**AUTOMATED IDENTIFICATION OF ABNORMAL ADULT EEGs**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**A Thesis Proposal**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**in Partial Fulfillment**

**of the Requirements for the Degree of**

**Master of Science in Electrical and Computer Engineering**

**By**

**Silvia López de Diego**

**December, 2016**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Dr. Joseph Picone**

**Thesis Advisor**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Dr. Pallavi Chitturi Dr. Iyad Obeid**

**Professor of Statistical Professor of Electrical &**

**Science Computer Engineering**

**Committee Member Committee Member**

# ABSTRACT

The interpretation of electroencephalograms (EEGs) is a process that is still dependent on the subjective analysis of the examiners. Though inter-rater agreement on critical events such as seizures can be high, it is much lower on subtler events (e.g., when there are benign variants). The focus of this study is to automatically classify normal and abnormal EEGs to provide neurologists with real-time decision support.

A demographically balanced subset of the TUH EEG Corpus was used to evaluate performance. The data, comprised of 200 abnormal EEGs and 202 normal EEGs was manually selected. This subset was partitioned into a training set (82 normal/80 abnormal) and an evaluation set (51 normal/55 abnormal). Principal Components Analysis (PCA) was used to reduce the dimensionality of the data. Two baseline classification algorithms were explored: k-Nearest Neighbor (kNN), Random Forest Ensemble Learning (RF). kNN achieved a 41.8% detection error rate while RF achieved an error rate of 31.7%. These error rates are significantly lower than those obtained by random guessing based on priors (49.5%). These algorithms were then compared to a Hidden Markov Models (HMM) based approach, which reduced the error rate to 17.0%, which is approaching human performance. Several deep learning architectures will also be explored in this thesis.

# TABLE OF CONTENTS

[ABSTRACT ii](#_Toc468487701)

[TABLE OF CONTENTS iii](#_Toc468487702)

[LIST OF TABLES v](#_Toc468487703)

[LIST OF FIGURES vi](#_Toc468487704)

[CHAPTER 1 1](#_Toc468487705)

[INTRODUCTION 1](#_Toc468487706)

[The Normal Adult EEG 2](#_Toc468487707)

[Visual Analysis of EEGs 4](#_Toc468487708)

[Automatic Abnormal EEG Classification 5](#_Toc468487709)

[Thesis Overview 6](#_Toc468487710)

[CHAPTER 2 7](#_Toc468487711)

[CLASSIFICATION OF SEQUENTIAL DATA 7](#_Toc468487712)

[2.1 Markov Models 7](#_Toc468487713)

[2.1.1 Hidden Markov Models (HMMs) 8](#_Toc468487714)

[2.1.2 Gaussian Mixture Models (GMMs) 10](#_Toc468487715)

[2.2 Performance of GMMs-HMMs Compared with Deep Neural Networks (DNNs) 11](#_Toc468487716)

[CHAPTER 3 12](#_Toc468487717)

[DATA & EXPERIMENTS 12](#_Toc468487718)

[3.1 Data 12](#_Toc468487719)

[3.1.1 Data Subset Selection 12](#_Toc468487720)

[3.1.2 Feature Extraction 13](#_Toc468487721)

[3.1.3 Dimensionality Reduction 15](#_Toc468487722)

[3.2 Experiments 15](#_Toc468487723)

[3.2.1 Baseline Experiments 16](#_Toc468487724)

[3.1.2 HMM Experiments 17](#_Toc468487725)

[CHAPTER 4 18](#_Toc468487726)

[PRELIMINARY EXPERIMENTS AND RESULTS 18](#_Toc468487727)

[4.1 Baseline Results 18](#_Toc468487728)

[4.1 GMM-HMM Results 21](#_Toc468487729)

[CHAPTER 5 24](#_Toc468487730)

[EXPECTATIONS AND FUTURE WORK 24](#_Toc468487731)

[5.1 Expected Outcomes 24](#_Toc468487732)

[5.2 Timeline for Future Work 24](#_Toc468487733)

[REFERENCES 26](#_Toc468487734)

# LIST OF TABLES

**Table 1.** Summary of word error rates for a subspace Gaussian Mixture Model and a Deep Neural Network. 11

**Table 2.** Comparison of the performance obtained with the two baseline systems 21

**Table 3.** Confusion matrix for the kNN system 21

**Table 4.** GMM-HMM correct detection rate for various numbers of Gaussian 22

**Table 5.** Correct detection rate for different signal input lengths 22

**Table 6.** Correct detection rate for different channels 23

**Table 8.** Confusion matrix for the best GMM-HMM system 23

**Table 7.** Summary of the performance for all the evaluated systems 23

# LIST OF FIGURES

**Figure 1.** Summary of the common steps that are followed for a clinical EEG recording and interpretation 2

**Figure 2.** Decission tree that shows the process that is generally followed for the abnormal identification of EEGs 5

**Figure 3.** Temporal Evolution of a seizure in the T4-A2 channel of an EEG. The top of the figure shows the spectrogram of the signal, while the bottom panel shows the signal in the time domain. 7

**Figure 4.** Example of a basic Markov model with states and transition probabilities 8

**Figure 5.** Example of a Hidden Markov model with states , transition probabilities , emission probabilities and visible stated . 9

**Figure 6.** HMM based phone model with transition probabilities and observation distributions 9

**Figure 7.** Distribution of the patients’ ages and genders for the selected dataset 12

**Figure 8.** Illustration of the base feature extraction process. 14

**Figure 9.** Location of studied channels in the 10-20 standard system of electrode placement for the TCP montage 17

**Figure 10.** Normal/abnormal classification error rate as a function of number of (trees ) 18

**Figure 11.** Error rate of the kNN and RF algorithms as a function of the PCA dimension 19

**Figure 12.** Errorrate as a function of the number of neighbors for PCA dimension of 20and 86 20

**Figure 13.** Classification error rate (for kNN) for a fronto-central (F4-C4) and a temporal-occipital (T5-O1) channel 20

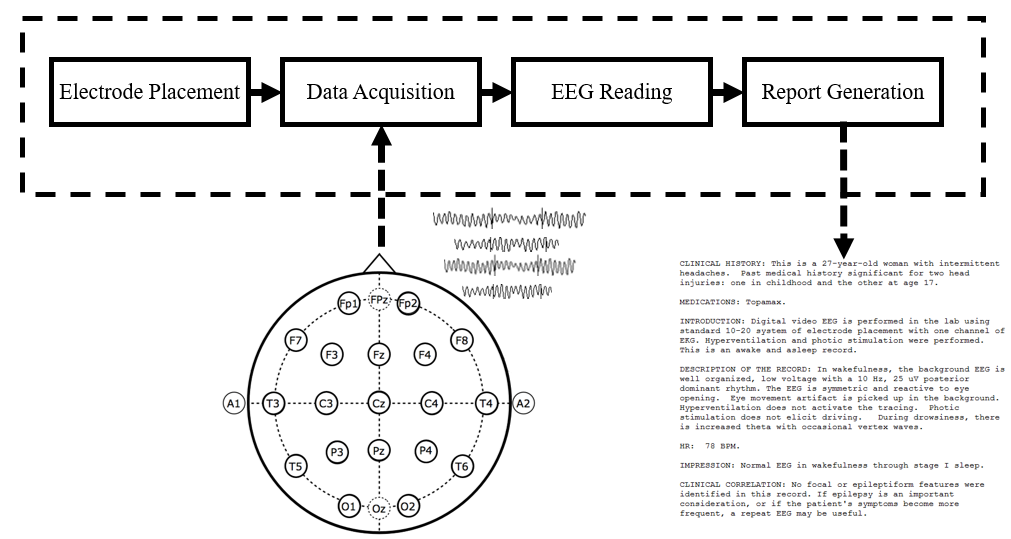
# CHAPTER 1

# INTRODUCTION

The recording of the neurons’ electrical activity along the scalp, known as electroencephalography (EEG), has been widely used for the diagnosis and management of conditions such as sleep disorders and epilepsy in the past 30 years. Despite the emergence of new technologies, such as Magnetic Resonance Imaging (MRI), the noninvasiveness and relative inexpensiveness of EEGs make this technique a popular choice as a diagnostics tool among physicians (Smith, 2005). A typical routine EEG has a duration of about 20 minutes, duration that is not always enough to record ictal (or interictal activity) in patients with seizure disorders. As a matter of fact, only 50% of patients with epilepsy show interictal epileptiform discharges (IED) in their first recording, and increasing the amount of recordings up to four seems to increase the yield of the tests (Smith, 2005). In this sense, the diagnosis and characterization of epilepsy usually require more than one routine EEG and/or long term monitoring (LTM) EEG recordings, which last for more extended periods of time (hours to days).

EEG records are interpreted by board certified physicians, process that, because of the demand for EEGs and the time that takes for their reading, can introduce a lag time, which ranges from days to weeks, to the diagnosis and decision making process. Additionally, reading EEGs depends heavily on the subjective judgement of the reader, which can lead to missed events or misdiagnosis of the patient (Azuma et al., 2003). Introducing a certain level of automation to the EEG interpretation task could potentially serve as an aid for the neurologists to speed up the reading process and ease some of the pressure that results from the high demand of EEGs by patients that are in the need of a diagnosis or the management of their conditions.

The EEG recording workflow involves the placement of the electrodes on the patient’s scalp (for scalp EEG) by the EEG technicians, the acquisition of the EEG data, the interpretation of the signals by a certified neurologists and the generation of the report that is presented to the patient (Amir Harati et al., 2014a). The EEG report contains a combination of the history, medications, description of the record and interesting findings. One portion of the report, however, contains the impression of the record, which shows whether the EEG is normal or abnormal given the EEG activity recorded in the session. **Figure 1** summarizes the procedure for a typical EEG recording.



**Figure 1.** Summary of the common steps that are followed for a clinical EEG recording and interpretation

The medical report that is produced for each EEG session describes the record, the recording conditions and summarizes the findings. One decision that is also shown in the report is whether the characteristics of the EEG was found to be within the normal limits for patients in a similar group of age and gender. The main objective for this study is to utilize machine learning techniques in order to generate the Normal/Abnormal decision automatically. If the proposed technology reaches clinically accepted performance, it could potentially serve as an aid to neurologists during the EEG interpretation task and reduce the lag time between EEG recording and reporting to the patients, establishing a more efficient workflow.

## The Normal Adult EEG

The EEG interpretation task can be broken down into two different parts: the analysis of the EEG background and the recognition of the transients (Finnigan & van Putten, 2013). The background pattern refers to the general characteristics of an EEG, which include the features that neurologists observe when making a normal/abnormal decision about the record. Some remarkable examples of the background pattern are the posterior dominant rhythm (PDR) and the frequency distributions of the signals throughout the scalp (Lodder & van Putten, 2013). The transient patterns, on the other hand, refer to rarer events that include pathological and physiological waveforms, such as spikes and sharp waves discharges.

The Characterization of the normal adult EEG has been based on a specific description of the background pattern and the presence—or lack thereof—of certain transient waveforms given the patient’s state of consciousness (awake EEG is different from drowsy EEG or comatose EEG). The main background characteristics of the normal adult EEG can be summarized as follows (Ebersole & Pedley, 2014):

1. Reactivity: Refers to the response to certain physiological changes. This changes could be eye opening and closing, sensory stimulation, etc.
2. Alpha Rhythm: This rhythm is the starting point for the visual analysis of EEGs. The presence, characteristics according to age and reactivity of this feature play an important role in the normal/abnormal classification of the EEG. The Alpha waves originate (predominantly) in the occipital lobe and are between 8-13 Hz in frequency and 15 to 45 V in amplitude.
3. Mu Rhythm: It is a central rhythm of frequencies between 8 to 10 Hz with amplitudes comparable to the alpha rhythm. This rhythm is suppressed unilaterally by the movement of the opposite extremity. This rhythm, however, is also suppressed by conditions such as fatigue, somatosensory and sensorimotor stimulation. In this sense, the Mu rhythm is not always detectable.
4. Beta Activity: Rhythm with frequencies of 18-25 Hz, 14-16 Hz and 35-40 Hz, with amplitudes between 5 and 20 mV. It is important to note, however, that it is rare to see activity higher than 25 Hz on scalp EEGs.
5. Theta Activity: Normal adults tend to show traces of less than 15 V 6-7 Hz activity in the frontal and frontocentral regions and occasionally in the midline central region. This rhythm, called Theta activity, usually becomes sustained and higher in voltage with the onset of drowsiness.

These features provide a description of the characteristics that are systematically observed by neurologists when evaluating an EEG. Characteristics such as the state and the age of the patient are also important considerations that may alter the characteristics presented above. These characteristics are mainly observed in normal adult EEGs (Ebersole & Pedley, 2014).

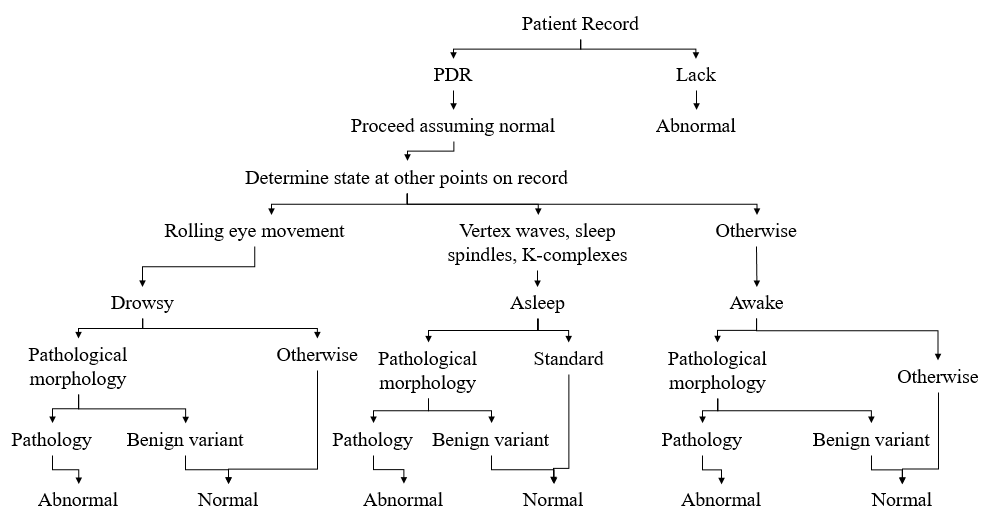
## Visual Analysis of EEGs

The previous section described the general features that characterize a normal EEG. The presence of these features does not necessarily guarantee the normality of the record. As it was explained before, these characteristics comprise what is defined as the background EEG, without taking into account transient patterns that could additionally be present in the record.

Neurologists analyze the records by evaluating the background EEG and determining whether the patient presents normal characteristics according to his or her state. If the patient did not present abnormal transients during the recording, and the background EEG was within normal limits, the recording is considered normal. The analysis of the background is broken down in steps that allow to take all of the characteristics into account in a systematic way. The analysis steps that ultimately lead to a decision about the normality of the record involve the observation of the following characteristics: frequency, voltage, waveform, regulation (make sure the alpha rhythm does not vary more than ±0.5 Hz ), locus, reactivity and interhemispheric coherence (Ebersole & Pedley, 2014).

The systematic visual analysis of EEGs usually starts with the evaluation of the occipital alpha rhythm, also called the Posterior Dominant Rhythm. This activity emerges in the occipital region when the eyes are closed, and fades as the patients enter a state of drowsiness. In this sense, the evaluation of the reactivity for the emergence of the PDR is one of the main features that are evaluated in order to make a decision about the normality of the record (Ebersole & Pedley, 2014). A decision tree for the evaluation of the normality of an EEG record is presented in **Figure 2** (Lopez, Suarez, Jungries, Obeid, & Picone, 2015).

In essence, a common characteristic that is evaluated in adults is the PDR and the reactivity of its occipital emergence. Throughout this study, the characteristics of the background EEG and, especially the PDR, are taken into consideration in order to make a normal/abnormal classification with the implementation of state-of-the-art machine learning techniques.



**Figure 2.** Decission tree that shows the process that is generally followed for the abnormal identification of EEGs

## Automatic Abnormal EEG Classification

A generalized algorithm or method for the classification of clinical abnormal EEGs is a task that has not been yet explored. While some work has been done in the identification of EEG abnormalities specific to certain pathological or physiological conditions, the study of the general background EEG as a resource for the classification of normal and abnormal records has not been investigated. For instance, studies have been done in order to classify athletes with residual functional deficits after the occurrence of a concussion with the help of EEG data and Support Vector Machines (SVMs) (Cao, Tutwiler, & Slobounov, 2008). These study, however, did not rely on clinical EEG data, and the classifier was designed to train a very specific condition.

In this work, baselines for the classification of normal and abnormal clinical EEG records are established. Two non-parametric algorithms, k-Nearest Neighbor and Random Forest Ensemble Learning, explained in more depth in Chapter 2, were used for the establishment of the baseline systems. Hidden Markov Models, also explained in Chapter 2, were then utilized for classification and comparison with the baselines.

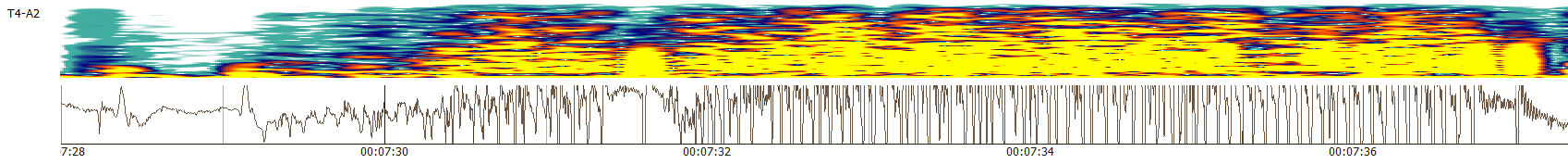
## Thesis Overview

This study presents the results of the classification of clinical EEGs obtained through a series of pilot studies, where the normal/abnormal classification problem was approached with K-nearest neighbors (KNN), and Random Forest Ensemble learning (RF). In addition, these preliminary studies are compared to a parametric approach based on Hidden Markov Models (HMMs). In Chapter 2, a description of the HMM utilized in this study will be presented in more depth. Chapter 3 will introduce the dataset utilized for this classification study, along with the subset selection process and the experimental setup. In Chapter 4, the results of the preliminary experiments will be presented and discussed. Finally, Chapter 5 will offer an idea of the expected outcomes of this study and a timeline in which the future work will be completed.

# CHAPTER 2

# CLASSIFICATION OF SEQUENTIAL DATA

Electroencephalography signals, like speech signals, are the product of a physiological process that unfolds in time. **Figure 3,** for instance, shows the temporal evolution of a seizure in one EEG channel and its respective spectrogram. In this sense, machine learning approaches that treat the observations in the data as independent and identically distributed (i.i.d.) would not successfully exploit the sequential nature of the data (Bishop, 2011). The inherent temporality of EEGs and the success that HMMs have shown in the area of speech recognition (Picone, 1990) serve as a motivation to select these models for the decoding and classification of EEG signals. Accordingly, the remainder of this chapter will offer the reader the necessary theoretical background for HMMs and a brief explanation of how these models have been implemented for continuous speech recognition. The following sections mostly follow theoretical explanations that can be found in (Duda, Hart, & Stork, 2001), (Rabiner, 1989) and (Bishop, 2011).



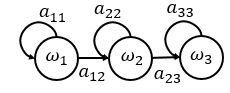
**Figure 3.** Temporal Evolution of a seizure in the T4-A2 channel of an EEG. The top of the figure shows the spectrogram of the signal, while the bottom panel shows the signal in the time domain.

## 2.1 Markov Models

If a sequence of states at subsequent times are considered, the state at a time is denoted as . The description of the model for a specific sequence (where T represents the length of the sequence) is then given by:

**(1)**

where represents a transition probability, or the probability of being in state at given that the state at is (Duda et al., 2001). The state at step in a first-order Markov model is a function that only depends on . Higher order Markov chains allow to consider states at earlier steps. So far, an observable Markov model, in which each step corresponds to an observable event, has been described. **Figure 4** shows an illustration of a three-state Markov model, with its respective states represented by nodes and the transition probabilities represented by links.



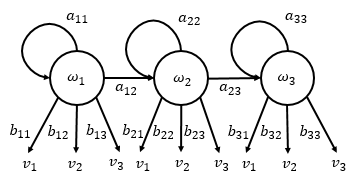
**Figure 4.** Example of a basic Markov model with states and transition probabilities

So far, the states of the model that has been described correspond to an observable event, which constitutes a restrictive model to be applied to problems such as speech recognition (Rabiner, 1989). As a matter of fact, in speech recognition systems, the perceiver does not have access to the Markov model states. On the contrary, spectral properties of the emitted sounds are measured and the outputs are analyzed in visible states , which represent a new set of stochastic processes that produce a sequence of directly accessible observations. This resulting augmented model describes a hidden Markov Model (HMM).

### 2.1.1 Hidden Markov Models (HMMs)

Hidden Markov models can be considered an extension, or augmentation, of the models that have been described to this point. In fact, in the case of HMMs, the visible observations are given by a probabilistic function of the state (Rabiner, 1989). In this augmented model, pictorially represented in **Figure 5**, the assumption that at every single time the system is at state is kept. However, for HMMs, the assumption that the system also emits a visible observation or symbol is also made. In this way (assuming a discrete symbol is emitted at each state), a probability of emitting a specific visible state is given by (Duda et al., 2001):

**(2)**



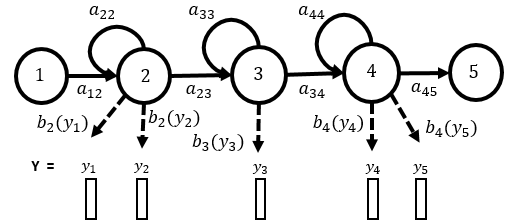
**Figure 5.** Example of a Hidden Markov model with states , transition probabilities , emission probabilities and visible stated .

In speech recognition, each spoken word is decomposed into a sequence of sounds (or base phones), which have pronunciation sequences . The likelihood of a word given an acoustic feature vector is given by:

**(3)**

Given the modeling difficulty for , Bayes Rule can be used to transform this equation into the equation specified by

**(4)**



**Figure 6.** HMM based phone model with transition probabilities and observation distributions

where represents the acoustic model and represents the language model (Gales & Young, 2007). For the explanation of pertinent concepts, the focus will be centered around the acoustic model.

If the decoding of the word “bat” is considered, for example, each of the valid pronunciations for the phones that comprise the word (/b/, /ae/ and /t/) would be represented by a continuous density HMM of the form shown by **Figure 6** (Gales & Young, 2007), and the likelihood would be given by

**(5)**

where represents a sequence of valid pronunciations. If the assumption of a single multivariate Gaussian is made for the output distribution, then would be given by:

**(6)**

where is the mean of state and representsits covariance. In this sense, the acoustic likelihood is described as follows:

**(7)**

where represents a state sequence through the model (Gales & Young, 2007).

The parameters for the acoustic model are commonly estimated through the forward-backward algorithm as explained in (Baum, Petrie, Soules, & Weiss, 1970). In general, this approach, which is a generalized instance of the Expectation Maximization (EM) algorithm, updates the weights of the system to better explain the observed training sequences.

### 2.1.2 Gaussian Mixture Models (GMMs)

The explanations that have been presented so far rely on the fact that a single Gaussian distribution models the state—output distributions. In problems like speech recognition, or electroencephalography in the present case, the utilization of a single Gaussian distribution is not necessary accurate, since this implies the assumption that the feature vectors are symmetric and unimodal. Variations such as speaker identity, accent, gender and others make this assumption rarely possible in practice (Gales & Young, 2007). To overcome this issue, several systems have successfully implemented mixtures of Gaussians (Steve Young, 1996), which are able to properly, and more accurately, model multi-modal data. If Gaussian mixture models are implemented, then the value for would be given by:

**(8)**

where represents the prior probability for mixture component of state . Commonly, the number of Gaussian mixtures is selected through the testing of models with different number of components and their evaluation in a held out set in order to find the optimal model (Gales & Young, 2007).

## 2.2 Performance of GMMs-HMMs Compared with Deep Neural Networks (DNNs)

Over the last decade, advances in computer hardware, machine learning, and deep learning algorithms have facilitated the faster and more accurate training of Deep Neural Networks (DNN) (Hinton et al., 2012). This technology has made a series of breakthroughs in the area of Automatic Speech Recognition (ASR) in the past few years, outperforming systems based on approaches such as HMM and GMM-HMM. The performance gap, however, gets smaller as the amount of training data decreases. This observation is evident from the results that have been obtained with the Kaldi Speech Recognition Toolkit (Povey et al., 2011) on the Intelligence Advanced Research Projects Activity (IARPA) provided database, BABEL, and the Fisher English Corpus. These results are summarized in **Table 1*.***

|  |  |  |  |
| --- | --- | --- | --- |
| **Corpus** | **Training Speech** | **SGMM WER** | **DNN WER** |
| BABEL Pashto | 10 hours | 69.20% | 67.60% |
| BABEL Pashto | 80 hours | 50.20% | 42.30% |
| Fisher English | 2000 hours | 15.40% | 10.30% |

**Table 1.** Summary of word error rates for a subspace Gaussian Mixture Model and a Deep Neural Network.

The results that are shown in **Table 1** show that, indeed, DNNs are capable of achieving significant improvements in the performance of a speech recognition system. However, the difference is not as significant as when the number of training observations is not large enough.

# CHAPTER 3

# DATA & EXPERIMENTS

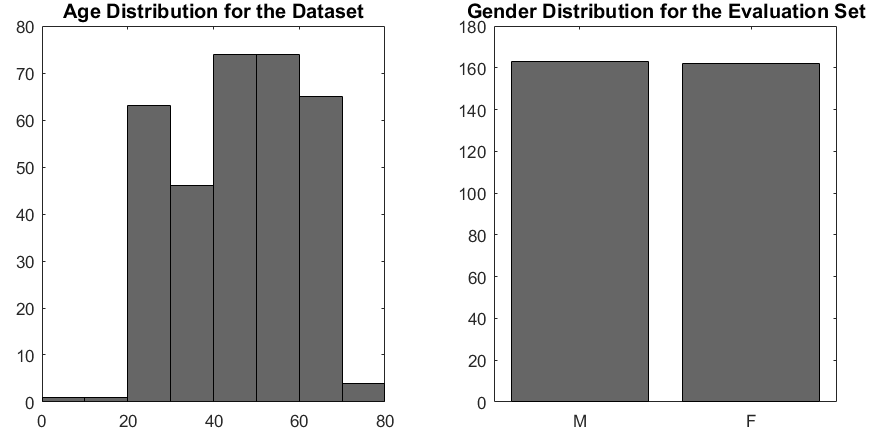
In this section, the data that was used for the experiments is characterized and the subset selection is described. Additionally, the setup for the baseline experiments, based on k-nearest neighbors and Random Forest, is presented. Finally, the experiments related to the establishment of a GMM-HMM based system for the abnormal identification of EEGs are described in detail.

## 3.1 Data

This study utilized a subset of the Temple University Hospital EEG (TUH EEG) data corpus, which represents the largest publicly available database of clinical EEGs (Amir Harati et al., 2014b). The database is currently comprised of more than 30,000 records from over 18,000 unique patients. Given the nature of this study, it is additionally important to point out the fact that around 75% of the records present in TUH EEG are abnormal.

### 3.1.1 Data Subset Selection

For the purposes of the study, a demographically balanced subset of the TUH EEG database was selected. The age and gender of the patients were considered for the selection of the data, and because pediatric EEGs are essentially very different in nature than adult EEGs (Ebersole & Pedley, 2014), the majority of the records utilized were obtained from patients that were older than 20 years old. **Figure 7** shows the histograms of ages for the training and evaluation sets respectively. It is possible to see that, excluding two outliers in the datasets, all of the patients in the age range of 20-90, with a mean of 45.72 and a standard deviation of 14.91. The genders of the patients, as it can be also seen in **Figure 7**, were also kept balanced.



**Figure 7.** Distribution of the patients’ ages and genders for the selected dataset

The final data subset that was divided in two sets: a training, which contained 80 abnormal and 82 normal EEGs, and an evaluation set, which contained 55 abnormal and 51 normal EEGs. From these recordings, only one channel was taken into consideration for the final analysis. The utilized channel was selected through experimental results, in a process that will be shown in a later section.

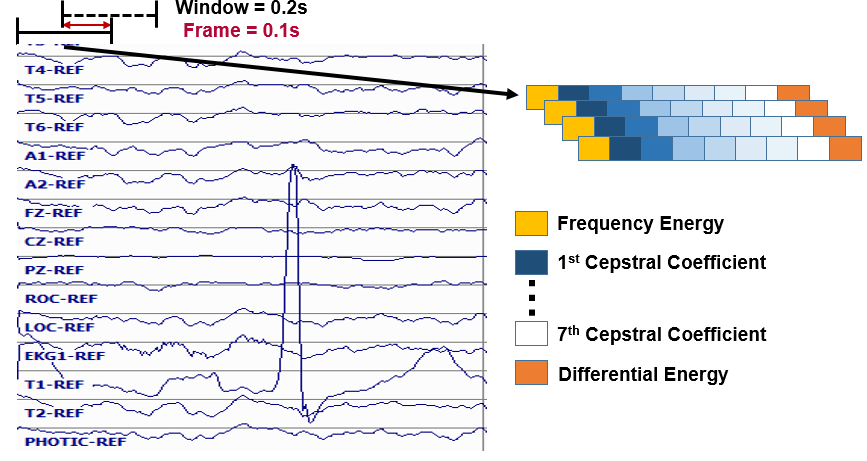
### 3.1.2 Feature Extraction

Feature extraction was performed on the EEG data in a pre-processing step. The feature extraction approach followed techniques that are similar to the ones based on Mel Frequency Cepstral Coefficients (MFCCs) that have been used for speech recognition (Picone, 1990). MFCCs are normally calculated through the computation of a high resolution Fast Fourier Transform and down-sampling the results with an oversampling approach that uses overlapping bandpass filters. The results obtained from this process are then transformed to the cepstral domain through a cosine transform (Huang, Acero, & Hon, 2001). When extracting the cepstral coefficients from EEG signals, a very similar approach is followed, with the exception that the filterbanks are linearly spaced, rather than Mel-spaced, as they usually are for speech recognition approaches. The first eight cepstral coefficients were kept, and the rest were discarded. Following the cepstral coefficients extraction, the frequency energy of the signal was calculated and used to replace the 0th cepstral coefficient. The calculation of the frequency energy is given by:

**(9)**

It is important to note that the frame and window duration for this portion of the feature extraction is 0.1 seconds and 0.2 seconds respectively (A. Harati, Golmohammadi, Lopez, Obeid, & Picone, 2015).

The extraction of the frequency energy and the cepstral coefficients is followed by the calculation of another type of energy: the differential energy (). The differential energy is a feature derived from the features that have been described to this point, and it is given by the difference between the largest and the smallest sample in a 0.9 seconds window. This feature is described as:



**Figure 8.** Illustration of the base feature extraction process.

**(10)**

Here, represents the number of frames. **Figure 8** shows an illustration of the feature extraction process that has been explained up to this point.

The first and second derivatives (differential and acceleration coefficients) of the base features explained to this point, are computed. These features represent the trajectory of the base features (Huang et al., 2001), and are calculated as follows:

**(11)**

In other words, represents a delta coefficient calculated for frame in terms of the static coefficients to . Similar to the calculation of the feature, the window used for the first and second derivatives is set to 0.9 seconds in this study. (A. Harati et al., 2015)

In summary, the features that are extracted from the EEG signals are the frequency energy (1 feature), 7 cepstral coefficients (7 features), a differential energy term (1 feature), and the first and second derivatives (18 features) of the base features. In this sense, each feature vector for each frame of data in every channel is represented by a 27 dimensional feature vector.

### 3.1.3 Dimensionality Reduction

As it was mentioned before, only one channel was utilized for the experiments. Only the beginning of the recording was used for the abnormal identifications if the EEGs. This experimental paradigm was majorly based on the fact that neurologists are reportedly able to distinguish a normal recording from a normal one by looking at the first few seconds of the files. To establish a baseline for the project, the first 60 seconds of the signal were considered, and the features across all the frames corresponding to this time were stacked together, forming a 16,200 () (Lopez et al., 2015). The dimensionality of the mentioned feature vectors was reduced, and only the most relevant eigenvectors of the covariance matrix for each class were retained (Jolliffe, 2002).

## 3.2 Experiments

The experiments are separated in two different sections: the establishment of a baseline through standard machine learning techniques, and the implementation and optimization of different GMM-HMM systems in order to solve the problem. This subsection goes over the details for the setup and execution of said experiments.

### 3.2.1 Baseline Experiments

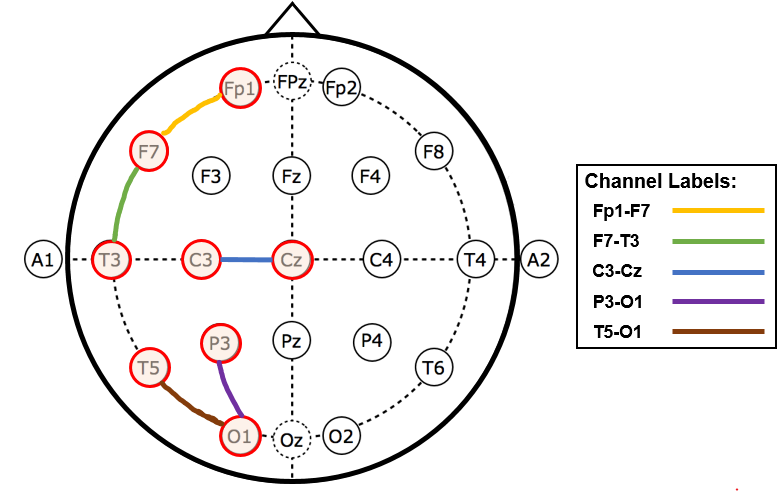
In order to establish a baseline for the problem, two standard machine learning algorithms were implemented and studied: k-Nearest Neighbors (kNN) and Random Forest Ensemble Learning (RF). In essence, the k-nearest neighbor algorithm assigns a class to an observation based on the distance measurements, Mahalanobis distance, in this case, between the observation and its k-nearest neighbors (Duda et al., 2001). While the Random Forest Ensemble Learning, makes a classification decision by considering the decisions made by all of the decision trees in the ensemble of trees (forest), and picking the class that received the majority of the votes.

The first set of experiments involved the variation of the dimensions of the features, in order to find the optimal dimension for the feature vectors. The following step was the individual optimization of the models by varying parameters specific to the algorithms. For kNN the number of nearest neighbors () was varied from 1 to 100 and for RF, the number of trees () was studied for values that ranged from 1 to 100. Once the parameters were properly optimized, a study about the relevance of the different channels for the normal/abnormal problem was conducted. Basically, the optimized systems were tested for all the 22 channels in the transverse central parietal (TCP) montage (ACNS, 2006), which accentuates spike activity, and the best performance was selected.

The results of these experiments helped to establish a baseline for the classification of normal and abnormal adult EEGs. The performance reported for the baselines was then compared to an HMM system, which was implemented with the hopes that the nature of the model and its reported usefulness on sequential data could help decrease the false alarm and detection rates.

### 3.1.2 HMM Experiments

Several experiments were conducted in order to optimize the HMM and find the proper number of Gaussian mixtures. The first step for these experiments was to find the optimal number of Gaussian mixtures and states for the HMM by running classification experiments with the full set of features and the first 10 minutes for each file. The features that were reduced through (PCA) were then used to test this model for a better comparison with the baseline. Once the system’s parameters were properly optimized and a pertinent comparison had been done with the baseline, the models were used to find the optimal amount of input time for the signal by varying the input time from 5 to 25 in steps of 5 minutes. Finally, the fully optimized model was implemented for different channels across the scalp. **Figure 9** shows the spatial information for the channels that were selected for the comparative study. It can be seen that approximately all the scalp regions (left side only, since symmetry is assumed) are represented by the channels selected for the study.



**Figure 9.** Location of studied channels in the 10-20 standard system of electrode placement for the TCP montage

The optimization for the HMM parameters was performed with the channel that showed the best performance for the baseline systems. The channel test with the HMM was conducted in order to verify whether the channel optimization could be generalized across the tested models.

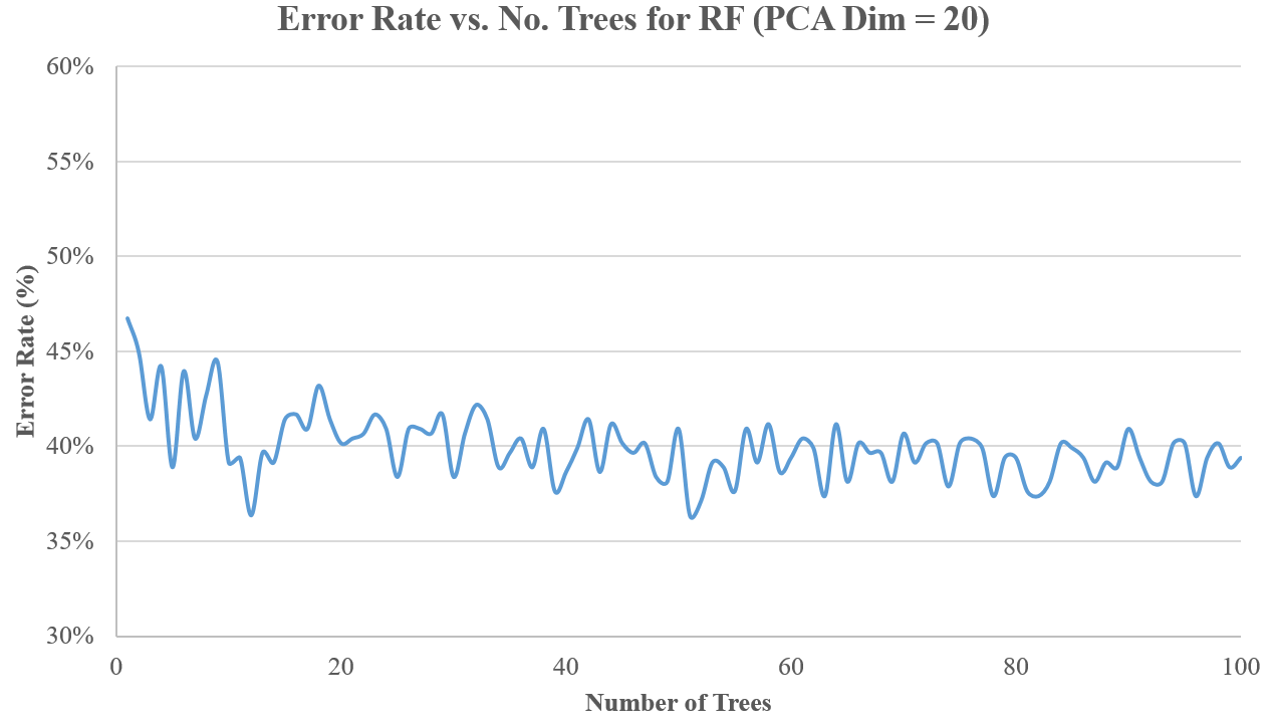
# CHAPTER 4

# PRELIMINARY EXPERIMENTS AND RESULTS

This section presents the results of the experiments that were described in Chapter 3. First, the results for the baseline experiments, which involved the classification of the normal and abnormal EEGs with the kNN and the RF algorithms, are presented. The last part of this chapter, presents the preliminary results that were obtained through HMMs.

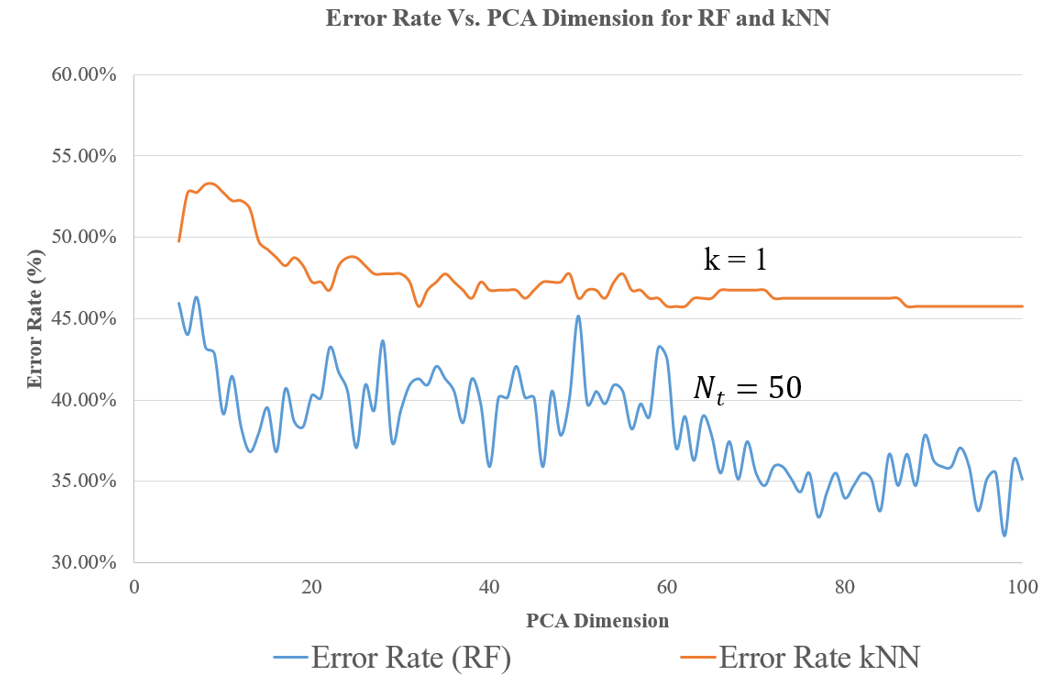
## 4.1 Baseline Results

The first step for the baseline experimental section was the selection of an optimal number of trees (). The error rate for the normal/abnormal classification was computed as a function of . **Figure 10** shows that the performance seems to saturate for . This observation motivated the selection of in a compromise for the tradeoff between training time and performance (Lopez et al., 2015).



**Figure 10.** Normal/abnormal classification error rate as a function of number of (trees )

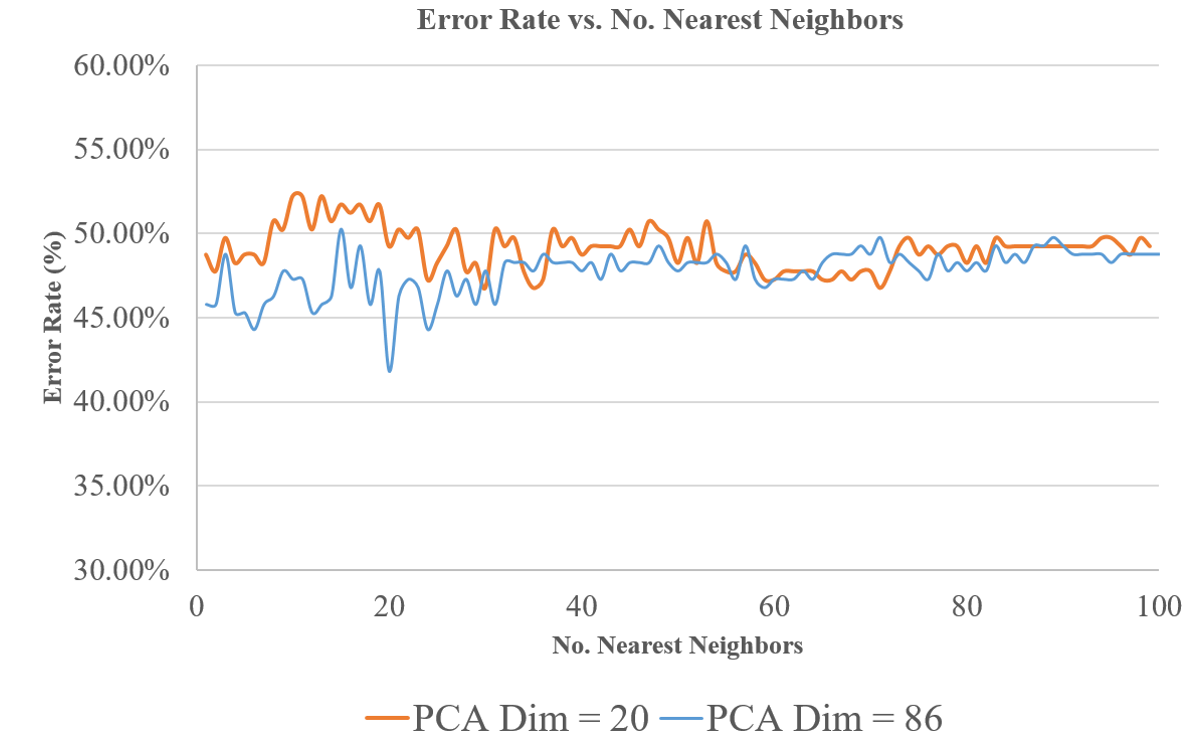
As it was discussed in chapter 3, the following experiments involved the study of the performance of both systems, kNN and RF, as a function of the PCA dimension of the features. **Figure 11** shows the results generated for each algorithm while varying the PCA dimension from 0 to 100. The previously selected number of trees was used for the RF implementation, while a value of was used for kNN (Lopez et al., 2015). From **Figure 11** it can be inferred that, for kNN, the performance is not heavily impacted by PCA values larger than 20, while for RF, the trend of the error seems to decrease up to the point where the PCA dimension is 86. Taking this information into account, 86 was selected to be the input dimension.



**Figure 11.** Error rate of the kNN and RF algorithms as a function of the PCA dimension

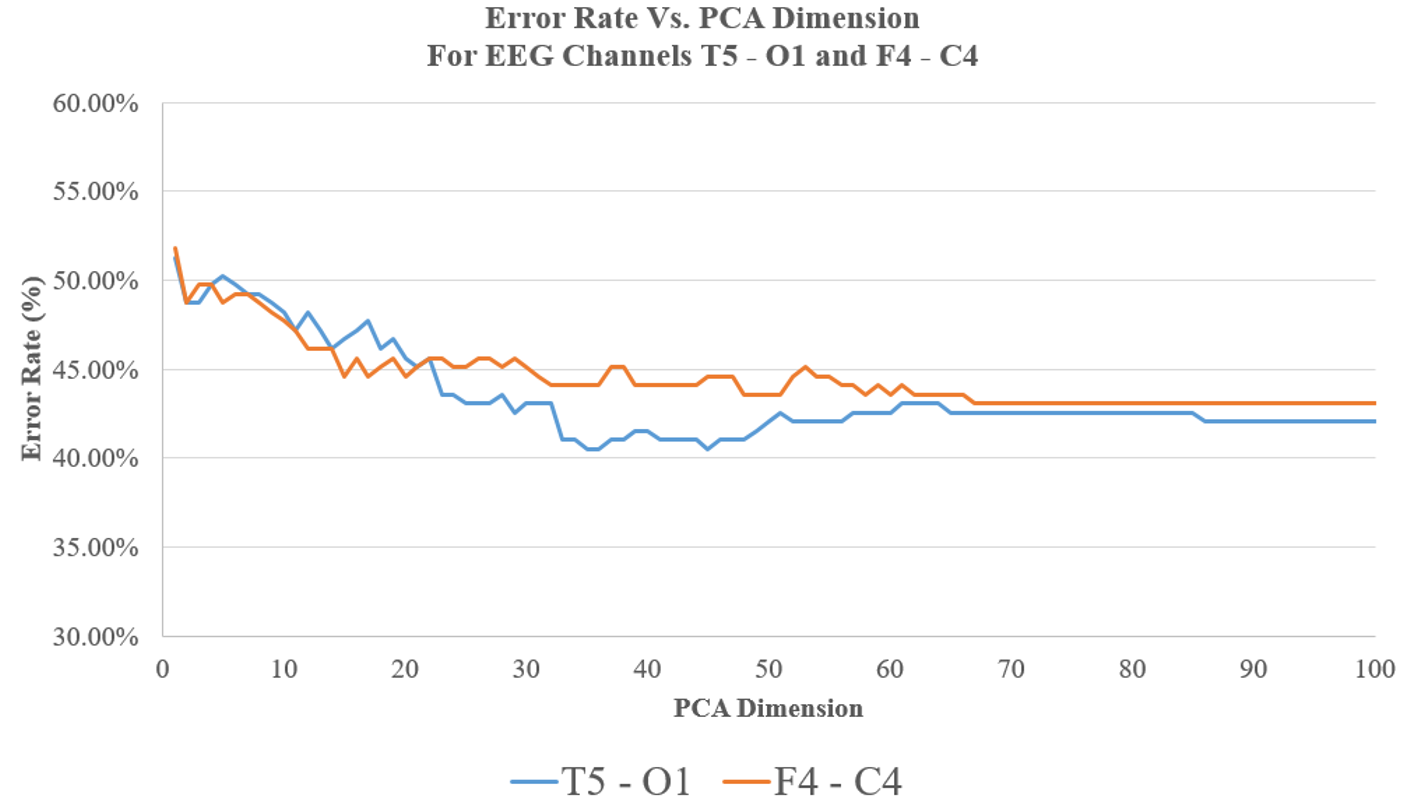
Considering the fact that the results obtained through kNN did not change much for PCA dimensions greater than 20, the optimization for the number of nearest neighbors () was studied separately for features that were reduced to a dimension of 20, and features that were reduced to a dimension of 86. **Figure 12** shows the results that were obtained through this experiment, and indicates that the performance has high variability, but considering its trend, the value of that shows a reasonable performance in comparison with other values is .

Finally, the optimized models were used to conduct an analysis about the channel relevance for the classification. The 22 channels of the TCP montage were individually used for the classification. In this way, it was possible to understand which channels (and what regions of the scalp) contribute the most for the identification of abnormal EEGs. **Figure 13** shows the most relevant results from this study. This analysis was performed through the implementation of the of the optimized kNN system (), since it presented less variance than the RF implementation (Lopez et al., 2015).



**Figure 12.** Errorrate as a function of the number of neighbors for PCA dimension of 20and 86

**.**



**Figure 13.** Classification error rate (for kNN) for a fronto-central (F4-C4) and a temporal-occipital (T5-O1) channel

The performance reported on **Figure 13** was shown for the channel that performed worse (F4-C4) and the one that showed the best performance (T5-O1). This observation is consistent with the way in which neurologists interpret the EEGs, which heavily involves the identification of abnormalities (slowing, lack of reactivity) in the posterior dominant rhythm, present in the posterior regions of the scalp (posterior, occipital-temporal channels in the TCP montage).

To summarize, the performance of the optimized version of both systems for input feature vectors of dimension 86 is presented in **Table 2**. The RF algorithm showed considerably higher variance than the kNN system. In order to obtain a better comparison with the HMM system, a confusion matrix for kNN is also shown in **Table 3**.

|  |  |  |
| --- | --- | --- |
| **No.** | **System Description** | **Error** |
| 1 | kNN () | 41.80% |
| 2 | RF () | 31.70% |

**Table 2.** Comparison of the performance obtained with the two baseline systems

|  |  |  |
| --- | --- | --- |
| **Ref/Hyp** | **Normal** | **Abnormal** |
| **Normal** | 50.50% | 49.50% |
| **Abnormal** | 34.00% | 66.00% |

**Table 3.** Confusion matrix for the kNN system

The results for the optimized baselines were then compared with a GMM-HMM system. The optimization details and performance of this system are presented in next section.

## 4.1 GMM-HMM Results

The optimization of the GMM-HMM system for this classification problem involves the selection of parameters such as the number of Gaussian mixtures and the number of HMM states. In order to do this, the first 10 minutes of data (features) for the T5-O1 channel were used as an input to the system. **Table 4** shows the summary of the results that were obtained through the evaluation of a number of system parameters. The closed loop performance for the best system (, ) reached a correct detection rate of 86.420%.

The information shown in **Table 4** shows that the optimal classification is obtained when the number of Gaussian mixtures is 3 and the number of HMM states is also 3 ().

|  |  |  |
| --- | --- | --- |
| **# Gaussian Mixtures** | **# HMM States** | **Correct Detection (%)** |
| 1 | 1 | 69.81% |
| 1 | 2 | 65.09% |
| 1 | 3 | 65.09% |
| 2 | 1 | 76.42% |
| 2 | 2 | 80.19% |
| 2 | 3 | 77.36% |
| 3 | 1 | 76.42% |
| 3 | 2 | 82.08% |
| **3** | **3** | **83.02%** |
| 4 | 1 | 82.08% |
| 4 | 2 | 64.15% |
| 4 | 3 | 77.36% |

**Table 4.** GMM-HMM correct detection rate for various numbers of Gaussian

To understand how much signal information would work better for the identification of abnormal EEGs, the optimized system was used to process different input lengths. **Table 5**, shows this analysis, and reveals that the best performance can be obtained for an input time of 10 minutes. The length of the majority of the recordings in the dataset are less than 25 minutes, so the performance saturates for durations longer than 25 minutes.

|  |  |  |
| --- | --- | --- |
| **Input (min)** | **#Gaussians/#HMM States** | **Correct Detection (%)** |
| 5 | 3/3 | 80.19% |
| **10** | **3/3** | **83.02%** |
| 15 | 3/3 | 80.19% |
| 20 | 3/3 | 79.25% |
| 25 | 3/3 | 76.42% |

**Table 5.** Correct detection rate for different signal input lengths

So far, the results that were presented were calculated with data from the T5-O1 channel, which was found to be optimal for the baseline systems. To make sure the channel selection could be generalized for the different systems an analysis was ran for several channels. **Table 6** shows the results of these experiments.

It is possible to observe that the channel that performed best for the GMM-HMM system is the same that was discovered through the baseline systems. In this sense, it can be said that this temporal-occipital channel has great relevance in the classification of abnormal EEGs.

|  |  |  |  |
| --- | --- | --- | --- |
| **Input (min)** | **#Gaussians/#HMM States** | **Channel** | **Correct Detection (%)** |
| 5 | 3/3 | Fp1-F7 | 80.19% |
| **10** | **3/3** | T5-O1 | **83.02%** |
| 15 | 3/3 | F7-T3 | 80.19% |
| 20 | 3/3 | C3-Cz | 79.25% |
| 25 | 3/3 | P3-O1 | 76.42% |

**Table 6.** Correct detection rate for different channels

The results that have been presented to this point, can be further summarized and compared to the baseline performance. **Table 7** shows the results of this comparison. The PCA-HMM experiment used the same exact inputs that were used for the baseline systems (for comparison) and the GMM-HMM classification system was use as the back-end. It can be seen that the best performance was achieved by both of the HMM systems, with the full feature system having the lowest overall error rate. **Table 8** shows the confusion matrix for the best reported system. It can be seen that the GMM-HMM system showed an improvement of 27.7% compared to the false alarm rate of the baseline kNN system.

|  |  |  |
| --- | --- | --- |
| **Ref/Hyp** | **Normal** | **Abnormal** |
| **Normal** | 78.18% | 21.82% |
| **Abnormal** | 11.76% | 88.24% |

**Table 8.** Confusion matrix for the best GMM-HMM system

|  |  |
| --- | --- |
| **System Description** | **Error (%)** |
| kNN (k=20) | 41.80% |
| RF (Nt=50) | 31.70% |
| PCA-HMM #GM = 3 #HMM States = 3) | 25.64% |
| **GMM-HMM (#GM = 3 #HMM States = 3)** | **16.98%** |

**Table 7.** Summary of the performance for all the evaluated systems

# CHAPTER 5

# EXPECTATIONS AND FUTURE WORK

## 5.1 Expected Outcomes

The classification experiments that have been presented in this study have shown that HMMs can be used in the abnormal identification of EEG signals, achieving performance that surpasses that of the other classification algorithms. Similarly, as it was discussed in Chapter 2, HMMs have been very successful in the area of automatic speech recognition and have recently been outperformed by deep learning models trained in sufficiently large amounts of data.

The main goal for the remaining research on this topic involves the implementation of deep learning techniques in order to decrease the reported false alarm rate and increase the sensitivity of the system, as has been done in the speech domain. In addition, given previous research, this task would involve the expansion and validation of a larger normal/abnormal database. The implementation of a deep learning system would allow to integrate the spatial and temporal context in order to take advantage of certain aspects of the domain knowledge, such as the fact that when the eyes are closed the PDR emerges, and improve the classification error rates.

## Timeline for Future Work

The timeline for the expected completion dates for the tasks that are necessary for the completion of this study are outlined as follows:

**December-January:**

1. Set up deep learning system for a second pass of deep learning after the GMM-HMM processing:

Implement and optimize a Stacked Denoising Autoencoders (SdA) system for the classification and increase the number of channels that are taken into account for the classification decision.

1. Expand and evaluate the normal/abnormal TUH database subset:

Generate simple natural language processing (NLP) scripts to obtain EEG sessions that have been evaluated and classified by neurologists and form a larger, demographically balanced, subset of the data.

**February**

1. Implement a long short term memory system for the normal/abnormal classification of EEGs in order to compare to the HMM-SdA implementation.

This system will be implemented with the Theano Python library for deep learning and evaluated in the expanded dataset.

1. Evaluate the SdA implementation on the expanded dataset.

Get results for the SdA implementation with the expanded dataset.

**March-May**

1. Complete the writing of the thesis and work on publications.
2. Defend this thesis.

# REFERENCES

ACNS. (2006). *Guideline 6: A Proposal for Standard Montages to Be Used in Clinical EEG [White Paper]. Retrieved from*. Retrieved from http://www.acns.org/pdf/guidelines/Guideline-6.pdf

Azuma, H., Hori, S., Nakanishi, M., Fujimoto, S., Ichikawa, N., & Furukawa, T. A. (2003). An intervention to improve the interrater reliability of clinical EEG interpretations. *Psychiatry and Clinical Neurosciences*, *57*(5), 485–489. http://doi.org/10.1046/j.1440-1819.2003.01152.x

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics*, *41*(1), 164–171. http://doi.org/10.1214/aoms/1177697196

Bishop, C. (2011). *Pattern Recognition and Machine Learning* (2nd ed.). New York, New York, USA: Springer. Retrieved from http://www.amazon.com/Pattern-Recognition-Learning-Information-Statistics/dp/0387310738/ref=zg\_bs\_3894\_1

Cao, C., Tutwiler, R. L., & Slobounov, S. (2008). Automatic classification of athletes with residual functional deficits following concussion by means of EEG signal using support vector machine. *IEEE Transactions on Neural Systems & Rehabilitation Engineering*, *16*(4), 327–335. http://doi.org/10.1109/TNSRE.2008.918422

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification*. *New York: John Wiley, Section*. http://doi.org/10.1007/BF01237942

Ebersole, J. S., & Pedley, T. A. (2014). *Current practice of clinical electroencephalography* (4th ed.). Philadelphia, Pennsylvania, USA: Wolters Kluwer. Retrieved from http://www.amazon.com/Current-Practice-Clinical-Electroencephalography-Ebersole/dp/145113195X

Finnigan, S., & van Putten, M. J. a M. (2013). EEG in ischaemic stroke: quantitative EEG can uniquely inform (sub-)acute prognoses and clinical management. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *124*(1), 10–19. http://doi.org/10.1016/j.clinph.2012.07.003

Gales, M., & Young, S. (2007). The Application of Hidden Markov Models in Speech Recognition. *Foundations and Trends® in Signal Processing*, *1*(3), 195–304. http://doi.org/10.1561/2000000004

Harati, A., Golmohammadi, M., Lopez, S., Obeid, I., & Picone, J. (2015). Improved EEG event classification using differential energy. In *2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)* (pp. 1–4). Philadelphia: IEEE. http://doi.org/10.1109/SPMB.2015.7405421

Harati, A., Lopez, S., Obeid, I., Jacobson, M., Tobochnik, S., & Picone, J. (2014a). The TUH EEG Corpus: A Big Data Resource for Automated EEG Interpretation. In *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium* (pp. 1–5). Philadelphia, Pennsylvania, USA. http://doi.org/http://www.isip.piconepress.com/publications/conference\_proceedings/2014/ieee\_spmb/tuh\_eeg

Harati, A., Lopez, S., Obeid, I., Jacobson, M., Tobochnik, S., & Picone, J. (2014b). THE TUH EEG CORPUS: A Big Data Resource for Automated EEG Interpretation. In *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium* (pp. 1–5). Philadelphia, Pennsylvania, USA. http://doi.org/http://www.isip.piconepress.com/publications/conference\_proceedings/2014/ieee\_spmb/tuh\_eeg

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., … Kingsbury, B. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine*, *29*(6), 82–97. http://doi.org/10.1109/MSP.2012.2205597

Huang, X., Acero, A., & Hon, H.-W. (2001). *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. Upper Saddle River, New Jersey, USA: Prentice Hall. Retrieved from http://www.amazon.com/Spoken-Language-Processing-Algorithm-Development/dp/0130226165

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). New York City, New York, USA: Springer-Verlag. Retrieved from https://www.springer.com/us/book/9780387954424

Lodder, S. S., & van Putten, M. J. A. M. (2013). Quantification of the adult EEG background pattern. *Clinical Neurophysiology*, *124*(2), 228–237. http://doi.org/10.1016/j.clinph.2012.07.007

Lopez, S., Suarez, G., Jungries, D., Obeid, I., & Picone, J. (2015). Automated Identification of Abnormal EEGs. In *IEEE Signal Processing in Medicine and Biology Symposium* (pp. 1–4). Philadelphia, Pennsylvania, USA. http://doi.org/http://www.isip.piconepress.com/publications/conference\_proceedings/2015/ieee\_spmb/abnormal/

Picone, J. (1990). Continuous Speech Recognition Using Hidden Markov Models. *IEEE ASSP Magazine*, *7*(3), 26–41. http://doi.org/10.1109/53.54527

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., … Vesely, K. (2011). The Kaldi Speech Recognition Toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society.

Rabiner, L. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, *77*(2), 257–286. http://doi.org/10.1109/5.18626

Smith, S. (2005). EEG in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery, and Psychiatry*, *76*(Suppl 2), ii2-ii7. http://doi.org/10.1136/jnnp.2005.069245

Steve Young. (1996). A review of large-vocabulary continuous-speech. *IEEE Signal Processing Magazine*, *13*(5), 45. http://doi.org/10.1109/79.536824