing provider can be left out of the troubleshooting process altogether before a patient has made the decision to discontinue therapy. This may be problematic, given some evidence that an existing relationship with a sleep specialist prior to diagnosis of sleep-disordered breathing (SDB) is associated with higher PAP adherence [65]. Similarly, adherent patients tend to be more satisfied with the level of communication and education they have received from the providers and staff providing their sleep care [98, 89]. Practices may also be in a better position to address the common comorbid conditions that can affect PAP adherence [30, 91]. As such, sleep specialty clinics may be well equipped to deliver timely interventions to promote PAP adherence, provided they effectively leverage their interdisciplinary clinical team to address some of the common barriers to PAP adherence.

The purpose of this study was to explore differences in PAP adherence between patients receiving PAP therapy from a traditional DME provider and those patients receiving PAP

therapy through an integrated sleep practice (ISP). We hypothesized superior PAP adherence among patients who receive PAP therapy directly from an ISP, relative to those receiving PAP therapy from a DME supplier.

2.2 Methods

This study was approved by the Institutional Review Board at Middle Tennessee State University.

2.2.1 Study population

This retrospective cohort study followed patients who received their sleep care from Sleep Centers of Middle Tennessee, a community-based sleep practice serving approximately 20,000 patients per year. Providers and staff at Sleep Centers of Middle Tennessee are provided specific training to support patients using principles of evidence-based behavioral sleep medicine [15, 13, 72]. The training was designed to be applied by the provider at initial evaluation and follow-up visits, by the polysomnographic technologist at sleep studies, and by DME staff during the initial PAP set-up visit.

We collected PAP adherence and demographic data from patients who initiated PAP therapy between January 1998 and June 2018. As such, those patients diagnosed with SDB but who pursued an alternative treatment modality or were lost to follow-up prior to receiving PAP therapy were not included. For a best attempt at including only the continuous PAP–naive population, patients with more than 1 continuous PAP provider over this 20-year interval were excluded from the final analysis. In addition, patients who started PAP therapy multiple times with the same DME provider had only data from the first trial analyzed. To exclude age confounders introduced by anti–self-referral (ie, Stark) regulations, patients 65 years of age or older at the time of PAP therapy initiation were excluded from the final analysis. Patients younger than 18 years of age or missing demographic characteristics were also excluded. Finally, to ensure that data were reliably transmitted via cellular modem, we included only patients from a subset of 4 DME companies that started treatment with PAP

after 1 January 2012. After this date, these DME providers agreed to keep wireless modems turned on indefinitely.

2.2.2 Procedures

All patients included in the study, regardless of the experimental group, had initial sleep clinic visits with the same providers from the same sleep practice. Both groups of patients received the same patient educational content during this visit. The procedures surrounding technical administration, interpretation, and results communication of diagnostic studies were also the same for each group of patients.

Upon receiving a diagnosis of SDB, patients were allowed to choose their DME provider, based on insurance compatibility and personal preference. Patients received their PAP supplies from either the ISP (ISP group) or 1 of 4 participating DME suppliers (DME group) who agreed to leave the wireless modem activated indefinitely. For patients initiating PAP therapy in the ISP group, DME staff were trained to elicit expectations and anticipated problems with continuous PAP therapy, then to counsel and educate patients with the principles of motivational enhancement in mind [13]. Participating DME providers were also provided with a written copy of the practice's PAP initiation protocols.

The protocol for scheduling follow-up visits was the same for patients from each group, regardless of the DME provider. In general, the interval for each follow-up visit was guided by the participants' success with PAP adherence. Patients were asked to return for short-term follow-up (ie, 3 months or less) until they achieved an average usage of at least 6 hours per night. Once a patient achieved an average of at least 6 hours of nightly adherence, they were scheduled for yearly follow-up.

2.2.3 Statistical analysis

Adherence rates for each of the groups were calculated at 30 days, 90 days, and 1 year after the initiation of PAP therapy. An individual patient was classified as being adherent during each of these intervals if he/she had at least 4 hours of PAP usage for at least 70%

over the respective interval. For example, a patient would be classified as adherent at 90 days if he/she used PAP for at least 4 hours on at least 63 of the first 90 days of use. Adherence rates between the experimental groups were compared with chi-square test at each of these intervals.

Average nightly durations of PAP use at 30 days, 90 days, and 1 year were also compared between groups. Each patient's PAP usage was averaged over the respective time interval. Given nonparametric distribution of the mean nightly usage for each group, the duration of PAP usage was compared between groups with a 1-sided Mann-Whitney U test.

2.3 Results

2.3.1 Study participants

Between January 1998 and June 2018, there were 15,853 patients in the original dataset who initiated PAP therapy for SDB. Of these, 967 patients were excluded because they had a PAP machine dispensed from both the ISP and a traditional DME company over this time period, leaving 7,448 participants in the ISP group and 6,471 participants in the DME group. After excluding patients who were started on PAP therapy outside of the target date range, 2,257 participants remained in the ISP group while 3,009 participants remained in the DME group. Patients without demographic data (235 excluded from ISP, 240 excluded from DME) and outside of the age interval of 18–64 years (189 excluded from ISP, 714 excluded from DME) were excluded. Of the 2,055 patients remaining in the DME group, 2,001 patients chose 1 of the 4 participating DME companies that agreed to keep the modems on indefinitely. Available characteristics from each of these final groups are shown in Table 8.

2.3.2 Rate of PAP adherence

As stated above, patients were classified as adherent if PAP therapy was used for at least 4 hours on at least 70% of days over 30-day, 90-day, and 1-year intervals. Table 9 shows the adherence rates for each of these groups at these intervals. Of the 2,001 patients included in the DME group, 1,327 patients (66%) were adherent at 30 days, 1,116 patients

Table 8: *Statistically significant at P <.05. BPAP = bilevel positive airway pressure; CPAP = continuous positive airway pressure; DME = durable medical equipment; ISP = integrated sleep practice.

Table 7. Baseline group characteristics.			
Characteristic	**ISP Group**	**DME Group**	**P**
Male, %	64.7	60.1	.003*
Mean age, years	46.7	47.4	.04*
White, %	84.5	80.5	.001*
Black, %	8.1	10.8	.004*
Other race/refused to report, %	7.4	8.6	0.2
Maintained on CPAP, %	92.5	92.8	0.8
Maintained on BPAP, %	6.2	5.6	0.5
*Statistically significant at P <.05. BPAP = bilevel positive airway pressure; CPAP = continuous positive airway pressure; DME = durable medical equipment; ISP = integrated sleep practice.			

(56%) were adherent at 90 days, and 665 patients (33%) were adherent at 1 year. Of the 1,833 patients included in the ISP group, 1,296 patients (71%) were adherent at 30 days, 1,212 patients were adherent at 90 days (66%), and 961 patients (52%) were adherent at 1 year. Chi-square analysis showed that the ISP group had a significantly higher rate of adherence at 30 days (P = .004), 90 days ($P < .00001$), and 1 year ($P < .00001$). Because of the significantly higher rate of Whites and lower rate of Blacks in the ISP group, the differences between experimental groups were also analyzed between these subgroups. The results of this subgroup analysis are displayed in Table 9.

Table 9: Values are n (%) unless otherwise indicated. For each group, patients were classified as being adherent if they achieved at least 4 hours of PAP use on at least 70% of nights over the respective time interval. P values were calculated with chi-square test. *Statistically significant at P <.05. DME = durable medical equipment; ISP = integrated sleep practice; PAP = positive airway pressure.

Number of patients adherent (adherence rate) at key time intervals: ISP vs DME groups.			
	ISP Group	**DME Group**	**P**
White, n	1549	1611	
30 days	1113 (73)	1107 (69)	.006*
90 days	1059 (68)	940 (58)	<.00001*
1 year	841 (54)	577 (36)	<.00001*
Black, n	148	217	
30 days	80 (54)	123 (57)	.62
90 days	70 (47)	99 (46)	.75
1 year	53 (35)	50 (23)	.008*
Total, n	1833	2001	
30 days	1296 (71)	1327 (66)	.004*
90 days	1212 (66)	1116 (56)	<.00001*
1 year	961 (52)	665 (33)	<.00001*
Values are n (%) unless otherwise indicated. For each group, patients were classified as being adherent if they achieved at least 4 hours of PAP use on at least 70% of nights over the respective time interval. P values were calculated with chi-square test. *Statistically significant at P <.05. DME = durable medical equipment; ISP = integrated sleep practice; PAP = positive airway pressure.			

2.3.3 Nightly duration of PAP usage

To determine whether there was a longer duration of PAP usage among patients in the ISP group, each participant's nightly usage was averaged over the respective 30 days, 90 days, and 1 year following PAP initiation. Of the 3,884 patients included in the final analysis, all patients had evidence of at least some PAP usage. The median durations of average nightly usage among the DME group at 30 days, 90 days, and 1 year were 345 minutes, 319 minutes, and 164 minutes, respectively. Median durations of average nightly usage among the ISP group at 30 days, 90 days, and 1 year were 357 minutes, 348 minutes, and 312 minutes, respectively. To test our hypothesis that the ISP group had superior usage duration among its participants, 1-sided Mann-Whitney U test showed significant differences between groups at 30 days ($p = .002$), 90 days ($p < .00001$), and 1 year ($p < .00001$). Comparisons between race subgroups are displayed in Table 10.

Table 10: Each patient's nightly duration of use was averaged over the respective intervals. The usage durations reported in the table describe the median (± SD) of each group's average nightly use, in minutes. P values were calculated with 1-sided Mann-Whitney U test. *Statistically significant at P <.05. DME = durable medical equipment; ISP = integrated sleep practice; PAP = positive airway pressure.

Median average nightly use of individuals still using PAP at key time intervals: ISP vs DME groups.			
	ISP Group	**DME Group**	**P**
White			
30 days	366 (± 123)	353 (± 125)	.004*
90 days	358 (± 132)	327 (± 133)	<.00001*
1 year	322 (± 157)	184 (± 163)	<.00001*
Black			
30 days	311 (± 125)	298 (± 117)	.62
90 days	284 (± 136)	279 (± 128)	.39
1 year	233 (± 150)	122 (± 143)	.001*
Total			
30 days	357 (± 125)	345 (± 126)	.002*
90 days	348 (± 134)	319 (± 134)	<.00001*
1 year	312 (± 157)	164 (± 160)	<.00001*
Each patient's nightly duration of use was averaged over the respective intervals. The usage durations reported in the table describe the median (± SD) of each group's average nightly use, in minutes. P values were calculated with 1-sided Mann-Whitney U test. *Statistically significant at $P < .05$. DME = durable medical equipment; ISP = integrated sleep practice; PAP = positive airway pressure.			

2.4 Discussion

The purpose of this study was to explore differences in PAP adherence between patients receiving PAP therapy from a traditional DME provider and those receiving PAP therapy through an ISP. To evaluate aggregated adherence differences, we first compared the proportion of adherent patients between each group. Our findings suggest superior PAP adherence rates in the ISP group, both in the short term (30 days, 90 days) and the long term (1 year). This difference remained significant when comparing the White subgroups. For the Black subgroup, this difference in adherence rate between ISP and DME groups was significant only at 1 year. In addition to superior rates of patient adherence, the average nightly usage was also found to be superior between experimental groups at all time points. This difference remained significant among all time points in the White subgroup and at 1 year in the Black subgroup. The long-term adherence rate of 52% in the ISP group was comparable to rates published previously [49, 98], while average nightly use of 5.2 hours in the ISP group was slightly better than averages previously published [49, 79].

Some may have noted differences in adherence rate between Black and White subgroups. These differences were found to be significant at all time points (1 month, 3 months, and 1 year) within both the ISP group and the DME group. Although this issue was not addressed in the initial set of hypotheses, we felt this was an important finding that should be added to the current body of literature [98, 79, 80, 23, 92]. There are likely multiple contributing factors through which race contributes to lower adherence rates, including shorter sleep duration [11] and differences in cultural attitudes about sleep [21]. It is also likely a proxy for other important latent variables such as socioeconomic status, which has also been shown to influence PAP adherence [98, 23, 18]. While monitoring these practice metrics in the future, we hope future work can uncover effective strategies for addressing these important mediating factors.

Many clinicians working in sleep medicine have experienced the suboptimal communica-

tion patterns and patient confusion related to fragmented care when dispensing PAP therapy through the heterogeneous landscape of traditional DME providers. As a result, some clinicians may find improved rates of adherence with integrated PAP care to be intuitive. Although interaction between PAP initiation and DME providers has not been well studied, we hypothesize that such integration has the potential to improve communication between the patient and his/her interdisciplinary team, a factor known to be important in promoting adherence to many interventions [40]. Second, an integrated care model fundamentally improves the degree of standardization over the people and processes required to initiate PAP therapy. As such, implementation and monitoring of new processes based on best-available evidence become more feasible. For example, over the period studied, the practice's PAP initiation protocol was provided to participating DME companies, but we were only able to confirm that the protocol was followed by those patients who chose the ISP as a PAP provider. Although some may see this as a limitation in the study design, we believe that such standardization could be an explanatory factor for the demonstrated differences in adherence rate. It is worth noting that, although the PAP initiation protocol was a practical application of evidence-based principles designed to improve attitudes and beliefs about treatment, it did not go through the rigor of formal validation.

Restrictions surrounding age of participant enrollment was another major study constraint. Because our study involved patient referral to the practice's internal DME company, federal laws prohibiting physician self-referral [3] (commonly known as the Stark law) precluded inclusion of Medicare patients into the ISP group. To avoid introducing age as a confounding factor between groups, patients eligible for Medicare on the basis of age were excluded from the study altogether, potentially limiting the generalizability to a typical sleep medicine clinic population.

This study has several other important limitations to keep in mind. First, as is often the case with retrospective cohort studies, there were statistically significant demographic

imbalances between groups that may have biased the results. Also, given the disparate adherence and clinical datasets, we were unable to explore other important confounding factors, such as socioeconomic status, SDB severity, specific SDB diagnosis, and baseline sleepiness. Admittedly, our methods to include only PAP-naive patients were imperfect because patients who previously attempted PAP therapy outside of our region's DME providers (including the ISP) could not be reliably excluded from our final analysis. However, we do not believe that there were systematic differences that would have introduced higher rates of previously PAP-intolerant individuals into either experimental group. Although only a subset of the above confounding factors have consistently proven to influence adherence [95, 60, 47, 66, 98, 79, 80, 23], future work should continue to focus on controlling for these important factors. Finally, along the same lines, a truly patient-centric study would have been able to explore differences in other important outcomes, such as improvements in sleepiness or neurocognitive symptoms. However, in light of previous evidence showing a linear dose–response between the duration of PAP usage and these outcomes [98, 79, 86, 96], the metrics available to us may provide a useful proxy to guide future patient-centered outcome studies.

Despite these limitations, the study explored a pervasive issue, affecting both academic and community-based sleep practices. The interactions between the traditional DME providers and PAP adherence have been underexplored in the peer-reviewed literature. The large difference in long-term (1-year) adherence rates between groups, in particular, warrants further exploration into the role that the heterogeneous landscape of traditional DME providers plays in long-term PAP adherence. As current health care trends drive sleep medicine practices to adopt remote monitoring and chronic care management models, further studies are needed to better understand the tradeoffs of receiving PAP therapy through traditional DME providers.

CHAPTER III.

REM Sleep Classification

3.1 Introduction

Sleep disorders are commonly seen in patients with neurodegenerative diseases. Specifically, (Rapid Eye Movement) REM sleep behavior disorder (RBD) may be the initial manifestation of neurodegenerative diseases such as Parkinson's disease, dementia, and multiple system atrophy.

Elements of the R&K standard that were defined because of the limitations of technology of the time, like the use of 30 second epochs, still exist in the AASM standard. At the time and to this day, there is a need to find a balance between interpretability (or explainability) and complexity. The more complex the data and analysis the more difficult it is to interpret and explain the results. Consequently, manually scored recordings are divided into 30 second epochs and scored with relatively few clinical markers (sleep spindles, k-complexes, sawtooth waves, etc) while automatically scored recordings must be reviewed manually. In manually scored recordings, the clinical markers and epochs can then be aggregated to define a complex and interpretable pathology requiring a trained technician and time-consuming methodologies. In automatically scored recordings, features defined by statisticians and data scientist in effective algorithms can be too abstract and removed from previously known clinical markers to carry meaning for clinical experts. The goal of this study is to find relatively few clinical markers or features used by machine learning that can later be aggregated to create an interpretable pathology of increased complexity.

The authors in [54] worked to make any sleep categorization algorithm more interpretable via pretraining on clinical markers [54] which focuses directly on explainabilty. They argue that algorithms trained on raw data learn more on noise than on actionable information which is a common problem with the interpretability of CNNs. Their algorithm starts with predefined clincial markers. The approach proposed in this paper differs by setting the

filter size for each convolutional neural network to be representative of 5 frequency bands explained in Table 11 and letting the algorithm find the clinical markers.

Table 11: Each convolutional network represents a different frequency band. The delta, theta, and alpha frequency bands are gathered from common definitions. The Beta frequency band is divided into 2 bands because it is the largest frequency band and noise in a signal can be found in the highest frequencies. A midpoint value in the frequency band is chosen to be the CNN Filter Frequency. The CNN Filter Size is calculated using the CNN Filter Frequency and the sample rate of the data set. $CNNFilterSize \approx (samplerate)/(CNNFilter)$

EEG Frequency Band to CNN Filter		
EEG Frequency Band (Hz)	**CNN Filter Frequency (Hz)**	**CNN Filter Size (samples)**
Delta (0-3.99)	2.8	35
Theta (4-7.99)	5.2	19
Alpha (8-13)	10	10
Beta1 (>13)	20	5
Beta2 (>13)	50	2

This approach is the first step in a series to accomplish multiple long-term goals. The first-term goal is to create a model that is accepted by the AASM to score sleep without the need for physician's review. By creating a CNN model with fewer filters based on EEG frequency bands, the filters and individual CNN sub models become relatable to clinical markers. This paper focuses on the first step in that process: reducing the number filters per EEG frequency band to a number that can be easily reviewed and understood. The second long-term goal is to create a clinical marker basis for machine intelligence and ingenuity. By allowing the algorithm to learn filters on raw data, a process can be created to learn filters that are few enough to be interpreted by a human and mathematically relevant to the machine learning process. The learned filters can then be used to by various algorithms as a basis of diagnosis and investigation.

3.2 Methods

The Sleep-EDF Database [6] is a commonly used sleep staging dataset. The dataset contains 197 whole-night PSG sleep recordings, containing EEG, EOG, chin EMG, and

event markers. The sampling rate is 100 Hz. From this dataset, up to 12 nights of EEG recordings from Fpz-Cz channel are used. To ensure the cleanest data set, the final weights were trained using only 2 nights with a highest number of REM epochs were used. The 30 second epochs are scored using the R&K standard and "not scored" epochs. The movement time and not scored epochs are removed because they are no longer in use according to the AASM. All the wake stages in the beginning of each night are removed except the last one because there could be hours of wake stages before sleep started. Because there were not hours of wake stages to end the study, there are not wake stages removed at the end. The stages were then labeled as REM or not REM.

The model used is a variation on the most common models, the CNN to RNN models [29, 93, 88, 87, 28]. The variation is a limit on the number of filters and a filter that matches a predefined frequency band explained below. The model in Figure 23 consists of a convolution part and recurrent or memory part. The model implements python 3.6.8, Tensorflow 1.14.0, and Tensorflow Keras 2.2.4.

The convolutional part consist of 5 CNN's. The GACNN model [37] found the most information can be extracted from an EEG signal using 4 different filter sizes for each CNN. This model adds an additional CNN with a different size filter. There is 1 CNN for each of the delta, theta, and alpha EEG bands and 2 CNNs for the beta EEG band for a total of 5 CNNs. Each CNN has a filter size that allows the frequency of the filter to fit within the specified frequency band as described in Table 11. The beta EEG band spans the largest frequency and is at the boundary of the sampling rate which can lead to noise in the signal. Splitting the beta EEG into 2 CNN filter sizes is an attempt to separate out the noise and extract more interpretable information.

Each CNN was trained independently on the 30 second epochs to score REM and not REM. Each CNN is associated with a different frequency range in an EEG described in Table 11. Each CNN has either 5 or 10 filters. The pooling layers pool size is on the left

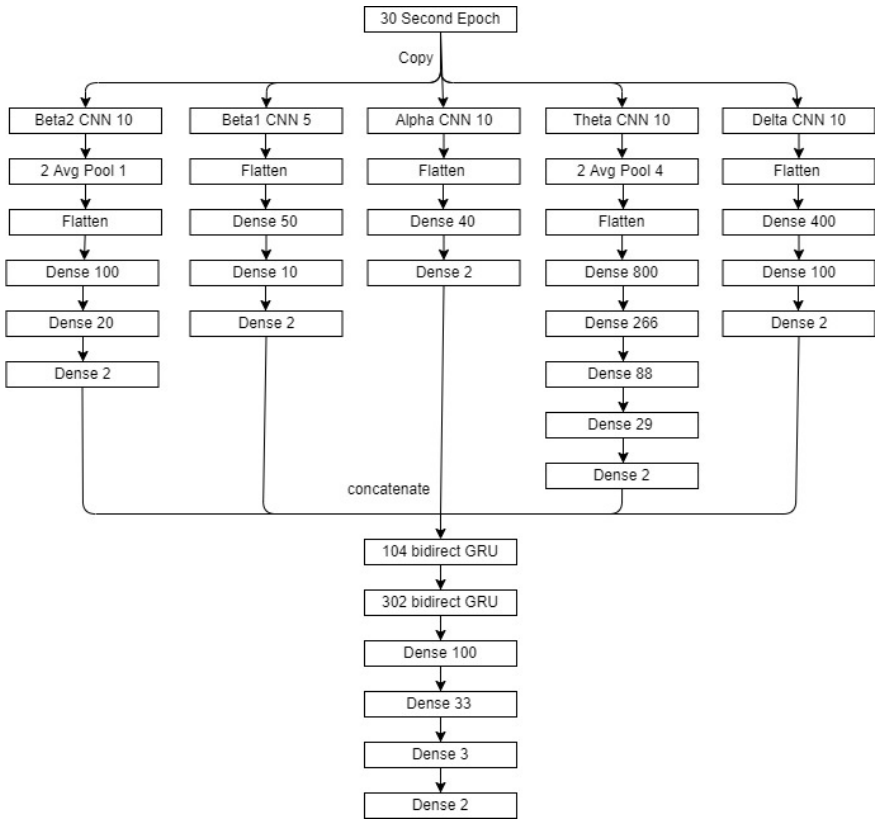


Figure 23: The figure labels the model input, the 5 convolutional networks applied to each input, and the recurrent network applied to 3 time steps of the input. The input is a 30 second epoch. The 30 second epoch is copied to provide input to each of the 5 convolutional networks. The first box of each convolutional network names the EEG band it represents and labels the number of filters. Each box following represents different information about the layer in the network it represents. Boxes about pooling layers define the scaling factor, pooling method, and stride. Boxes about dense layers and recurrent layers define the output size. The concatenate is meant to identify the concatenation of the output of all convolutional networks as well as the concatenation of 3 input time steps.

and the factor by which to downscale for strides is on the right in Figure 23. In each CNN, the same activation, weight initializer, and bias initializer are used on each layer except the final layer which always uses softmax. The activation function is always tanh. The weight initializer and bias initializer is described in Table 11.

The output of the 5 CNNs are concatenated without any adjustment. A new CNN data set is created with each record consisting of 3 epochs of the concatenated output of the CNNs. Each record is labeled with the middle or current epoch. This duplicates the pre and post epochs used to score the current epoch by technicians.

The RNN has 2 layers of bidirectional Gated Recurrent Units (GRU) cells, 3 fully connected layers, and the final softmax layer. The input to the first GRU is the new CNN data set. As can be seen in Figure 23, the first GRU layer outputs 104 dimensions and the second 302 dimensions. All the other parameters were set to the default Keras parameters. No optimization was required.

A random search algorithm for preprocessing and hyperparameter optimization searched over 100,000 different configurations including GRU, Long Short Term Memory (LSTM), bidirectional, and time distributed RNN models. The model presented here is the model with the highest F1-score. While there were very few CNN models that worked, there were many recurrent networks with high accuracy.

The algorithm consists of a signal processing, layers, and optimizer portions. The signal processing portion applied frequency filters. The final model did not use any frequency filters. The layers and optimizer used JSON structures to store and build models. The layers available to search included 1-D convolution, flattening, fully connected, time distributed, max pooling, average pooling, long short term memory, GRU, and bidirectional. The optimizers available include stochastic gradient descent, Adam, Nadam, and root mean square propagation.

Model selection training and final training were completed with different configurations

to reduce model selection time and increase final training accuracy. For both, each CNN and the RNN use 5 nights of the original data set for training data. They both use 10% of the training data as validation data. For the CNNs, they both normalized the data with the standard deviation and mean across the entire training set.

Model selection training included the number of nights as a hyperparameter and batch size as function of epochs in those nights. The number of nights included in training was restricted to leave at least 1 night for testing from a patient not included in the training set. The batch size was calculated using the count of epochs in the training data divided by the cross validation folds. There are always 10 cross validation folds.

Final training removed the cross validation and set the batch size to 32.

3.3 Results

An algorithm or tool that identifies a deep learning model for signal processing, a deep learning model for categorizing REM sleep, and measurements for the deep learning model are the three outcomes of this study.

The algorithm or tool can and has been reused for many different signal processing categorization projects. Even though the current version lacks a user interface, it serves as a proof of concept. With just a few parameters, this program can do a random search through more than a trillion distinct configurations for deep learning models for signal processing. It also has the ability to apply Fourier analysis as a preprocessing step. While it is a work in progress, this strategy can be part of the learning ability of algorithms.

The deep learning model is described in Figure 23. The input to this model is 5 copies of 3 consecutive 30 second epochs of a RAW EEG recording. The model assigns a label of REM or not REM to the 2nd epoch where the epochs are ordered oldest to newest in time. Each copy of an epoch is input to 1 of 5 CNNs. Each CNN is representative of a specified frequency range. Each CNN model produces a one hot encoded output for REM and not REM. The number for REM and not REM are individually less than 1 and combined add up

to 1. For each epoch, these outputs are concatenated into a list. The concatenated output for all 3 epochs is used as input to the GRU portion of the model. The GRU portion of the model produces a one hot encoded output for REM and not REM. The output label is REM when the one hot encoded REM output is greater than or equal to 0.5. Otherwise, the output label is not REM.

Table 12 shows the number of true and false predictions for each part of the model. The Delta CNN model had the most true REM. The Delta CNN model predicted the most actual REM the CNN models. Since eye movement was not removed from the signal, it is likely that eye movements are detected in this frequency. Note that eye movement has been detected as noise in the Fpz-Cz EEG channel. Beta2 predicted the most False REM. This is expected because N1 shares this frequency range with REM. Beta 1 predicted the fewest number of actual REM. Finally, the RNN did the best at reducing False Non REM and increasing REM.

Table 12: The top row is the count of REM and not REM epochs. Each of the following row shows the correct REM, False REM, False not REM, and True not REM predictions for each network in the model. The last row (RNN) takes 3 epochs for each input therefore it has two fewer predictions.

Prediction Results Comparison				
	REM	False REM	False NotREM	NotREM
Actual	1344	0	0	8330
Beta2	1007	585	337	7745
Beta1	5	40	1339	8290
Alpha	781	90	563	8240
Theta	143	34	1201	8296
Delta	1134	195	210	8135
RNN	1212	107	132	8221

Table 13 shows the percent of total epochs for each part of the model. The actual REM is 13.89% of the total epochs. The actual NotREM is 86.11% of the total epochs. The RNN predicted true REM as 12.53% of the total epochs and true NotREM as 85.00% of the the

total epochs. The RNN produced the metrics described in Table 14.

Table 13: This table is intended to show the counts as a percentage of the total number of epochs.

Prediction Results Comparison				
	REM	False REM	False NotREM	NotREM
Actual	13.89%	0.00%	0.00%	86.11%
Beta2	10.41%	6.05%	3.48%	80.06%
Beta1	0.05%	0.41%	13.84%	85.69%
Alpha	8.07%	0.93%	5.82%	85.18%
Theta	1.48%	0.35%	12.41%	85.76%
Delta	11.72%	2.02%	2.17%	84.09%
RNN	12.53%	1.11%	1.36%	85.00%

Table 14 shows the comparison to other models. The models used for comparison use the exact same dataset. These models attempted to score all stages of sleep, and while this could be the end goal, focusing on one stage of sleep can help the model be more explainable. In addition, other stages of sleep can have very specific clinical markers which can be easily identified with simpler methods.

Table 14: Each model represented here posted calculated results using the same type of output and the same dataset. While some models did look for all sleep stages and used additional data sets, this table only shows comparisons of REM vs not REM on the Sleep-EDF Database Expanded [37, 87, 63].

Model Comparisons					
Model	Accuracy	Precision	Recall	F1-Score	Kappa
GACNN	91.00	92.00	91.00	91.00	90.00
DeepSleepNet		80.90	83.90	82.40	
SleepEEGNet		81.63	88.71	85.02	
This Model	97.30	93.08	89.32	90.86	90.00

The GACNN model has a higher Recall and F1-Score 22. The GACNN model is very similar, combining four CNN models into an RNN. To specify the difference, this model contains much fewer CNN layers in each CNN model and much fewer filters in each CNN

layer. This concept of using fewer filters is intended to help in the explainability of the model.

3.4 Discussion

The objective of this research is to enable a neural network to produce findings that can be explained. There are only 45 filters to interpret, which is easier than explaining the many layers of fully connected networks. It is demonstrated that fewer filters may be employed to stage sleep because the model produces results that are comparable to those of other high accuracy models..

The results confirm other aspects of the definition of REM. Rapid eye movement can be seen as noise in the Fpz-Cz channel with extremely low frequency. Since that noise was not removed, the Delta CNN was able to detect it producing the highest accuracy results. In manual scoring, REM stage typically has the highest frequency and Alpha frequency sometimes. This correlates with Beta2 and Alpha CNN which both have high accuracy results. Beta2 results are higher than Alpha. The CNN filter between Alpha and Beta2 has very low accuracy. This frequency range can be attributed to the idea that many sleep stages can be represented in this range.

There are numerous ways to proceed with this subject. For instance, attention layers can be used to enhance this model. This might make the important filters easier to understand. The frequency ranges could be further divided for good reason. Beta can be separated specifically to narrow down the significant frequency range. It may be necessary to filter the initial signal so that only certain frequency ranges exhibit the p value for each CNN filter. Finally, other algorithms could be able to use CNN filters. These filters could be employed in a reinforcement learner to comprehend brain wave sequences if the algorithms can only recognize a small number of filters.

The final model reported in this paper is 1 of many that could have been used. The GRU is used instead of LSTM to reduce the complexity of the model in addition to it's improved

results. While complexity is a major concern, complexity limitations were not considered in place of producing a model that has learned clinical markers. The clinical markers in this model can be analyzed and compared to clinical markers used today, but they were not similar upon a visual inspection. Future versions of this model can be produced with the automated tool to focus on clincial markers for more specific scenarios.

The algorithm or tool can be considered part of the model. Since models must be improved and those improvements are strongly influenced by the input parameters, there could be many models for various disease state categorizations. This is one reason to focus on 1 sleep stage at a time. Each sleep stage could have it's own set of models for various disease states. Since degenerative brain disease is related to REM sleep and REM sleep disorder, creating one model to categorize REM sleep and one for REM sleep disorder would result in comparable clinical markers defined by the various CNNs. Since these clinical markers can be explained in the same way as sawtooth waves, it creates a point of discussion between clincians and statisticians that was not there before.

There are many additions to the tool or algorithm that are in process. This includes the ability to speed up training by using transfer learning across similar models, the addition of attention layers, and the addition of wavelet decomposition as an optional preprocessing step. This will eventually be a tool that builds models from scratch. This differentiates other deep learning automations that start with several pre-existing models.

CHAPTER IV.

Wavelet Deep Learner

4.1 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 [22]. The electrode positions are well-defined by the International 10/20 system [22] [2] [1]. The EEG channels that are best to use were standardized originally by Rechtscheffen and Kales [71] 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) [22]. Elements from all 3 are used to identify sleep stages [22]. 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 [22]. 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 [83, 38, 70, 50, 81] or a single channel EEG [90, 36, 31, 102]. Ebrahimi et al. achieved 93% accuracy using the wavelet decomposition of a single channel EEG and a neural network [36]. Kurt et al. achieved 97-98% accuracy using the wavelet decomposition of EEG, EOG, and chin-EMG [50].

In order to automate sleep staging with an at home study, wavelet decomposition of a single channel EEG will be used as a preprocessing step to the previously built CNN to GRU neural network. The previous model will need to be optimized. Additional layers such as attention layers can be added. Alvarez et al. [102] 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. [57] 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. [57] 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 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 will use the DWT to decompose the EEG signals. Since Daubechies wavelet order 2 produced high accuracy results in multiple papers [31, 36], this will use Daubechies wavelet order 2. This study used the multilevel discrete wavelet

transform.

4.2 Methods

4.2.1 Data

This study used the ”Sleep Recording and Hypnograms in European Data Format (EDF)” data set or ”The Sleep-EDF Database [Expanded]” from physionet.org [6]. 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.

Each patient had two relevant files: a polysomnogram (PSG) recording in EDF format and a hypnogram of annotations in EDF+. EDF is a standard format for exchanging EEG recordings [4]. EDF+ has all the capabilities of EDF and the ability to contain annotations [4]. 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. All the awake epochs except the last one are removed prior to the first epoch that is not a wake epcoch. The count of each annotation per

file is printed to standard out.

4.2.2 Wavelet Decomposition

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 15.

Table 15: 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

During the decomposition process, the wavelet functions are convolved with the signal at different scales and positions, resulting in a set of coefficients that represent the contribution of each wavelet function to the signal at each scale and position. The coefficients at each scale represent a different frequency band of the signal. After each convolution, the resulting coefficients are down sampled, which means that the number of samples is reduced by a factor of two by discarding redundant samples. The sampling rate after each level of decomposition can be calculated using the following formula where the original signal has a sample rate of f_s and n is the level of decomposition

$$f_{sn} = \frac{f_s}{2^n}.$$

The resulting sample rates are matched to EEG frequencies used in the previous defined model in Table 16.

Table 16: DWT frequency bands and sample rates

Frequency Band	**DWT Level**	**Sample Rate (Hz)**
Beta2 (>13 Hz)	1	50.0
Beta1 (>13 Hz)	2	25.0
Alpha (8-13 Hz)	3	12.5
Theta (4-7.99 Hz)	4	6.25
Delta (0-3.99 Hz)	5	3.125

The reconstructed signal after each of five DWT levels on a sample 30 second epoch is visualized in Figure 24.

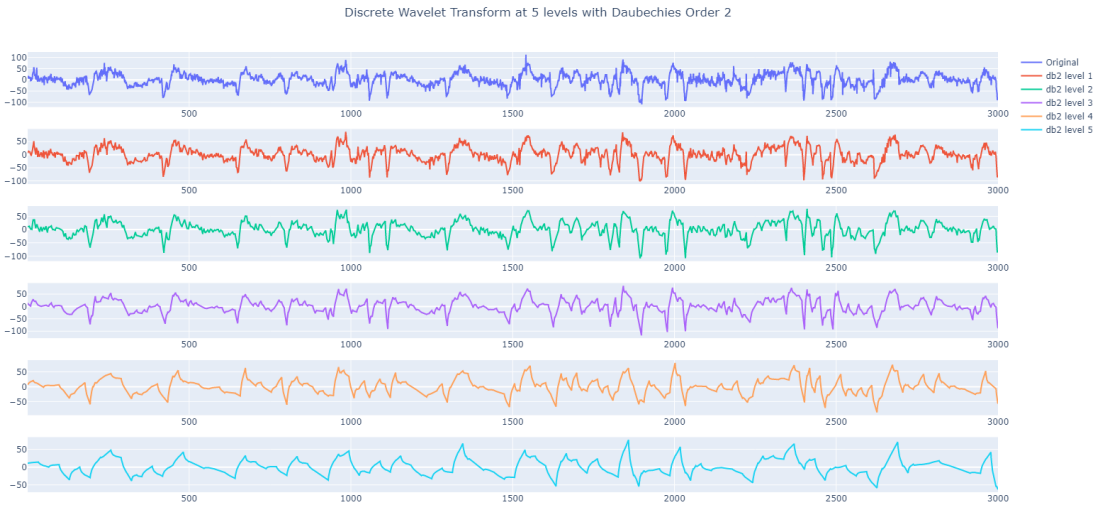


Figure 24: Each plot shows a sample 30 second epoch at a different of level of wavelet decomposition using Daubechies order 2 mother wavelet. The top plot is the original signal. Each 1 down after that goes to a higher level of decomposition starting at level 1 and going to level 5 respectively.

4.2.3 Model

Each of these reconstructed signals is used as input to a CNN. The hyperparameters of the neural network are tested using a custom hyperparameter random search. The random search is divided into 4 steps: preprocessing/training parameters, the candidate builder, the model builder, and the model trainer/tester. These are managed with the random search python script.

A. Preprocessing/Training Parameters

The random search starts building a model by defining a few parameters based on the preprocessing of data and training of the model. The preprocessing parameters include the number and set of input files, the normalization standard deviation and mean, the mother wavelet, the decomposition level, and the expected frequency band. The training parameters include the number of cross validation folds, the validation split percentage, the max number of training epochs, and the batch size. The following is a sample of the preprocessing and training parameters.

dataFiles: input152.csv,input042.csv

cvFolds: 10

validation_split: 0.1

epoch: 100

batchSize: 381

normSTD: 19.476044983423417

normMean: 0.017218097667286797

mother wavelet: db2

decomp level: 5

frequency band: delta

The data files are preprocessed into 30 second epochs with labels. All the movement time are removed. All wake stages except the last one before the first stage that is not wake. The files use the following naming format with the patient number in 2 digits (PP) and the night number in 1 digit (n)

$$inputPPn.csv.$$

The other preprocessing arguments are the normalization standard deviation (normSTD), the normalization mean (normMean), the mother wavelet used in multilevel discrete wavelet transform (mother wavelet), the decomposition level used in multilevel discrete wavelet transform (decomp level), and the frequency band being represented by this CNN. The normalization process is calculated across the entire dataset in the dataFiles parameter. The mother wavelet is manually chosen.

The training arguments include the batch size of each epoch (batchSize), the number maximum number of epochs for training (epoch), the number of cross validation folds (cvFolds), and the validation split percentage in decimal (validation_split). The batch size is calculated to be the entire training data set given the record count in dataFiles (R), the number of cross validation folds (C), and validation split (V) with the following formula. The best epoch size is found through testing. The validation split and cross validation folds are chosen as a baseline for showing generalization

$$batchSize = int((R*(1-V))/C+1).$$

B. The Candidate Builder

The candidate builder creates a json structure with all the necessary information to build a model. Each CNN candidate is expected to have a CNN layer followed by a collection of dense layers. Hyperparameters specific to the CNN include the number of filters (defaulted

to 5 or 10) and the kernal size (defaulted to 3, 5, 10, 20, or 66). The number of dense layers and the number of nodes in each layer is a hyperparameter. The activation function, kernel initializers, bias initializers, and optimizers are also all hyperparameters. The following is an example candidate in json.

```
[
  {
    "layer": "conv1d",
    "no_filters": 5,
    "kernal_size": 19,
    "padding": "same",
    "activation": "tanh",
    "kernel_initializer": "he_normal",
    "bias_initializer": "random_normal"
  },
  {
    "layer": "dense",
    "output": 5,
    "activation": "tanh",
    "kernel_initializer": "he_normal",
    "bias_initializer": "random_normal"
  },
  {
    "layer": "dense",
    "output": 2,
    "activation": "softmax",
    "kernel_initializer": "he_normal",
```

```
    "bias_initializer": "random_normal"
  },
  {
    "layer": "compile",
    "optimizer": "adam",
    "loss": "categorical_crossentropy",
    "metrics": [
      "acc"
    ]
  }
]
```

C. The Model Builder

The model builder reads the json structure and creates a Tensorflow Keras model. A sample CNN model is shown in Figure 25.

D. The Model Trainer/Tester

The model trainer/tester trains the model, tests the model, and saves all relevant data. The stopping condition is always based on the validation loss. There is an option to explore the models with a tensor board. Each cross validation training and test set is created using a the Scikit-Learn StratifiedKFold function. Each training set is trained using the batch size, validation percentage, and maximum number of epochs parameters. This training is allowed to save a large collection of metrics including all the fit history, a plot of the loss versus validation loss, and a plot of the area under the ROC curve versus validation area under the ROC curve. The fit history can include accuracy, precision, recall, f1 score, and area under the ROC curve. Each test set is then used to get a collection of metrics that is always saved. The test set collection of metrics inlcude true positive, false negative, false positive, true negative, true positive, accuracy, sensitivity, specificity, recall, precision, f1 score, and the

```
Model: "model"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 3000, 1)]         0
_________________________________________________________________
conv1d1 (Conv1D)             (None, 3000, 5)           100
_________________________________________________________________
flatten (Flatten)            (None, 15000)             0
_________________________________________________________________
dense1 (Dense)               (None, 10)                150010
_________________________________________________________________
dense2 (Dense)               (None, 5)                 55
_________________________________________________________________
dense3 (Dense)               (None, 2)                 12
=================================================================
Total params: 150,177
Trainable params: 150,177
Non-trainable params: 0
```

Figure 25: The output of model.summary() from Tensorflow Keras of a model built by reading a candidate builder json structure.

last loss calculated in the fit history.

4.2.4 Filter Comparisons: The Frequency Band Wavelet Model

Once all the models are trained, the CNN filters are extracted and examined for interpretability. Each frequency band representative model trained on the data decomposed by each mother wavelet (the frequency band wavelet model) produces a different set of filters on each cross validation step. The filters extracted are the filters from the first cross validation step with the highest f1 score for each frequency band wavelet model. Two methods of comparison are used: visual comparison and spearman rank correlation.

The visual comparison is intended to see if there are any relationship to known clincial markers in provided by the AASM. The clinical markers include but are not limited to k-complexes, sleep spindles, and slow wave activity. K-complexes are sharp negative EEG waves that occur during N2 sleep in response to external stimuli. The presence of K-complexes can be used as a marker of sleep depth and stability. Sleep spindles are brief bursts of EEG activity in the frequency range of 11-16 Hz that occur during N2 sleep. The presence and density of sleep spindles can be used as a marker of sleep quality. Slow wave activity is EEG activity in the frequency range of 0.5-4 Hz that occurs during N3 sleep. The amount of SWA is an important marker of sleep quality, with higher levels of SWA indicating deeper and more restorative sleep.

The spearman rank correlation is used as a first step numerical analysis of filters. The assumption is that a filter can be found by every wavelet model is a high priority. The Spearman's rank correlation is calculated for the filters from each extracted frequency band wavelet model against each of the other extracted filters for each wavelet models associated with the same frequency band. For example, the theta coiflet 4 model filters are compared against all the other models representing the theta frequency band. There are 5 or 10 filters for each model. An example single comparison would be the 1st filter of the theta coiflet 4 model against the 10th filter of the theta haar model.

The Spearman's rank coefficient is then used to find filters that are found by other training. A formula is used to find the filters with the most and highest correlations. The formula uses the 95th percentile of all the Spearman's rank coefficients as a threshold, the filter threshold. The top 5 filters with the most Spearman's rank coefficients greater than the filter threshold are found to be the most relevant filters.

4.3 Results

There are tens of gigabytes of results that are available for review in a result set. Things such as the exact model definition, the training parameters, and the results of each training can be found in the result set. The following is a collection of sample and high priority results. The wavelet model with the highest f1 score is coiflet 4. The examples are primary based on that. Note that the full model is described in greater detail in a previous study so it is not described here. The results described below include: training parameters, the fit history with validation of a single training, the cross validation results of a single training, the average cross validation results for all wavelet models, a plot of filters from the coiflet 4 model and none model, the plot of of the top 5 correlated filters from each frequency band, and a description of those comparisons.

4.3.1 Training Parameters

The first part of the model completed through the random search is the model for the delta frequency band. The preprocessing and training arguments are:

dataFiles: input152.csv,input042.csv,input171.csv,input161.csv,
input091.csv,input002.csv,input142.csv,input031.csv,
input082.csv,input151.csv,input101.csv,input032.csv

cvFolds: 10

validation_split: 0.1

epoch: 100

batchSize: 2054

normSTD: 20.366845241085922

normMean: -0.4573919127663568

mother wavelet: coif4

decomp level: 5

frequency band: delta

4.3.2 Fit History - Data

The training then creates the following ”fit history”.

Table 17: Fit History

loss	accuracy	precision	recall	f1_score	auc_roc
0.0018	1	1	1	1	1
0.0012	1	1	1	1	1
0.0008	1	1	1	1	1
0.0007	1	1	1	1	1
0.0006	1	1	1	1	1
0.0005	1	1	1	1	1
0.0005	1	1	1	1	1

Table 18: Validation Fit History

val_loss	val_accuracy	val_precision	val_recall	val_f1_score	val_auc_roc
0.1018	0.9606	0.9606	0.9606	0.9606	0.9945
0.0978	0.9685	0.9685	0.9685	0.9685	0.9947
0.0797	0.9790	0.9790	0.9790	0.9790	0.9967
0.0773	0.9790	0.9790	0.9790	0.9790	0.9967
0.0812	0.9764	0.9764	0.9764	0.9764	0.9963
0.0844	0.9738	0.9738	0.9738	0.9738	0.9960
0.0843	0.9738	0.9738	0.9738	0.9738	0.9960

The fit history for the train data shows that the model prediction has 100 percent on all metrics while the loss starts low and steadily decreases. The fit history for the validation start of with all metrics in the 96% range and moving up while the loss starts low and steadily decreases. This suggest the model is accurate with the ability to generalize.

4.3.3 Fit History - Plots

This fit history is then plotted in two different plots: loss vs validation loss and auc_roc vs validation auc_roc. See Figure 26.

This visualization of the fit history is a simpler way to confirm this model is a good fit for this data.

4.3.4 Cross Validation

After the training, the cross validation test data is used to create a different set of metrics. This cross validation data is aggregated accross all cross validation folds in Table 19.

Table 19 identifies the true positives, false positive, false negatives, and true negatives. Through 6 of the 10 cross validation fold there are 0 false positives and false negatives. Table 19 identifies that these 6 folds are 100% accurate and the average accuracy is 97%. Despite that the f1 score gets as low as 11% and the average f1 score is 85%. This is not

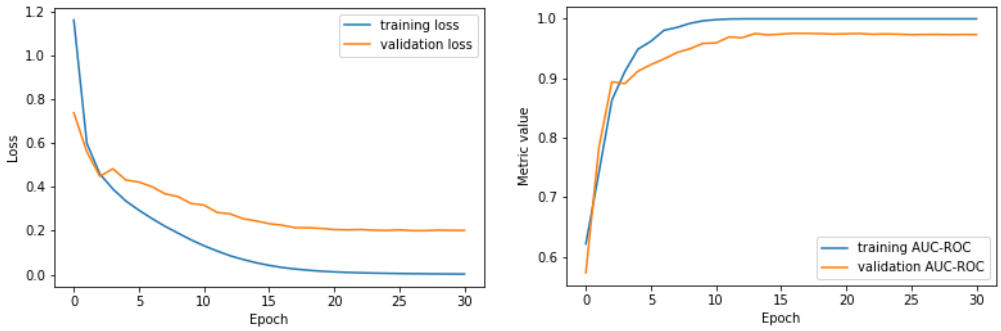


Figure 26: The loss and validation loss of the first cross validation fold with data that has been preprocessed with the multilevel discrete wavelet transform level 5 with mother wavelet Daubechies order 2. The AUC ROC and validation AUC ROC of the first cross validation fold with data that has been preprocessed with the multilevel discrete wavelet transform level 5 with mother wavelet Daubechies order 2.

as good as other models and the other parts of this complete model need to be added to improve the model.

4.3.5 Average Cross Validation

The same model is tested with wavelet decomposition using 28 different mother wavelets and 1 without any wavelet decomposition. All the results produced 98% accuracy or higher accuracy with an f1 score between 93% and 94%. Table **??** shows the average cross validation results from all 22 mother wavelets. The coif4 mother wavelet has the highest f1 score with 0.9456. The db5 mother wavelet had the lowest f1 score with 0.9374. The model without any wavelet decomposition is closer to the lowest f1 score with 0.9392 but notably not the lowest.

Table 20: The average with confidence interval cross validation results of the model trained with 29 different data sets. 28 of the data sets were preprocessed by 28 different mother wavelets. 1 does not have any preprocessing. These are listed from highest f1score to lowest f1score with the one without preprocessing highlighted in red.

Wavelet	Recall	Precision	F1 Score
coif4	0.991 (0.9711, 1.0109)	0.9996 (0.9987, 1.0005)	0.9456 (0.8253, 1.0659)

sym3	0.9911 (0.9715, 1.0107)	0.9999 (0.9997, 1.0001)	0.9451 (0.8237, 1.0665)
db4	0.9911 (0.9715, 1.0107)	0.9999 (0.9997, 1.0001)	0.9448 (0.8232, 1.0664)
haar	0.9911 (0.9715, 1.0107)	1 (1, 1)	0.9445 (0.8215, 1.0675)
coif5	0.991 (0.9711, 1.0109)	1 (1, 1)	0.9441 (0.8202, 1.068)
db6	0.99 (0.9708, 1.0092)	0.9981 (0.9959, 1.0003)	0.941 (0.8241, 1.0579)
sym7	0.9902 (0.9705, 1.0099)	0.9993 (0.9979, 1.0007)	0.941 (0.8177, 1.0643)
sym4	0.9897 (0.9708, 1.0086)	0.9981 (0.9957, 1.0005)	0.9407 (0.8248, 1.0566)
sym8	0.99 (0.9708, 1.0092)	0.9985 (0.9969, 1.0001)	0.9407 (0.8212, 1.0602)
db7	0.9901 (0.9704, 1.0098)	0.999 (0.9974, 1.0006)	0.9401 (0.817, 1.0632)
db1	0.9896 (0.97, 1.0092)	0.998 (0.9958, 1.0002)	0.9397 (0.8201, 1.0593)
sym6	0.9898 (0.9704, 1.0092)	0.9984 (0.9965, 1.0003)	0.9397 (0.8191, 1.0603)
db2	0.9899 (0.9705, 1.0093)	0.9985 (0.9967, 1.0003)	0.9397 (0.8187, 1.0607)
sym2	0.9899 (0.9704, 1.0094)	0.9986 (0.9964, 1.0008)	0.9397 (0.8176, 1.0618)
coif2	0.9901 (0.9704, 1.0098)	0.9988 (0.9972, 1.0004)	0.9397 (0.8167, 1.0627)
sym5	0.9898 (0.9699, 1.0097)	0.9984 (0.9968, 1)	0.9394 (0.8165, 1.0623)
None	0.9894 (0.9691, 1.0097)	0.9977 (0.9954, 1)	0.9392 (0.8173, 1.0611)
coif1	0.9898 (0.9699, 1.0097)	0.9985 (0.9969, 1.0001)	0.9391 (0.8155, 1.0627)
gaus1	0.9892 (0.9691, 1.0093)	0.9976 (0.995, 1.0002)	0.939 (0.8188, 1.0592)
db8	0.9896 (0.9698, 1.0094)	0.9983 (0.9967, 0.9999)	0.9388 (0.8155, 1.0621)
db3	0.9895 (0.9696, 1.0094)	0.9983 (0.9963, 1.0003)	0.9387 (0.8163, 1.0611)
gaus3	0.9892 (0.9697, 1.0087)	0.9976 (0.9952, 1)	0.938 (0.8193, 1.0567)
mexh	0.989 (0.9685, 1.0095)	0.9973 (0.9945, 1.0001)	0.938 (0.8165, 1.0595)
dmey	0.9895 (0.9696, 1.0094)	0.9981 (0.9959, 1.0003)	0.938 (0.815, 1.061)
gaus2	0.9891 (0.9683, 1.0099)	0.9976 (0.9948, 1.0004)	0.938 (0.8148, 1.0612)
gaus4	0.9891 (0.9686, 1.0096)	0.9975 (0.995, 1)	0.9378 (0.8149, 1.0607)
coif3	0.9891 (0.9696, 1.0086)	0.998 (0.9956, 1.0004)	0.9377 (0.8175, 1.0579)

morl	0.9889 (0.9687, 1.0091)	0.9976 (0.995, 1.0002)	0.9376 (0.8165, 1.0587)
db5	0.9892 (0.9696, 1.0088)	0.9979 (0.9955, 1.0003)	0.9374 (0.8168, 1.058)

Table 21: The average with confidence interval cross validation results of the model trained with 29 different data sets. 28 of the data sets were preprocessed by 28 different mother wavelets. 1 does not have any preprocessing. These are listed from highest f1score to lowest f1score with the one without preprocessing highlighted in red.

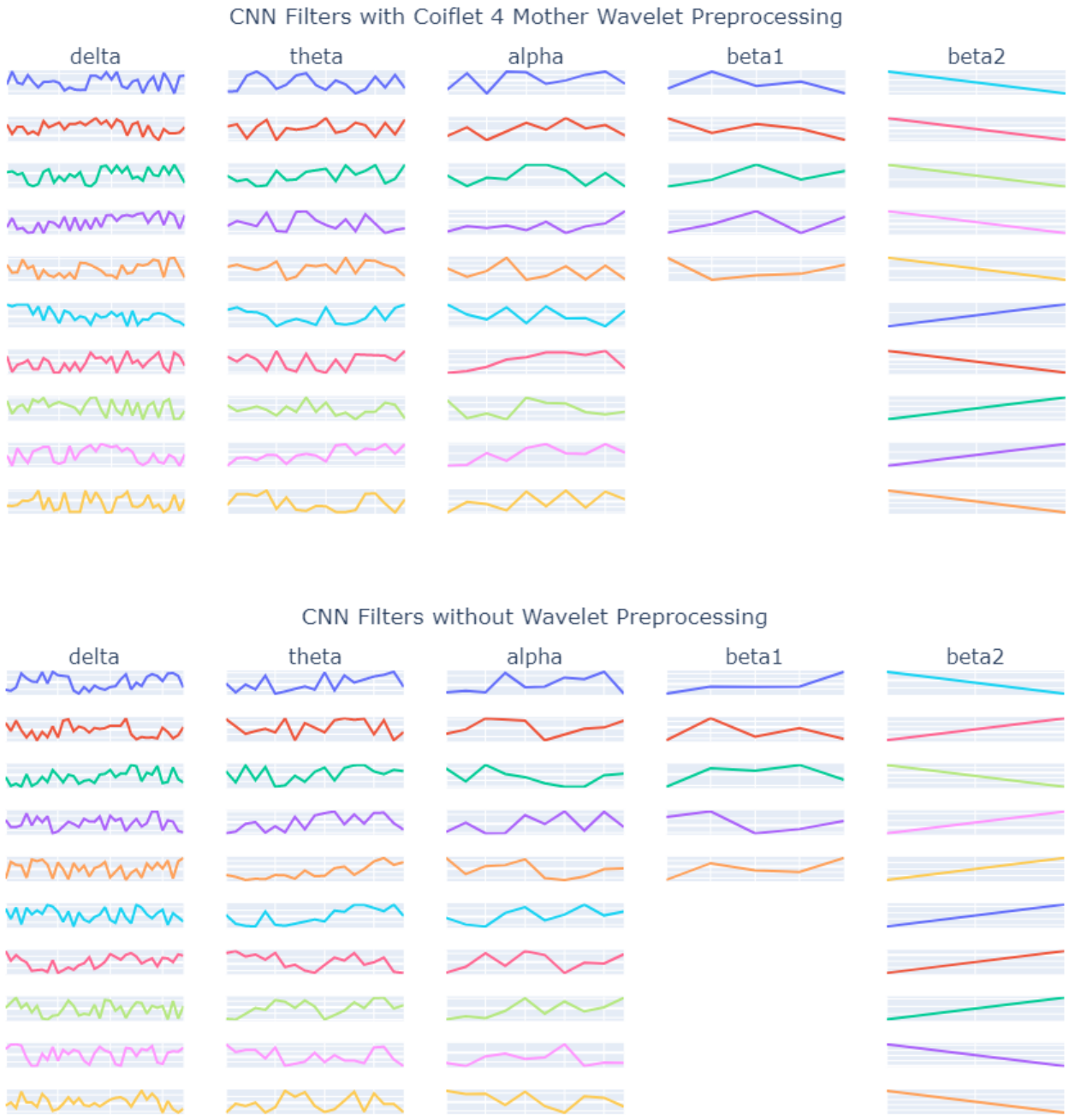
Wavelet	Acc	Spec	F1 Score
coif4	0.991 (0.9711, 1.0109)	0.9996 (0.9987, 1.0005)	0.9456 (0.8253, 1.0659)
sym3	0.9911 (0.9715, 1.0107)	0.9999 (0.9997, 1.0001)	0.9451 (0.8237, 1.0665)
db4	0.9911 (0.9715, 1.0107)	0.9999 (0.9997, 1.0001)	0.9448 (0.8232, 1.0664)
haar	0.9911 (0.9715, 1.0107)	1 (1, 1)	0.9445 (0.8215, 1.0675)
coif5	0.991 (0.9711, 1.0109)	1 (1, 1)	0.9441 (0.8202, 1.068)
db6	0.99 (0.9708, 1.0092)	0.9981 (0.9959, 1.0003)	0.941 (0.8241, 1.0579)
sym7	0.9902 (0.9705, 1.0099)	0.9993 (0.9979, 1.0007)	0.941 (0.8177, 1.0643)
sym4	0.9897 (0.9708, 1.0086)	0.9981 (0.9957, 1.0005)	0.9407 (0.8248, 1.0566)
sym8	0.99 (0.9708, 1.0092)	0.9985 (0.9969, 1.0001)	0.9407 (0.8212, 1.0602)
db7	0.9901 (0.9704, 1.0098)	0.999 (0.9974, 1.0006)	0.9401 (0.817, 1.0632)
db1	0.9896 (0.97, 1.0092)	0.998 (0.9958, 1.0002)	0.9397 (0.8201, 1.0593)
sym6	0.9898 (0.9704, 1.0092)	0.9984 (0.9965, 1.0003)	0.9397 (0.8191, 1.0603)
db2	0.9899 (0.9705, 1.0093)	0.9985 (0.9967, 1.0003)	0.9397 (0.8187, 1.0607)
sym2	0.9899 (0.9704, 1.0094)	0.9986 (0.9964, 1.0008)	0.9397 (0.8176, 1.0618)
coif2	0.9901 (0.9704, 1.0098)	0.9988 (0.9972, 1.0004)	0.9397 (0.8167, 1.0627)
sym5	0.9898 (0.9699, 1.0097)	0.9984 (0.9968, 1)	0.9394 (0.8165, 1.0623)
None	0.9894 (0.9691, 1.0097)	0.9977 (0.9954, 1)	0.9392 (0.8173, 1.0611)
coif1	0.9898 (0.9699, 1.0097)	0.9985 (0.9969, 1.0001)	0.9391 (0.8155, 1.0627)
gaus1	0.9892 (0.9691, 1.0093)	0.9976 (0.995, 1.0002)	0.939 (0.8188, 1.0592)

db8	0.9896 (0.9698, 1.0094)	0.9983 (0.9967, 0.9999)	0.9388 (0.8155, 1.0621)
db3	0.9895 (0.9696, 1.0094)	0.9983 (0.9963, 1.0003)	0.9387 (0.8163, 1.0611)
gaus3	0.9892 (0.9697, 1.0087)	0.9976 (0.9952, 1)	0.938 (0.8193, 1.0567)
mexh	0.989 (0.9685, 1.0095)	0.9973 (0.9945, 1.0001)	0.938 (0.8165, 1.0595)
dmey	0.9895 (0.9696, 1.0094)	0.9981 (0.9959, 1.0003)	0.938 (0.815, 1.061)
gaus2	0.9891 (0.9683, 1.0099)	0.9976 (0.9948, 1.0004)	0.938 (0.8148, 1.0612)
gaus4	0.9891 (0.9686, 1.0096)	0.9975 (0.995, 1)	0.9378 (0.8149, 1.0607)
coif3	0.9891 (0.9696, 1.0086)	0.998 (0.9956, 1.0004)	0.9377 (0.8175, 1.0579)
morl	0.9889 (0.9687, 1.0091)	0.9976 (0.995, 1.0002)	0.9376 (0.8165, 1.0587)
db5	0.9892 (0.9696, 1.0088)	0.9979 (0.9955, 1.0003)	0.9374 (0.8168, 1.058)

4.3.6 CNN Filters

The filters from each frequency band wavelet model are extracted and plotted for visual inspection. The clinical markers such as k-complexes, v-waves, and spindles can be seen in the filters. Slow waves can not be seen as easily. Figure 27 has the filters from coiflet 4 model and the none model (the model trained on data without preprocessing). The CNN filter size makes it necessary to consider each frequency band independently. The CNN filter size for each band is Delta (35), Theta (19), Alpha (10), Beta1 (5), and Beta2 (2).

The Spearman's rank coefficient is then calculated for all wavelet models in the same frequency band. This coefficient has more implications for the delta, theta, and alpha frequency bands because of their filter size. The 95th percentile for the Spearman's rank coefficient is 0.34145658263305323. The percent of coefficients for the top 5 of each frequency band that are higher than the 95th percentile are in Table 22.



.

Figure 27: Filters from coiflet 4 wavelet model and the no wavlet model are plotted. The filters are from the weights created during the cross validation with the highest f1 score. Clinical markers such as k-complexes, v-waves, and spindles can be seen in the filters. Slow waves can not be seen as easily.

Table 19: Cross validation results and the average across those cross validations with data that has been preprocessed with the multilevel discrete wavelet transform level 5 with mother wavelet Daubechies order 2. The results identify the true positives, false positives, true negatives, false negatives, accuracy, specificity, recall, precision, and f1 score of each cross validation test set.

CV Fold	**True REM**	**False REM**	**False NonREM**	**True NonREM**	
1	7	17	275	1983	
2	228	11	54	1989	
3	227	0	55	2000	
4	223	63	59	1937	
5	214	0	67	2000	
6	270	2	12	1997	
7	279	8	3	1991	
8	278	0	4	1999	
9	281	1	1	1998	
10	99	25	183	1974	
Average	210.6	12.7	71.3	1986.8	
CV Fold	**Acc**	**Spec**	**Recall**	**Precision**	**F1 Score**
1	0.872	0.992	0.025	0.292	0.046
2	0.972	0.995	0.809	0.954	0.875
3	0.976	1	0.805	1	0.892
4	0.947	0.969	0.791	0.78	0.785
5	0.971	1	0.762	1	0.865
6	0.994	0.999	0.957	0.993	0.975
7	0.995	0.996	0.989	0.972	0.981
8	0.998	1	0.986	1	0.993
9	0.999	0.999	0.996	0.996	0.996
10	0.909	0.987	0.351	0.798	0.488
Average	0.9633	0.9937	0.7471	0.8785	0.7896

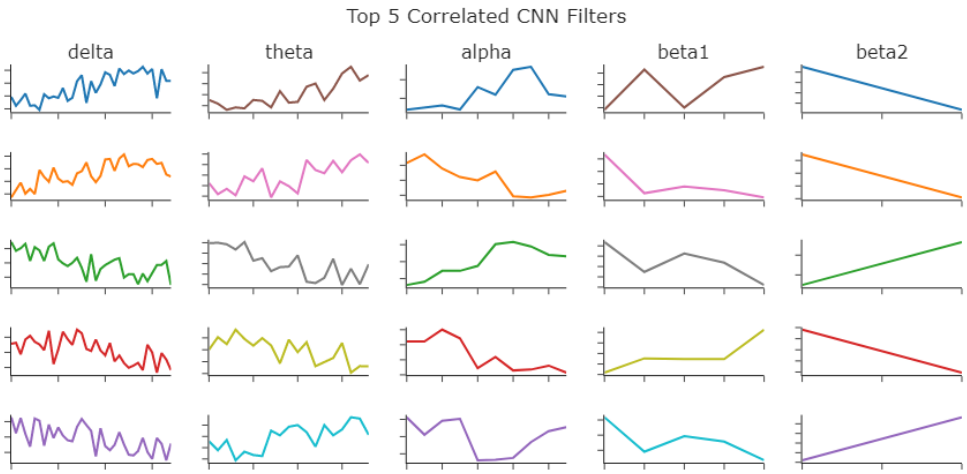
Table 22: The frequency band wavelet models that have the biggest percent of Spearman's coefficients above the 95th percentile of Spearman's coefficients. The 95th percentile of Spearman's coefficients is calculated across all models. Every wavelet model in each frequency band is then ranked by the number of wavelet models it has a higher Spearman's coefficient with than the 95th percentile.

Freq Band	**wavelet**	**95th %**	**% above 95th**
delta	1st db8	0.3415	15.17

delta	9th sym8	0.3415	14.83
delta	3rd db3	0.3415	11.72
delta	6th sym5	0.3415	11.72
delta	4th db7	0.3415	11.38
theta	4th None	0.4684	11.72
theta	9th db8	0.4684	11.03
theta	7th sym6	0.4684	11.03
theta	4th sym2	0.4684	10.69
theta	6th mexh	0.4684	10.00
alpha	6th db4	0.6485	10.00
alpha	5th coif2	0.6485	10.00
alpha	0th sym5	0.6485	10.00
alpha	0th dmey	0.6485	10.00
alpha	5th gaus3	0.6485	9.31
beta1	0th db2	0.9000	3.45
beta1	1st sym7	0.9000	3.45
beta1	1st coif4	0.9000	3.45
beta1	0th None	0.9000	3.45
beta1	3rd db7	0.9000	3.45
beta2	9th None	1.0000	0.00
beta2	2nd db7	1.0000	0.00
beta2	2nd sym2	1.0000	0.00
beta2	3rd sym2	1.0000	0.00
beta2	4th sym2	1.0000	0.00

Each of the top 5 models are then plotted in Figure 28 to be visually inspected. The clinical markers such as k-complexes, v-waves, and spindles can be seen in the filters. Slow

waves can not be seen as easily.



.

Figure 28: The plot of the top 5 wavelet models as identified with the calculation represented in table 22

4.4 Discussion

Preprocessing with wavelet decomposition does provider marginal improvement in f1 score in most cases. From visual inspection, clinical markers can be seen in the CNN filters. Preprocessing with wavelet decomposition does not definitively improve the ability to see those clinical markers. Analysis on a collection of CNN filters from models trained on data preprocessed with wavelet decomposition may provide insights in the future.

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.

CHAPTER V.

Discussion

This paper focuses on PAP therapy compliance of ISP versus DME groups, a CNN model for REM sleep stage classification with a single channel EEG, and interpretability of the CNN model based on preprocessing data with many mother wavelets for DWT. These studies are intended to have tangible impacts on clinical sleep medicine. The ISP group has better PAP compliance rates at 30 days, 90 days, and 1 year. The CNN model produces comparable fit metrics to current models. The DWT preprocessing allows for more comparison to clinical markers.

These studies provide platforms for using data science to improve clinical sleep medicine and research. STARK laws could be reviewed to reduce the need for patients to use a DME PAP therapy dispenser and improve PAP therapy compliance resulting in better population health. Methodologies of monitoring compliant use of PAP therapy could be improved. The level of detail of clinical markers can be improved by investigating the CNN filters of the models created. Various preprocessing techniques can be applied to study different diseases.

STARK law prevents ISPs from dispensing PAP therapy to patients. STARK law prevents providers from profiting off of prescription drugs used within their own practice. Without this law, providers could prescribe drugs based on the profit that they would receive rather than the patient's health. While this is important for prescription drugs, treatments such as PAP therapy are not usually classified as prescription drugs. It is more common for them to be classified as durable medical equipment. switching the current classification from prescription drug to durable medical equipment would allow for providers to dispense PAP therapy to their own patients resulting in better health outcomes for patients.

The monitoring of PAP therapy is completed by the PAP therapy dispenser. When patients are not compliant, PAP therapy monitoring can initiate interventions to encourage patients to be compliant. Interventions are generally better accepted by a trusted provider

rather than a 3rd party. Monitoring is initially completed for insurance companies in the first 90 days while long term compliance monitoring is deprioritized. Long term compliance could help improve patient sleep health and over all health.

An EEG was originally divided into 30 second epochs according to the R&K standard partially because a 30 second epoch fits on a piece of paper. Today, we no longer use paper to review EEG recordings. Smaller epochs take more time to investigate by a technician but provide a more detailed analysis. The clinical markers that we see today are used to grade these 30 second epochs. EEG recordings can record beyond 1024 hertz. This level of detail could provide insights into brain activity that we previously did not know existed. Using a technician to investigate this level of granularity would be too time consuming and difficult for clinical practitioners to interpret. Relevant interpretable information needs to be discovered through data science methodologies like those described in this paper.

Patient populations with various understood characteristics need to be investigated. This can be interpreted as patients of with various dieseases and/or brain types. A standard set of CNN filters that are consistently learned across many models could be used as a description of a healthy patient. Deviations from this standard set could be interpreted as indications of various diseases. This standard set could be discovered by the investigation of CNN filters across many model and preprocessing techniques.

Data science can provide many benefits to general human population health through the investigation of clinical sleep medicine. Our ability to provide multi year monitoring and interventions with data science techniques is growing. Our ability to provide granular and interpretable investigations of medical imaging data such as EEG data is growing. As a result of data science studies, patient health is improving.

BIBLIOGRAPHY

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

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

[3] 42 U.S. Code § 1395nn - Limitation on certain physician referrals.

[4] European Data Format (EDF).

[5] Machine Learning FAQ.

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

[7] Time-frequency Signal Analysis with Applications.

[8] AASM Home Sleep Apnea Testing (HSAT) Position Statement, October 2017. Section: Press Releases.

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

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

[11] Adenekan, B., Pandey, A., McKenzie, S., Zizi, F., Casimir, G. J., and Jean-Louis, G. Sleep in America: role of racial/ethnic differences. *Sleep Medicine Reviews*, 17(4):255–262, August 2013.

[12] Alickovic, E. and Subasi, A. Ensemble SVM Method for Automatic Sleep Stage Classification. *IEEE Transactions on Instrumentation and Measurement*, 67(6):1258–1265, June 2018. Conference Name: IEEE Transactions on Instrumentation and Measurement.

[13] Aloia, M. S., Arnedt, J. T., Riggs, R. L., Hecht, J., and Borrelli, B. Clinical Management of Poor Adherence to CPAP: Motivational Enhancement. *Behavioral Sleep*

Medicine, 2(4):205–222, November 2004. Publisher: Taylor & Francis _eprint: https://doi.org/10.1207/s15402010bsm0204_3.

[14] Aloia, M. S., Arnedt, J. T., Stanchina, M., and Millman, R. P. How Early in Treatment is PAP Adherence Established? Revisiting Night-to-Night Variability. *Behavioral Sleep Medicine*, 5(3):229–240, August 2007. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15402000701264005.

[15] Aloia, M. S., Arnedt, J. T., Stepnowsky, C., Hecht, J., and Borrelli, B. Predicting treatment adherence in obstructive sleep apnea using principles of behavior change. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 1(4):346–353, October 2005.

[16] Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 2390–2397, June 2008. ISSN: 2161-4407.

[17] 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.

[18] Bakker, J. P., O'Keeffe, K. M., Neill, A. M., and Campbell, A. J. Ethnic Disparities in CPAP Adherence in New Zealand: Effects of Socioeconomic Status, Health Literacy and Self-Efficacy. *Sleep*, 34(11):1595–1603, November 2011.

[19] Bandura, A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84:191–215, 1977. Place: US Publisher: American Psychological Association.

[20] Bandura, A. Health promotion by social cognitive means. *Health Education & Behavior: The Official Publication of the Society for Public Health Education*, 31(2):143–164, April 2004.

[21] Baron, K. G., Gilyard, S. G., Williams, J. L., Lindich, D., Koralnik, L., and Lynch, E. B. Sleep-related attitudes, beliefs, and practices among an urban-dwelling African American community: a qualitative study. *Sleep Health*, 5(4):418–425, August 2019.

[22] 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.

[23] Billings, M. E., Auckley, D., Benca, R., Foldvary-Schaefer, N., Iber, C., Redline, S., Rosen, C. L., Zee, P., and Kapur, V. K. Race and Residential Socioeconomics as Predictors of CPAP Adherence. *Sleep*, 34(12):1653–1658, December 2011.

[24] Budhiraja, R., Parthasarathy, S., Drake, C. L., Roth, T., Sharief, I., Budhiraja, P., Saunders, V., and Hudgel, D. W. Early CPAP use identifies subsequent adherence to CPAP therapy. *Sleep*, 30(3):320–324, March 2007.

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

[26] Cai, Q., Gao, Z., An, J., Gao, S., and Grebogi, C. A Graph-Temporal fused dual-input Convolutional Neural Network for Detecting Sleep Stages from EEG Signals. *IEEE Transactions on Circuits and Systems II: Express Briefs*, pages 1–1, 2020. Conference Name: IEEE Transactions on Circuits and Systems II: Express Briefs.

[27] Chai-Coetzer, C. L., Luo, Y.-M., Antic, N. A., Zhang, X.-L., Chen, B.-Y., He, Q.-Y., Heeley, E., Huang, S.-G., Anderson, C., Zhong, N.-S., and McEvoy, R. D. Predictors

of long-term adherence to continuous positive airway pressure therapy in patients with obstructive sleep apnea and cardiovascular disease in the SAVE study. *Sleep*, 36(12):1929–1937, December 2013.

[28] Chen, Q., Yue, H., Pang, X., Lei, W., Zhao, G., Liao, E., and Wang, Y. Mr-ResNeXt: A Multi-resolution Network Architecture for Detection of Obstructive Sleep Apnea. In Zhang, H., Zhang, Z., Wu, Z., and Hao, T., editors, *Neural Computing for Advanced Applications*, Communications in Computer and Information Science, pages 420–432, Singapore, 2020. Springer.

[29] Chen, Z., Wu, M., Cui, W., Liu, C., and Li, X. An Attention Based CNN-LSTM Approach for Sleep-Wake Detection with Heterogeneous Sensors. *IEEE Journal of Biomedical and Health Informatics*, pages 1–1, 2020. Conference Name: IEEE Journal of Biomedical and Health Informatics.

[30] Collen, J. F., Lettieri, C. J., and Hoffman, M. The Impact of Posttraumatic Stress Disorder on CPAP Adherence in Patients with Obstructive Sleep Apnea. *Journal of Clinical Sleep Medicine : JCSM : Official Publication of the American Academy of Sleep Medicine*, 8(6):667–672, December 2012.

[31] 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.

[32] Daubechies, I. *Ten Lectures on Wavelets*. SIAM, 1992.

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

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

[35] DiMatteo, M. R. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. *Medical Care*, 42(3):200–209, March 2004.

[36] 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.

[37] Esfahanian, P. and Akhavan, M. GACNN: Training Deep Convolutional Neural Networks with Genetic Algorithm. *arXiv:1909.13354 [cs]*, September 2019. arXiv: 1909.13354.

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

[39] Goshtasbi, N., Boostani, R., and Sanei, S. SleepFCN: A Fully Convolutional Deep Learning Framework for Sleep Stage Classification Using Single-Channel Electroencephalograms. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:2088–2096, 2022. Conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering.

[40] Haskard Zolnierek, K. B. and DiMatteo, M. R. Physician Communication and Patient Adherence to Treatment: A Meta-Analysis. *Medical Care*, 47(8):826, August 2009.

[41] Hassan, A. R. and Bhuiyan, M. I. H. A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features. *Journal of Neuroscience Methods*, 271:107–118, September 2016.

[42] Haynes, R. B., Montague, P., Oliver, T., McKibbon, K. A., Brouwers, M. C., and Kanani, R. Interventions for helping patients to follow prescriptions for medications. *The Cochrane Database of Systematic Reviews*, (2):CD000011, 2000.

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

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

[45] Hong, D., Wang, J., and Gardner, R. *Real Analysis with an Introduction to Wavelets and Applications*. Elsevier Academic Press, 2005.

[46] Huang, S., Zhu, J., Chen, Y., Wang, T., and Ma, T. Analysis and Classification of Sleep Stages Based on Common Frequency Pattern From a Single-Channel EEG Signal. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 3711–3714, July 2020. ISSN: 1558-4615.

[47] Jacobsen, A. R., Eriksen, F., Hansen, R. W., Erlandsen, M., Thorup, L., Damgård, M. B., Kirkegaard, M. G., and Hansen, K. W. Determinants for adherence to continuous positive airway pressure therapy in obstructive sleep apnea. *PloS One*, 12(12):e0189614, 2017.

[48] Kemp, B., Zwinderman, A., Tuk, B., Kamphuisen, H., and Oberyé, J. The Sleep-EDF Database [Expanded], 2018. Type: dataset.

[49] Kribbs, N. B., Pack, A. I., Kline, L. R., Smith, P. L., Schwartz, A. R., Schubert, N. M., Redline, S., Henry, J. N., Getsy, J. E., and Dinges, D. F. Objective measurement of

patterns of nasal CPAP use by patients with obstructive sleep apnea. *The American Review of Respiratory Disease*, 147(4):887–895, April 1993.

[50] 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.

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

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

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

[54] Lee, T., Hwang, J., and Lee, H. TRIER: Template-Guided Neural Networks for Robust and Interpretable Sleep Stage Identification from EEG Recordings. *arXiv:2009.05407 [cs, eess]*, September 2020. arXiv: 2009.05407.

[55] Li, J., Struzik, Z., Zhang, L., and Cichocki, A. Feature learning from incomplete EEG with denoising autoencoder. *Neurocomputing*, 165:23–31, October 2015.

[56] Li, Q., Wang, B., Jin, J., and Wang, X. Comparison of CNN-Uni-LSTM and CNN-Bi-LSTM based on single-channel EEG for sleep staging. In *2020 5th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, pages 76–80, November 2020. ISSN: 2189-8723.

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

[58] Mallat, S. *A Wavelet Tour of Signal Processing, (3rd Edition)*. Elsevier, 2008.

[59] Marin, J. M., Carrizo, S. J., Vicente, E., and Agusti, A. G. N. Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without

treatment with continuous positive airway pressure: an observational study. *Lancet (London, England)*, 365(9464):1046–1053, March 2005.

[60] McArdle, N., Devereux, G., Heidarnejad, H., Engleman, H. M., Mackay, T. W., and Douglas, N. J. Long-term use of CPAP therapy for sleep apnea/hypopnea syndrome. *American Journal of Respiratory and Critical Care Medicine*, 159(4 Pt 1):1108–1114, April 1999.

[61] McEvoy, R. D., Antic, N. A., Heeley, E., Luo, Y., Ou, Q., Zhang, X., Mediano, O., Chen, R., Drager, L. F., Liu, Z., Chen, G., Du, B., McArdle, N., Mukherjee, S., Tripathi, M., Billot, L., Li, Q., Lorenzi-Filho, G., Barbe, F., Redline, S., Wang, J., Arima, H., Neal, B., White, D. P., Grunstein, R. R., Zhong, N., Anderson, C. S., and SAVE Investigators and Coordinators,. CPAP for Prevention of Cardiovascular Events in Obstructive Sleep Apnea. *The New England Journal of Medicine*, 375(10):919–931, September 2016.

[62] Meurice, J.-C., Dore, P., Paquereau, J., Neau, J.-P., Ingrand, P., Chavagnat, J.-J., and Patte, F. Predictive Factors of Long-term Compliance With Nasal Continuous Positive Airway Pressure Treatment in Sleep Apnea Syndrome. *Chest*, 105(2):429–433, February 1994.

[63] Mousavi, S., Afghah, F., and Acharya, U. R. SleepEEGNet: Automated Sleep Stage Scoring with Sequence to Sequence Deep Learning Approach. *PLOS ONE*, 14(5):e0216456, May 2019. arXiv:1903.02108 [cs, eess, q-bio].

[64] Nikhil Chandran, A., Sreekumar, K., and Subha, D. P. EEG-Based Automated Detection of Schizophrenia Using Long Short-Term Memory (LSTM) Network. In Patnaik, S., Yang, X.-S., and Sethi, I. K., editors, *Advances in Machine Learning and Computational Intelligence*, Algorithms for Intelligent Systems, pages 229–236, Singapore, 2021. Springer.

[65] Pamidi, S., Knutson, K. L., Ghods, F., and Mokhlesi, B. The impact of sleep consultation prior to a diagnostic polysomnogram on continuous positive airway pressure adherence. *Chest*, 141(1):51–57, January 2012.

[66] Pelletier-Fleury, N., Rakotonanahary, D., and Fleury, B. The age and other factors in the evaluation of compliance with nasal continuous positive airway pressure for obstructive sleep apnea syndrome. A Cox's proportional hazard analysis. *Sleep Medicine*, 2(3):225–232, May 2001.

[67] Prochaska, J. O. Strong and weak principles for progressing from precontemplation to action on the basis of twelve problem behaviors. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, 13(1):47–51, January 1994.

[68] Prochaska, J. O. and DiClemente, C. C. Stages and processes of self-change of smoking: toward an integrative model of change. *Journal of Consulting and Clinical Psychology*, 51(3):390–395, June 1983.

[69] Prochaska, J. O. and Velicer, W. F. The transtheoretical model of health behavior change. *American journal of health promotion: AJHP*, 12(1):38–48, 1997.

[70] 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.

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

[72] Richards, D., Bartlett, D. J., Wong, K., Malouff, J., and Grunstein, R. R. Increased adherence to CPAP with a group cognitive behavioral treatment intervention: a randomized trial. *Sleep*, 30(5):635–640, May 2007.

[73] Rosenthal, L., Gerhardstein, R., Lumley, A., Guido, P., Day, R., Syron, M. L., and Roth, T. CPAP therapy in patients with mild OSA: implementation and treatment outcome. *Sleep Medicine*, 1(3):215–220, July 2000.

[74] Rotenberg, B. W., Murariu, D., and Pang, K. P. Trends in CPAP adherence over twenty years of data collection: a flattened curve. *Journal of Otolaryngology - Head & Neck Surgery = Le Journal D'oto-Rhino-Laryngologie Et De Chirurgie Cervico-Faciale*, 45(1):43, August 2016.

[75] Sackett, D. L., Haynes, R. B., Gibson, E. S., Taylor, D. W., Roberts, R. S., and Johnson, A. L. Patient compliance with antihypertensive regimens. *Patient Counselling and Health Education*, 1(1):18–21, 1978.

[76] Santaji, S. and Desai, V. Analysis of EEG Signal to Classify Sleep Stages Using Machine Learning. *Sleep and Vigilance*, August 2020.

[77] santé, O. m. d. l. and Organization, W. H. *Adherence to Long-term Therapies: Evidence for Action*. World Health Organization, 2003. Google-Books-ID: kcYUTH8rPiwC.

[78] Sawyer, A. M., Canamucio, A., Moriarty, H., Weaver, T. E., Richards, K. C., and Kuna, S. T. Do cognitive perceptions influence CPAP use? *Patient Education and Counseling*, 85(1):85–91, October 2011.

[79] Sawyer, A. M., Gooneratne, N. S., Marcus, C. L., Ofer, D., Richards, K. C., and Weaver, T. E. A systematic review of CPAP adherence across age groups: clinical and empiric insights for developing CPAP adherence interventions. *Sleep Medicine Reviews*, 15(6):343–356, December 2011.

[80] Scharf, S. M., Seiden, L., DeMore, J., and Carter-Pokras, O. Racial differences in clinical presentation of patients with sleep-disordered breathing. *Sleep & Breathing = Schlaf & Atmung*, 8(4):173–183, December 2004.

[81] Shambroom, J. R., Fábregas, 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.

[82] Sin, D. D., Mayers, I., Man, G. C. W., and Pawluk, L. Long-term compliance rates to continuous positive airway pressure in obstructive sleep apnea: a population-based study. *Chest*, 121(2):430–435, February 2002.

[83] 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.

[84] Smith, J. R., Negin, M., and Nevis, A. H. Automatic Analysis of Sleep Electroencephalograms by Hybrid Computation. *IEEE Transactions on Systems Science and Cybernetics*, 5(4):278–284, October 1969. Conference Name: IEEE Transactions on Systems Science and Cybernetics.

[85] Stepnowsky, C. J., Marler, M. R., and Ancoli-Israel, S. Determinants of nasal CPAP compliance. *Sleep Medicine*, 3(3):239–247, May 2002.

[86] Stradling, J. R. and Davies, R. J. Is more NCPAP better? *Sleep*, 23 Suppl 4:S150–153, June 2000.

[87] Supratak, A., Dong, H., Wu, C., and Guo, Y. DeepSleepNet: a Model for Automatic Sleep Stage Scoring based on Raw Single-Channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11):1998–2008, November 2017. arXiv: 1703.04046.

[88] Supratak, A. and Guo, Y. TinySleepNet: An Efficient Deep Learning Model for Sleep Stage Scoring based on Raw Single-Channel EEG. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 641–644, July 2020. ISSN: 1558-4615.

[89] van de Mortel, T. F., Laird, P., and Jarrett, C. Client perceptions of the polysomnography experience and compliance with therapy. *Contemporary Nurse*, 9(2):161–168, June 2000.

[90] 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.

[91] Wallace, D. M., Sawyer, A. M., and Shafazand, S. Comorbid insomnia symptoms predict lower 6-month adherence to CPAP in US veterans with obstructive sleep apnea. *Sleep & Breathing = Schlaf & Atmung*, 22(1):5–15, March 2018.

[92] Wallace, D. M., Shafazand, S., Aloia, M. S., and Wohlgemuth, W. K. The Association of Age, Insomnia, and Self-Efficacy with Continuous Positive Airway Pressure Adherence in Black, White, and Hispanic US Veterans. *Journal of Clinical Sleep Medicine : JCSM : Official Publication of the American Academy of Sleep Medicine*, 9(9):885–895, September 2013.

[93] Wang, Q., Wang, S., Wu, N., and Xu, L.-Q. An Automatic Sleep Staging Method Using a Multi-head and Sequence Network. In *Proceedings of the Fourth International Conference on Biological Information and Biomedical Engineering*, BIBE2020, pages 1–7, New York, NY, USA, July 2020. Association for Computing Machinery.

[94] Weaver, T. E., Kribbs, N. B., Pack, A. I., Kline, L. R., Chugh, D. K., Maislin, G., Smith, P. L., Schwartz, A. R., Schubert, N. M., Gillen, K. A., and Dinges, D. F. Night-to-night variability in CPAP use over the first three months of treatment. *Sleep*, 20(4):278–283, April 1997.

[95] Weaver, T. E. and Grunstein, R. R. Adherence to continuous positive airway pressure therapy: the challenge to effective treatment. *Proceedings of the American Thoracic Society*, 5(2):173–178, February 2008.

[96] Weaver, T. E., Maislin, G., Dinges, D. F., Bloxham, T., George, C. F. P., Greenberg, H., Kader, G., Mahowald, M., Younger, J., and Pack, A. I. Relationship Between Hours of CPAP Use and Achieving Normal Levels of Sleepiness and Daily Functioning. *Sleep*, 30(6):711–719, June 2007.

[97] Wei, L., Ventura, S., Lowery, M., Ryan, M. A., Mathieson, S., Boylan, G. B., and Mooney, C. Random Forest-based Algorithm for Sleep Spindle Detection in Infant EEG. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 58–61, July 2020. ISSN: 1558-4615.

[98] Wickwire, E. M., Lettieri, C. J., Cairns, A. A., and Collop, N. A. Maximizing positive airway pressure adherence in adults: a common-sense approach. *Chest*, 144(2):680–693, August 2013.

[99] Wolpert, E. A. A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. *Archives of General Psychiatry*, 20(2):246, February 1969.

[100] Xu, M., Wang, X., Zhangt, X., Bin, G., Jia, Z., and Chen, K. Computation-Efficient Multi-Model Deep Neural Network for Sleep Stage Classification. In *Proceedings*

of the 2020 Asia Service Sciences and Software Engineering Conference, ASSE '20, pages 1–8, New York, NY, USA, May 2020. Association for Computing Machinery.

[101] Zolnierek, K. B. H. and Dimatteo, M. R. Physician communication and patient adherence to treatment: a meta-analysis. *Medical Care*, 47(8):826–834, August 2009.

[102] Álvarez, D., Cerezo-Hernández, A., López-Muñiz, 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.