



FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

Niina Mäkinen

Deep Learning Approach for Epileptic Seizure Detection

Master's Thesis
Degree Programme in Biomedical Engineering
June 2020

Mäkinen N. (2020) Deep Learning Approach for Epileptic Seizure Detection. University of Oulu, Degree Programme in Biomedical Engineering. Master's Thesis, 104 p.

ABSTRACT

Epilepsy is the most common brain disorder that affects approximately fifty million people worldwide, according to the World Health Organization. The diagnosis of epilepsy relies on manual inspection of EEG, which is error-prone and time-consuming. Automated epileptic seizure detection of EEG signal can reduce the diagnosis time and facilitate targeting of treatment for patients. Current detection approaches mainly rely on the features that are designed manually by domain experts. The features are inflexible for the detection of a variety of complex patterns in a large amount of EEG data. Moreover, the EEG is non-stationary signal and seizure patterns vary across patients and recording sessions. EEG data always contain numerous noise types that negatively affect the detection accuracy of epileptic seizures. To address these challenges deep learning approaches are examined in this paper.

Deep learning methods were applied to a large publicly available dataset, the Children's Hospital of Boston-Massachusetts Institute of Technology dataset (CHB-MIT). The present study includes three experimental groups that are grouped based on the pre-processing steps. The experimental groups contain 3-4 experiments that differ between their objectives. The time-series EEG data is first pre-processed by certain filters and normalization techniques, and then the pre-processed signal was segmented into a sequence of non-overlapping epochs. Second, time series data were transformed into different representations of input signals. In this study time-series EEG signal, magnitude spectrograms, 1D-FFT, 2D-FFT, 2D-FFT magnitude spectrum and 2D-FFT phase spectrum were investigated and compared with each other. Third, time-domain or frequency-domain signals were used separately as a representation of input data of VGG or DenseNet 1D.

The best result was achieved with magnitude spectrograms used as representation of input data in VGG model: accuracy of 0.98, sensitivity of 0.71 and specificity of 0.998 with subject dependent data.

VGG along with magnitude spectrograms produced promising results for building personalized epileptic seizure detector. There was not enough data for VGG and DenseNet 1D to build subject-dependent classifier.

Keywords: Electroencephalogram (EEG), Epilepsy, Seizure detection, Deep learning, raw EEG, 1D-FFT, 2D-FFT, magnitude spectrogram, DenseNet 1D, VGG

Mäkinen N. (2020) Epileptisten kohtausten havaitseminen syväoppimisella lähestymistavalla. Oulun yliopisto, lääketieteen tekniikan tutkinto-ohjelma. Diplomityö, 104 s.

TIIVISTELMÄ

Epilepsia on yleisin aivosairaus, joka Maailman terveysjärjestön mukaan vaikuttaa noin viiteenkymmeneen miljoonaan ihmiseen maailmanlaajuisesti. Epilepsian diagnosointi perustuu EEG:n manuaaliseen tarkastamiseen, mikä on virheeltistä ja aikaa vievää. Automaattinen epileptisten kohtausten havaitseminen EEG-signaalista voi potentiaalisesti vähentää diagnoosiaikaa ja helpottaa potilaan hoidon kohdentamista. Nykyiset tunnistusmenetelmät tukeutuvat pääasiassa piirteisiin, jotka asiantuntijat ovat määritelleet manuaalisesti, mutta ne ovat joustamattomia monimutkaisten ilmiöiden havaitsemiseksi suuresta määrästä EEG-dataa. Lisäksi, EEG on epästationäärinen signaali ja kohtauspiirteet vaihtelevat potilaiden ja tallennusten välillä ja EEG-data sisältää aina useita kohinatyyppisiä, jotka huonontavat epilepsiakohtauksen havaitsemisen tarkkuutta. Näihin haasteisiin vastaamiseksi tässä diplomityössä tarkastellaan soveltuvatko syväoppivat menetelmät epilepsian havaitsemiseen EEG-tallenteista.

Aineistona käytettiin suurta julkisesti saatavilla olevaa Bostonin Massachusetts Institute of Technology lastenklinikan tietoaainestoa (CHB-MIT). Tämän työn tutkimus sisältää kolme koeryhmää, jotka eroavat toisistaan esikäsittelyvaiheiden osalta: aikasarja-EEG-data esikäsiteltiin perinteisten suodattimien ja normalisointitekniikoiden avulla, ja näin esikäsitelty signaali segmentoitui epookkeihin. Kukin koeryhmä sisältää 3-4 koetta, jotka eroavat menetelmiltään ja tavoitteiltaan. Kussakin niistä epookkeihin jaettu aikasarjadata muutettiin syötesignaalien erilaisiksi esitysmuodoiksi. Tässä tutkimuksessa tutkittiin ja verrattiin keskenään EEG-signaalia sellaisenaan, EEG-signaalin amplitudi-spektrogrammeja, 1D-FFT-, 2D-FFT-, 2D-FFT-amplitudi- ja 2D-FFT -vaihespektriä. Näin saatuja aika- ja taajuusalueen signaaleja käytettiin erikseen VGG- tai DenseNet 1D -mallien syötetietoina.

Paras tulos saatiin VGG-mallilla kun syötetietona oli amplitudi-spektrogrammi ja tällöin tarkkuus oli 0,98, herkkyys 0,71 ja spesifisyys 0,99 henkilöstä riippuvaisella EEG-datalla.

VGG yhdessä amplitudi-spektrogrammien kanssa tuottivat lupaavia tuloksia henkilökohtaisen epilepsiakohtausdetektorin rakentamiselle. VGG- ja DenseNet 1D -malleille ei ollut tarpeeksi EEG-dataa henkilöstä riippumattoman luokittelijan opettamiseksi.

Avainsanat: Electroencephalogrammi (EEG), Epilepsia, Kohtauksen tunnistaminen, Syväoppiminen, 1D-FFT, 2D-FFT, amplitudi-spektrogrammi, DenseNet 1D, VGG

TABLE OF CONTENTS

ABSTRACT

TIIIVISTELMÄ

TABLE OF CONTENTS

FOREWORD

ABBREVIATIONS

1.	INTRODUCTION.....	8
2.	LITERATURE REVIEW.....	9
2.1.	Epilepsy and EEG	9
2.1.1.	Definition of epilepsy and epileptic seizure	9
2.1.2.	Etiology of epilepsy	9
2.1.3.	Seizure classification.....	10
2.1.4.	Status epilepticus.....	13
2.1.5.	Origin of EEG	15
2.1.6.	EEG registration and activations.....	16
2.1.7.	EEG in diagnostics of epilepsy and status epilepticus	18
2.1.8.	Interpretation of EEG	20
2.1.9.	Normal EEG	21
2.1.10.	Epileptiform abnormalities.....	23
2.1.11.	EEG artifacts	27
2.2.	Deep learning	28
2.2.1.	Overview	28
2.2.2.	Artificial neural networks.....	29
2.2.3.	Categorization of deep neural networks.....	30
2.2.4.	Architectures	32
2.2.5.	Tuning deep neural networks	37
2.2.1.	Deep learning workflow in signal processing	38
3.	RELATED WORK.....	40
4.	METHODS.....	42
4.1.	Designed system.....	42
4.2.	Input formats	44
4.3.	Used CNN architectures.....	46
4.4.	Bayesian optimization	48
4.5.	Validation methods and evaluation metrics	48
5.	IMPLEMENTATION	53
5.1.	Workflow.....	53
5.1.1.	Data preparation	53
5.1.2.	Feature engineering	54
5.1.1.	Model selection	56
5.1.2.	Hyperparameter tuning.....	57
5.2.	Experiments.....	58
5.2.1.	Experimental group 1	59
5.2.2.	Experimental group 2	60
5.2.3.	Experimental group 3	61
6.	RESULTS.....	63
6.1.	Experimental group 1	63
6.2.	Experimental group 2	64

6.3.	Experimental group 3	65
7.	DISCUSSION	66
8.	CONCLUSION	70
9.	REFERENCES	72
10.	APPENDICES	90

FOREWORD

This work was carried out during the year 2019 as a master's thesis of Biomedical engineering at the department of computer science and engineering, University of Oulu, Finland.

I would like to express my sincere gratitude to my supervisor, Prof. Tapio Seppänen for his support and guidance throughout this project. I want to thank my another supervisor, PhD Antti Ruha whose support and guidance were crucial during difficult times. I would like to thank MD, PhD Hanna Ansakorpi and MD, PhD Erkki Liimatta for the time and effort to pre-reviewing my thesis. I also want to thank my colleagues Markus Nykänen and Ossi Kumpula for peer support and insights in technical aspects. Especially, I want to thank Markus Nykänen for a solid implementation of DenseNet 1D that was used as one neural network solution in this study.

For the financial support, I am grateful to my workplace Bittium for this possibility to make research on this difficult but interesting topic.

Finally, I would like to thank my family and friends, especially Camilla, for the support and encouragement.

Oulu, 5.5.2020

Niina Mäkinen

ABBREVIATIONS

ANN	Artificial neural network
AUC	Area under the curve
BiPEDs	Bilateral periodic epileptiform discharges
BIPLEDs	Bilateral independent periodic epileptiform discharges
CNN	Convolutional neural network
DBN	Deep belief network
DBM	Deep Boltzmann machine
DBS	Deep brain stimulation
ECG	Electrocardiogram
ED	Ictal epileptiform discharge
ER	Error rate
EEG	Electroencephalogram
EMG	Electromyography
EOG	Electrooculography
FFT	Fast Fourier transform
FN	False negatives
FNR	False negative rate
FP	False positives
FPR	False positive rate
GPFA	Generalized paroxysmal fast activity
GSW	Generalized spikes and slow-wave complexes
IED	Interictal epileptiform discharge
ILAE	International League Against Epilepsy
IMSD	Independent multifocal spike discharges
LSTM	Long short term memory
MISD	Multifocal independent spike discharges
MLPs	Multi-Layer Perceptrons
MRI	Magnetic resonance imaging
RNN	Recurrent neural network
PEDs	Periodic epileptiform discharges
PFA	Paroxysmal fast activity
PLEDs	Periodic lateralized epileptiform discharges
PR	Precision-recall
RBM	Restricted Boltzmann machine
ROC	Receiver operating characteristics curve
SE	Status epilepticus
TN	True negatives
TP	True positives
VNS	Vagus nerve stimulation
π	Pi
ϕ	Phase
i	Imaginary unit
C	Complex numbers

1. INTRODUCTION

Epilepsy is the most common serious neurological disorder. The number of people affected by epilepsy is three out of every thousand in worldwide. In several areas, the percentage of epilepsy patients in a population is about 4%, which means 40 people per thousand. Every year, among every 100 000 people there will be 40 – 70 new people who are affected by epilepsy. About 1% of the Finnish population are affected by epilepsy. 1% equals to 56 000 Finnish citizens and 5 000 of them are children. About 36 000 people need active drug treatment (about 0,7% of the population) and about 9 000 people are affected by severe epilepsy in Finland (about 20-25% of all epilepsy patients). People from all age groups can suffer from epilepsy, but the age group of over 65 years old has over doubled risk to get affected by epilepsy compared to 25-55 years old age group. Each year about 3 000 people start drug treatment for epilepsy and 800 of them are younger than 15 years old [1 p. 8, 2, 3].

Electroencephalogram (EEG) is a technique commonly used for monitoring the brain activity and diagnosis of epilepsy. EEG recordings are analyzed by neurologists to detect and identify epileptic seizures. The visual examination is a laborious and time-consuming task. Identification of epilepsy, minimization of delay in treatment and finding the optimal level of healthcare are critical aspects in the treatment of epilepsy [4]. There is a lot of research work carried out to automatically classify the epileptic and non-epileptic signals [5-9]. From the machine learning point of view, classification of epileptic and non-epileptic signals is a challenging task. The reason is a lack of available data for training a classifier. Moreover, the presence of noise, interference, and artefacts in EEG signal causes challenges in learning the brain patterns associated with seizure and non-seizure EEG signals [10]. Epilepsy is not a singular disease entity, but a variety of disorders caused by different brain dysfunctions [11]. Thus, there are variety of complex patterns of different seizure types. Furthermore, seizures may not only vary among the patients but also between seizures. [12]

Researches have proposed approaches for the detection of seizures using hand-engineered features extracted from EEG signals [6]. Deep learning is an approach that automatically extracts features from the data [13]. Automatically extracted features have shown promising results as compared to hand-engineered features [13]. Deep convolutional neural network variants such as AlexNet [14], VGG [15], ResNet [16] and DenseNet [17] have shown good performance in many fields. Moreover, CNN-LSTM [18], 3DCNN [19] and C3D [20] networks have shown significant results with time-domain data.

The aim of this work was to build a binary classifier for epileptic seizure detection to distinct seizure or non-seizure EEG. Specifically, time-series EEG signal, magnitude spectrograms, 1D-FFT, 2D-FFT, 2D-FFT magnitude spectrum and 2D-FFT phase spectrum were investigated as a representation of input data of a deep learning model. VGG and DenseNet 1D were examined as deep learning classifiers for automatic epileptic seizure detection. Deep learning methods were applied to a large publicly available dataset, the Children's Hospital of Boston-Massachusetts Institute of Technology dataset (CHB-MIT). Moreover, the trained classifiers were evaluated with subject dependent and subject dependent data.

2. LITERATURE REVIEW

2.1. Epilepsy and EEG

2.1.1. *Definition of epilepsy and epileptic seizure*

Epilepsy is defined as a chronic neurological brain disorder where a brain has a long-lasting abnormal tendency to cause epileptic seizures [1 p. 8, 3]. Epilepsy is not a singular disease entity, but a variety of disorders caused by different brain dysfunctions [11]. An epileptic seizure is defined as “*a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain*” [1 p. 8, 3, 11, 21, 22].

A diagnosis of epilepsy requires the occurrence of at least one epileptic seizure and abnormal change found in brain magnetic resonance imaging or in EEG recording [1 p. 9, 11]. The duration of epileptic seizure varies from a few seconds to several minutes [23 p.15]. A person that suffers from epilepsy tends to have problems with neurological, cognitive, psychological or social consequence of this condition [1 p. 8, 3, 11].

2.1.2. *Etiology of epilepsy*

Epilepsy is more than just spontaneously occurring seizures, epilepsy is a neurological disorder where a person has observable and abnormal long-lasting changes in a brain. Finding the cause of epilepsy is the most important part of diagnostics because it allows more efficient treatment for a brain disorder behind epilepsy. In some cases, epilepsy is genetic, but most often it is caused by brain damage, infections, trauma, stroke, brain tumor or development of abnormalities. Still, the majority of the causes of epilepsy stay unknown. [2, 23]

The etiology of epilepsy can be divided into genetic, structural, infectious, metabolic, immune and unknown causes (Figure 1.) [24]. Epilepsy caused by genetic causes is resulted directly from a known or presumed genetic mutation. More specifically, these genetic causes can be a chromosomal defect or a single genetic error. In most cases, underlying genes reflecting epilepsy are not yet known. A genetic does not mean the same than inheritable, because new mutations are constantly evolving. It is well known that environmental factors contribute to the development of epilepsy and therefore, a genetic etiology does not exclude the importance of environmental contributions. [23 p.16-18, 24]

In structural etiology, a structural abnormality has a substantially increasing risk to cause epilepsy. The underlying basis for a structural abnormality are genetic, acquired, or both. Acquired causes are such as stroke, trauma, infection or cerebrovascular disorder. However, in some cases structural abnormalities may be also associated with genetic mutations. Structural abnormalities can be visibly seen and examined via a brain neuroimaging. [23 p.19, 24]

Metabolic causes of epilepsy are metabolic state or metabolic disorder in a brain. The understanding of metabolic disorders of the brain and the association between metabolic abnormalities and epilepsy is constantly evolving. Recently it is known that metabolic epilepsy results directly from a known or presumed metabolic disorder. Nonetheless, often metabolic disorders are also genetic, but the factor that causes an epileptic seizure is a metabolic defect in the background. [23 p.16-18, 24]

Currently abnormal activity of immunological defense mechanism has been identified as one cause of epilepsy. Immune epilepsy results directly from an immune disorder. In an immune disorder, there is autoimmune-mediated central nervous system inflammation, which leads a defense mechanism of a body to attack against itself. The field of immune disorders is also emerging and causes rapid increase in diagnosis of these disorders. [23 p.18, 24]

The most common cause of the development of epilepsy is an infection. In infectious etiology epilepsy is directly resulted from a known infection in which seizures are a core symptom of the disorder. Bacteria, viruses, fungi and protozoans can cause infection in a brain and thus damage a brain. Eventually, a brain damage caused by infection can lead to the development of epilepsy. The development of epilepsy can occur during an acute or a post-acute phase. [23 p.18, 24, 25]

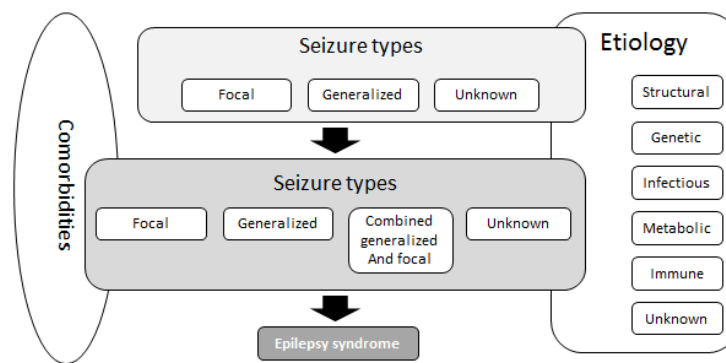


Figure 1. Framework for classification of the epilepsies.

2.1.3. Seizure classification

Understanding of epileptic seizure networks is developing constantly, but there is still a lot to do to serve sufficient seizure classification. In 1981, International League Against Epilepsy (ILAE) Commission conducted a study where hundreds of video-EEG recordings of seizures were evaluated and applied to develop recommendations that divided seizures into partial and generalized onset, simple and complex partial seizures, and various specific types. The classification remains as a base for seizure classification in widespread use. Previous seizure classifications have been based on anatomy, but recently have been shifted to the modern approach in which epilepsy is defined as a network disease. Network perspective allows seizures to be categorized in neocortical, thalamocortical, limbic, and brainstem networks. Recently the classification of seizures has been operational rather than based on fundamental mechanisms as previously. [12, 24, 26, 27]

Multiple variations of seizure classification exist for various purposes, but this document takes under consideration only clinical perspective based on EEG patterns and uses current epileptic seizure classification and terminology of ILAE [12] (Figure 2.). In clinical purposes, classification of seizure type is an important factor for diagnostics and treatment of epilepsy. The first task in diagnostics of epileptic syndrome is to determine that event has the characteristics of seizure and the second phase is to diagnose a seizure type. The classification of epileptic seizures is still

mostly based on neurologist's opinion and observation. It also needs to be noted that EEG cannot be the only asset to classify epileptic seizures. [12, 23, 24, 25, 28 p.22-25, 29].

The current seizure classification (Figure 2.) is based on the following features: where the seizure begins in the brain, level of awareness during a seizure and other features of seizures (Figure 3.). Defining where the seizure begins is an important factor because it affects the choice of seizure medication and possible surgery. Seizures defined by where the seizure begins are divided into focal, generalized, unknown onset and focal to bilateral seizures. Awareness in the beginning and during a seizure is divided into focal aware, focal impaired awareness, awareness not known and generalized seizures. Describing awareness is an important factor for a person's safety [30]. Motor symptoms in focal and general seizures are divided into motor and non-motor seizures. Sometimes the onset of a seizure is not known then the seizure is described as an unknown onset seizure. Unknown onset seizures are divided into motor and non-motor seizures. [12, 30]

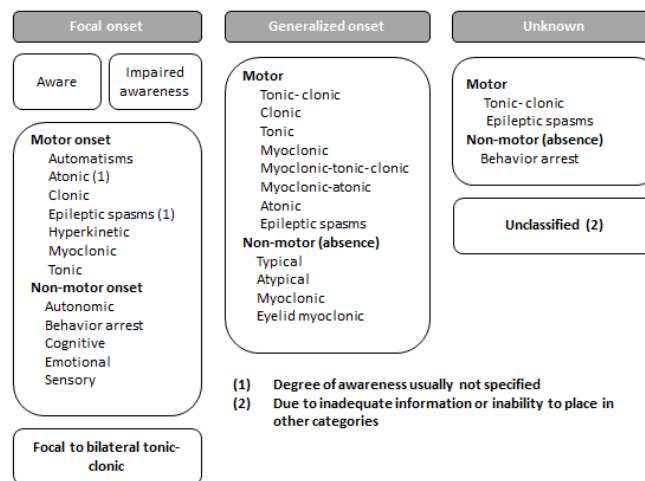


Figure 2. ILAE 2017 classification of seizure type expanded version.

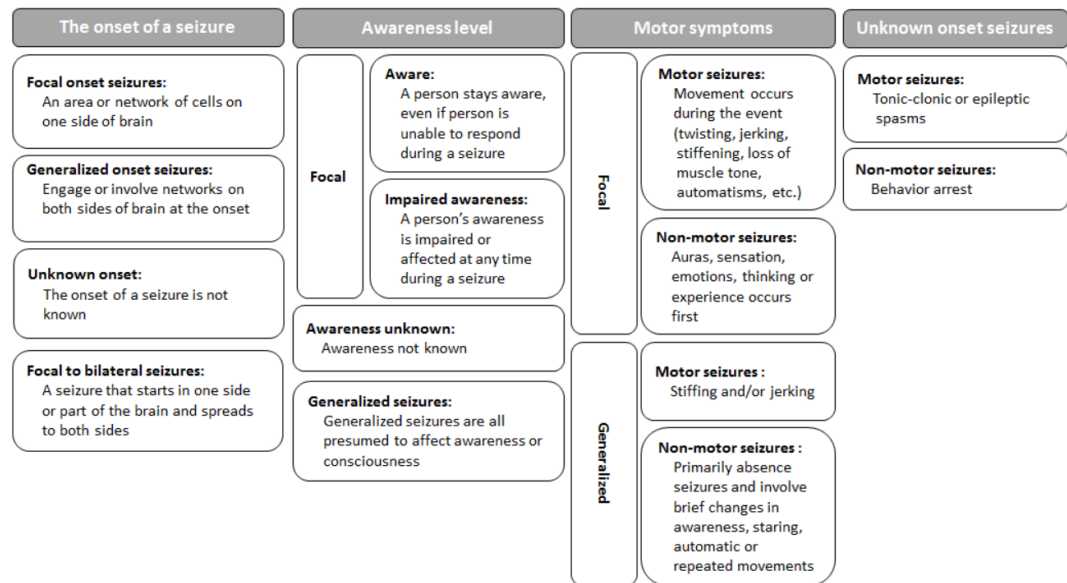


Figure 3. The symptom features of an epileptic seizure

Described features are the key factors for the seizure type classification. The seizure types based on the features are divided into focal, general and unknown types (Figure 2.). Epilepsy types that have the characteristics of generalized seizures are typically genetic and begin in childhood. The characteristic of general seizure type is that the electrical discharge starts from a local starting point and spreads to the wider area of a brain. General seizures engage networks on both sides of the brain at the onset. The electrical discharge may spread rapidly from one hemisphere to another via the corpus callosum. All generalized seizure types have an effect on awareness or consciousness. That is the reason why only motor and non-motor features are applied to divide generalized seizure types. Generalized motor seizures are divided into tonic-clonic, clonic, tonic, myoclonic, myoclonic-tonic-clonic, myoclonic-atonic, atonic and epileptic spasms seizure types. (Figure 2.) The primary feature of generalized motor seizures is stiffening and/or jerking of a body. The specific characteristics of each generalized motor seizure type are presented in Appendix 1. [12, 31, 32]

Generalized non-motor seizures are divided into typical absence, atypical absence, myoclonic absence and eyelid myoclonic seizure types. The primary feature of generalized non-motor seizures is absence. Absence involves brief changes in awareness, staring, automatic or repeated movements. The specific characteristics and common causes of each generalized non-motor seizure type are presented in Appendix 2. [12]

Epilepsy types that have focal seizure characteristics begin typically in adulthood. Focal seizures start in an area or network of cells on one side of the brain. Focal seizures may occur over any lobe or hemisphere, but are most commonly seen in either temporal or frontal lobe epilepsy. The location and the spreading area of the electrical discharge in a brain are related to symptoms that are caused by the seizure (Figure 4.). Focal seizures are divided into seizure types based on awareness and motor symptoms. Focal seizures based on awareness or consciousness is grouped into focal aware, focal impaired awareness and focal seizures that turn into bilateral

tonic-clonic seizure. The characteristics of focal seizures are presented in Appendix 3. [12, 23, 33, 34]

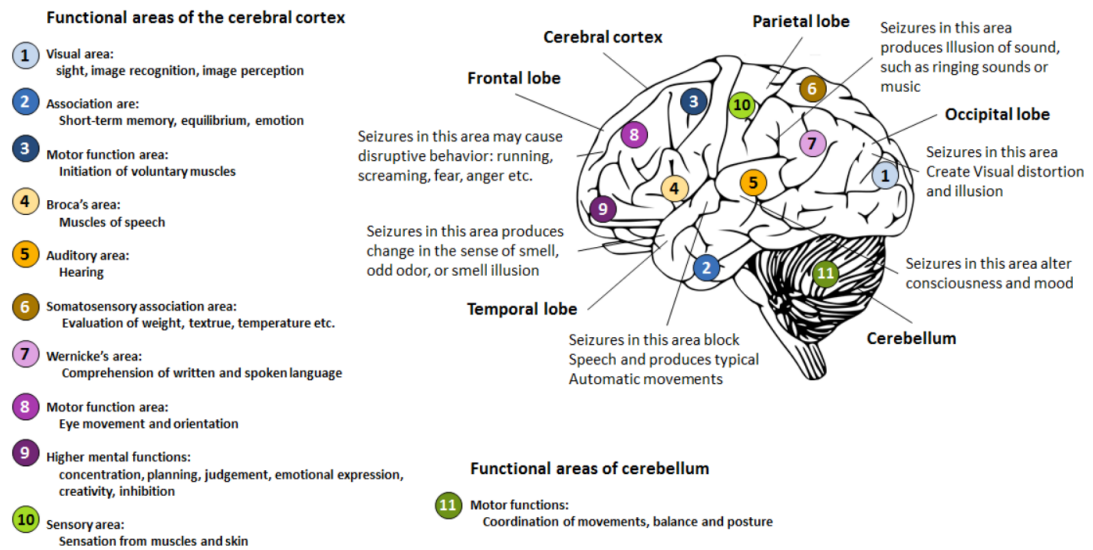


Figure 4. Functional areas of the cerebral cortex, the cerebellum and their relationship between focal seizures and symptoms

Focal motor seizures are divided into automatisms, atonic, clonic, epileptic spasms, hyperkinetic, myoclonic and tonic seizure types (Figure 2.). The primary features of focal motor seizures are stiffness or jerking of body, loss of muscle tone or automatisms. The characteristics of focal motor seizures are described in Appendix 4. Focal non-motor seizures are divided into autonomic, behavior arrest, cognitive, emotional or sensory seizure types. The primary features of non-motor seizures are related to sensation, emotions, thinking or experience and they occur at the beginning of non-motor seizures. The characteristics of focal non-motor seizure types are described in Appendix 5. [12, 30]

2.1.4. *Status epilepticus*

Status epilepticus (SE) is a serious condition where epileptic seizure or a series of seizures without recovery between them is lasting more than defined time period. Generally, if epileptic seizure lasts more than 5 minutes the condition requires intensive care. Status epilepticus is an epileptic seizure that lasts more than 30 minutes. Prolonged seizures are caused by the failure of the mechanism responsible for seizure termination or the initiation of abnormal mechanisms. Classically status epilepticus is defined as “*a condition characterized by an epileptic seizure that is sufficiently prolonged or repeated at sufficiently brief intervals so as to produce an unvarying and enduring epileptic condition*”. [35, 36, 37, 38 p.155]

Status epilepticus may appear after acute or chronic disorder that affects the central nervous system such as cerebral infarction, central nervous system infection, brain damage, metabolic syndrome, electrolyte imbalance or hypoxia. Status epilepticus may lead to neuronal death, neuronal injury, and alteration of neuronal networks, depending on the type and duration of seizures, therefore status epilepticus

must be terminated quickly with aggressive therapy. It has been reported that on average 20% of cases of status epilepticus are fatal and long-term mortality rates in children is 22% and 57% in adults. Approximately one third of all status epilepticus patients die, one third gets permanent various degrees of neurological deficiencies and one third recovery completely. [35-37, 39]

Status epilepticus can be divided into convulsive and non-convulsive status epilepticus. Convulsive status epilepticus includes motor symptoms and non-convulsive status epilepticus non-motor. Non-convulsive status epilepticus includes loss of consciousness, hallucinations or symptoms that are more difficult to observe than motor symptoms. Convulsive and non-convulsive status epilepticus are divided into generalized and focal based on the EEG registration. More specifically status epilepticus are divided into five manifestation types: tonic-clonic SE, absence SE, myoclonic SE, focal SE with impaired consciousness and focal SE without impaired consciousness. The characteristics and the EEG features of each status epilepticus type are presented in Appendix 6. Tonic-clonic status epilepticus is the most common status epilepticus type and is 70% of all treatment cases. [37, 40, 41 p.199-200]

Only for 30 – 60 % of status epilepticus patients have been diagnosed with epilepsy earlier. In Finland appearance of status epilepticus is about 340/100 000 annually. Annually the appearance of number of status epilepticus that needs hospital care is 20/100 000 in Finland and the risk is biggest in children and in elder. The factors that most often causes the condition of status epilepticus for adults are poor balance in treatment of epilepsy, earlier diagnosed neurological diseases and acute symptomatic factors such as transient ischemic attack or severe cardiac arrhythmia attack. For children the most common causes are fever, earlier diagnosed neurological diseases and infection in central nervous system. Annually in Finnish intensive care units 3.4/100 000 patients were treated with severe status epilepticus and 0.7/100 000 patients with extremely severe status epilepticus. [36]

Status epilepticus is a neurological emergency condition which needs quick aggressive therapy and even general anesthesia to terminate a seizure [41 p.202]. Still, majority of epileptic seizures are short-lasting and terminate without special treatment. In status epilepticus there are two operational dimensions: the first is the time point (t_1) where the seizure should be regarded as continuous seizure activity and the time point (t_2) where the risk of long-term consequences start to increase dramatically. The current time points t_1 and t_2 are defined only for tonic-clonic, focal status epilepticus with impaired consciousness and absence status epilepticus (Table 4.). The operational times of tonic-clonic status epilepticus, focal status epilepticus with impaired consciousness and absence status epilepticus are presented in appendix 6. For other status epilepticus types the time points have not yet been defined. In clinical perspective t_1 determines the time when treatment should be started and t_2 determines how aggressive treatment should be implemented to prevent long-term consequences. [35, 36]

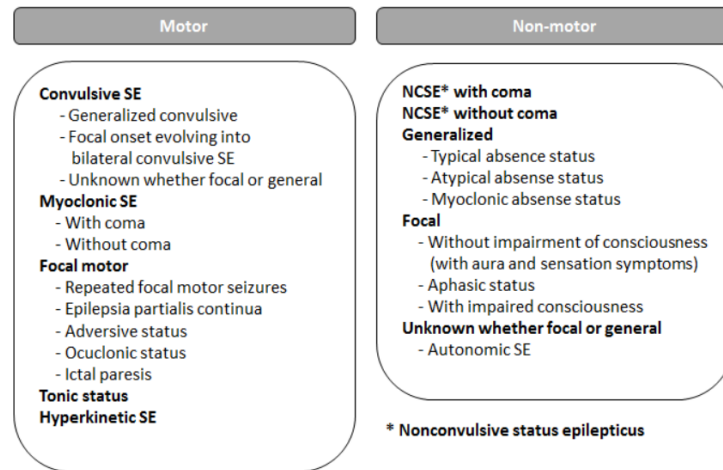


Figure 5. Classification of status epilepticus [35].

In more details, status epilepticus is classified in four different axes semiology, etiology, EEG correlates and age. Based on the semiology status epilepticus is classified into two classes the presence of absence of prominent motor symptoms and the degree of impaired consciousness. Status epilepticus is not a disease entity but rather a symptom that has numerous different etiologies. A classification of seizure types cannot be simply applied for classification of status epilepticus because the symptoms during the status epilepticus may differ a lot compared to short-lasting seizures. The classification of status epilepticus based on etiology is presented in figure 3. The ictal EEG pattern of any type of status epilepticus are not specific. Therefore, there are no EEG criteria for status epilepticus, but there are features such as location, name of the pattern, morphology, time-related features, modulation and effect of intervention which describe EEG patterns in status epilepticus (Table 9.). The different age groups of status epilepticus are neonatal (0 to 30 days), infancy (1 month to 2 years), childhood (>2 to 12 years), adolescence and adulthood (> 12 to 59 years) and elderly (≥ 60 years). [35]

Status epilepticus has four phases based on the degree of severity: early status epilepticus, established status epilepticus, refractory status epilepticus and super refractory status epilepticus (Appendix 7.). The objectives of treatment of status epilepticus are maintenance of vital functions, quick termination of clinical seizure symptoms and deviant electrical discharge activity of the brain. After the acute phase, the most important objectives of treatment are to prevent recurrence of seizures, observation of systemic complications, detection of etiological factors and minimization of mortality and secondary symptoms. [36]

2.1.5. Origin of EEG

The electroencephalogram (EEG) is a neurophysiological research for measurement of an electrical activity of brain cells [42 p.1]. In EEG recording electric potential differences are measured between electrodes that are attached invasively or noninvasively to the scalp [43]. The origins of cerebral potentials reflect the intrinsic electrophysiological properties of the nervous system [42 p.2]. EEG is primarily generated by cortical pyramidal neurons in the cerebral cortex that are oriented

perpendicularly to the brain's surface [44]. Creation of EEG signals needs simultaneous activation of millions nerve cells in the same area of a brain. Recently it has been reconsidered that EEG may contain components from other cellular structures. [43]

EEG is created when electrical charges move within the cerebral cortex. Neuronal function of the cerebral cortex is maintained with ionic gradients established by neuronal membrane. Electrical currents of cerebral activity are required to be amplified and displayed for interpretation. Synaptic potentials are the primary source of the extracellular current flow that is seen in the EEG. These synaptic potentials are called as excitatory postsynaptic potentials that flow inwardly to other parts of the cell via sodium and calcium ions and inhibitory postsynaptic potentials which instead flow outward in the opposite direction and involve chloride or potassium ions. The EEG provides evidence of the continuous and changing voltage fields varying with different locations on the scalp over time. [42 p.2-3]

The most important measured quantities of EEG are amplitude and frequency. The amplitude of EEG is typically around tens to hundreds of microvolts and reflects the size of synchronously activating group of nerve cells. Measuring technology is an important factor in EEG but also the thickness, electrical conductivity and size of the scalp and distance and position of electrodes have significant impact on measurement. Frequency changes slowly in EEG from 1 to several hundreds of Hertz. The spontaneous electrical brain activity consists of voltage fluctuation in different frequencies. Traditionally the frequency bands are divided into delta (less than 4 Hz), theta (4-8 Hz), alpha (8-13 Hz) and beta (over 13 Hz) bands. Fast brain activity at a frequency of around 40 Hz is called as gamma activity. Practically frequency of normal EEG is around 0.5 – 25 Hz, but epileptic seizures may cause rapid bursts that reach even over 200 Hz frequency in EEG signal. [43, 45 p.50-51]

The setup of EEG varies depending on use, therefore there are several EEG paradigms for different targets (Appendix 8.). The basic EEG measurement setup is called as 10-20 international system (Figure 6.). Earlier typical EEG research included only 19 channels, but currently International Federation of Clinical Neurophysiology (IFCN) has given a new recommendation to increase the number of electrodes to 26. A larger number of electrodes disclose brain activities that were not seen via a smaller number of channels. Most of EEG measurements are recorded in intensive care or in emergency duty. The number of electrodes is decreased to most important channels in intensive care and emergency duty because of the circumstances. Often synchronized video recording and polygraph channels are added to the EEG measurement. Typical polygraph channels are ECG, EOG, EMG, respiration measured with respiration belt, oxygen saturation measured with pulse oximetry and movement interference signals measured with piezoelectric sensors. [43, 46]

2.1.6. EEG registration and activations

The most often EEG is measured from electrodes that are attached noninvasively to the scalp's skin. The names of electrodes indicate the lobe of the cerebral cortex where the electrodes are placed (Figure 4.). Electrodes vary between different materials, shapes and the way of them are attached. Electrodes can be attached to the skin of head separately, with a hat or a web along with a gel that ensures better contact between the electrode and skin. The type of electrodes affect for a little bit on

the susceptibility to interference, frequency response and field of application. An electrical connection is created from brain tissue to amplifier via electrodes and cables. An amplifier measures EEG signals based on the chosen montage. Functionality of channels should be tested with calibration signal at the beginning and in the end of registration and during registration if necessary. [46]

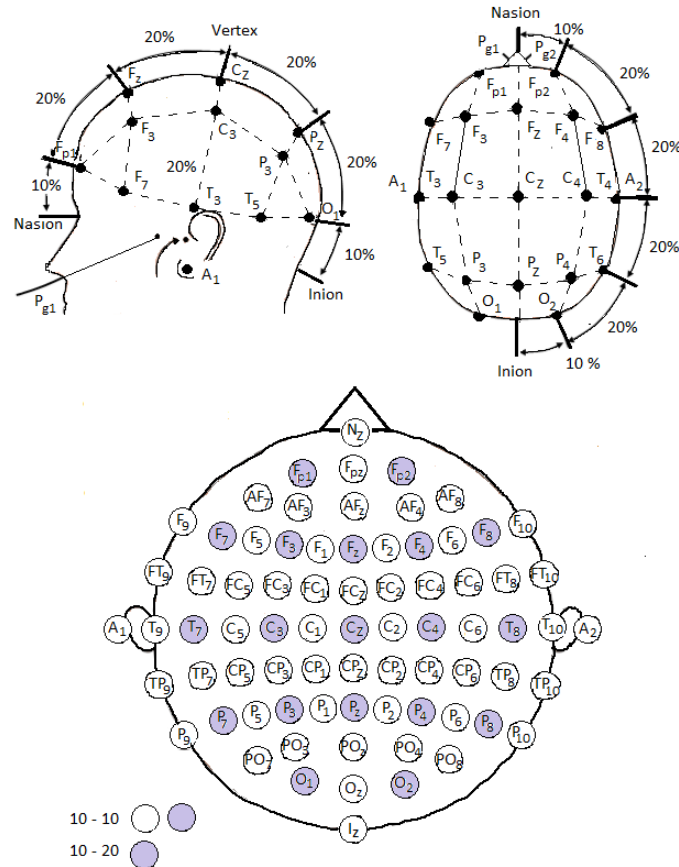


Figure 6. The international 10-20 system seen from (figure on left) left and (figure on right) above the head. A = Ear lobe, C = central, Pg = nasopharyngeal, P = parietal, F = frontal, Fp = frontal polar, O = occipital.

The EEG montage is a standardized arrangement and selection of channel pairs and chains for display and review. The montages are divided into two categories: bipolar and referential/monopolar. In bipolar montages, one electrode is connected to one or two neighboring electrodes forming a chain of electrodes. The bipolar longitudinal pattern is the most commonly used bipolar montage. In longitudinal pattern each channel connects adjacent electrodes from anterior to posterior in two lines (Figure 6.). The bipolar traverse montage connects adjacent electrodes in a chain from left to right. In monopolar montage each electrode is connected to a single referential point (Figure 7.). A referential point of monopolar montage can be the vertex, the mastoid process or common average reference. The Cz electrode indicates the vertex and the mastoid process includes electrodes from either individual ears or a mathematical derivation of both sides. Another commonly used montage in the evaluation of epilepsy is the Laplacian montage. The Laplacian

means a source, derivation montage where each electrode is compared with a weighted average of the neighboring electrodes. [47]

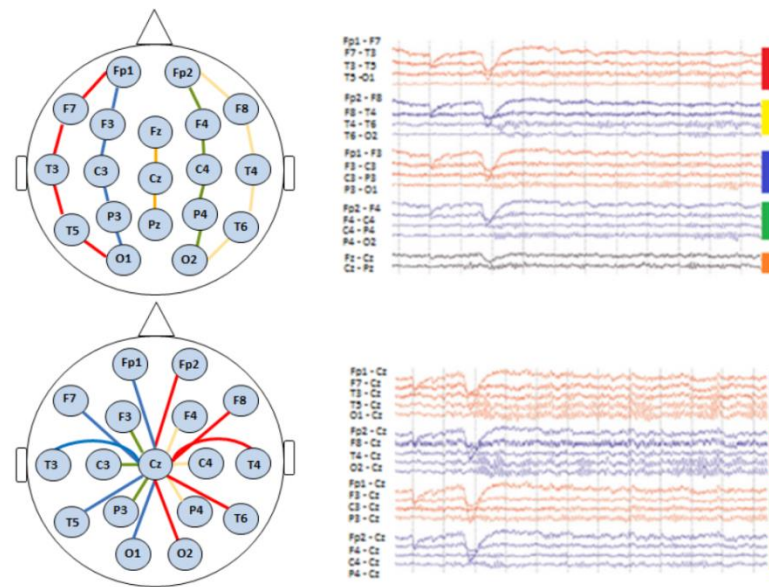


Figure 7. Longitudinal bipolar montage on top and common electrode reference monopolar montage on below.

Various types of activations are utilized during EEG registration to get useful information about the function of neural network. Routine activation is eyes closed – eyes open activation. On examination of epilepsy, hyperventilation and flashing light provocation are the key activations to provoke epileptiform activity. During hyperventilation test, a patient takes deep breaths 3 – 5 minutes, causing decrease in concentration of carbon dioxide in the blood. Concentration of carbon dioxide is a primary regulator of the cerebrovascular system. In flashing light provocation, a patient is predisposed with flashing light of different frequencies while a patient is having eyes open and alternately eyes closed. A sleep deprivation is also used to provoke epileptiform activity. In sleep deprivation, a patient is given instructions not to sleep at night before EEG registration. Sleep deprivation requires deep sleep hence patients are requested to stay awake overnight to guarantee better sleep quality during EEG registration. [48 p.81]

2.1.7. EEG in diagnostics of epilepsy and status epilepticus

EEG measurement is the primary clinical tool for diagnostics of epilepsy and status epilepticus. EEG is used for neurophysiological identification, classification, quantification and localization of epileptiform discharges (EDs). In more details, EEG is utilized to indicate the existence of a finding that correlates with epilepsy, which epilepsy type is in question and is epileptic seizure focal or generalized. EEG is used also to quantify the number and duration of epileptiform abnormalities. Epilepsy is an electrical dysfunction of the brain, and hence EEG is the most suitable research method for evaluation of electrical activity of a brain. Possibility to monitor dynamically electrical activity of the brain is also necessary for presurgical evaluation, considering stopping AED therapy, monitoring of seizure frequency,

monitoring effect of medicine, and deciding can a patient have a driver's license or flight certificate. [41 p.200-201, 49 p.144-145, 50, 51, 52]

In basic EEG findings that correlate with epilepsy are found in only 50% of epilepsy patients during the first registration. By repeating EEG recording with a sleep deprivation, 80 - 90 % of epilepsy findings are discovered. In 1% of population can be seen features of epileptiform activity in the EEG without epilepsy. An ambulatory EEG provides a possibility to monitor electrical activity of the brain in the patient's normal home environment for natural recording. The advantages of ambulatory EEG are to quantify the number and duration of subclinical seizures and seizures without awareness. In some cases, focal epileptic seizures are difficult to be identified. Utilization of invasive EEG monitoring with intracranial electrodes is necessary if the location of seizure onset has not been found with non-invasive presurgical evaluation or the goal is to carefully define cortical function in an area of planned surgical resection. [10, 49 p.145, 50]

The EEG is the most important research method in diagnostics and monitoring treatment of status epilepticus. In case of status epilepticus, EEG is especially used for suspicion of non-convulsive status epilepticus. In emergency situations, EEG is a primary clinical tool in the classification of status epilepticus by defining is a patient suffering from tonic-clonic SE, absence SE, focal SE or psychomotor SE. EEG is required to evaluate an appropriate choice of antiepileptic-drug. A patient that is suffering from status epilepticus is in ongoing EEG monitoring therefore EEG helps also to determine the size of the dose of antiepileptic-drug and to indicate when status epilepticus seizure ends. [49 p.144-146, 52]

The symptoms of convulsive status epilepticus are easier to observe than non-convulsive because of motor symptoms. Both types require EEG monitoring, but non-convulsive is more critical because the symptoms of non-convulsive status epilepticus might not be that clearly observable [37]. General status epilepticus is always visible in EEG signals, but in some cases of focal status epilepticus EEG registration appears to be normal. In focal seizures epileptiform activity may be in the too small area of a brain or too deep in brain tissue so it cannot be seen in EEG signal. [49 p.145-146]

Diagnosing epilepsy without EEG is more harmful than not diagnosing epilepsy even though the patient is suffering from it. Diagnosing epilepsy requires always other clinical information along with EEG therefore the diagnosis of epilepsy cannot be fully based on the EEG interpretation. In some cases interictal epileptiform features in EEG do not necessarily always indicate epilepsy. Incorrect diagnostics are often caused by wrong EEG interpretation when EEG contains artefacts or doctors have dissenting opinions. [41 p.199-202]

There are conditions that are related to cardiac, toxic/metabolic, pulmonary, movement disorders, migraine headache, cerebral diseases, autonomic disorders, sleep disorders, vestibular dysfunctions or transient global amnesia that may mimic epileptic seizure in EEG recording. There are also psychiatric events that may mimic symptoms of epileptic seizures such as psychogenic non-epileptic seizures, malingering, catatonia, panic attack, hallucinations/psychosis, episodic dyscontrol, fugue state and Munchausen syndrome/ Munchausen syndrome by proxy. [10]

2.1.8. Interpretation of EEG

EEG interpretation requires understanding of EEG's basis, recording technology, descriptive terminology, and general capabilities and limitations. By understanding these factors, the EEG interpretation and diagnostics can be more clearly addressed and leads to appropriate prognosis and diagnosis. EEG interpretation is all about identifying the patterns and defining relevant patterns to the patient. The essential EEG feature components provide insights about the cerebral electrical activity. Becoming familiar with the component features that characterize each pattern is an important aspect in EEG interpretation. [53, 54]

The EEG signal consists of several patterns of rhythmic or periodic activity [55 p.37]. EEG waveforms are divided into the delta (0.1–<4 Hz), theta (4–<8 Hz), alpha (8–13 Hz) and beta (14–30 Hz) frequency bands. There is also a high frequency wave pattern called as gamma (30 – 80 Hz) and mu wave pattern that overlaps with other frequency bands (7 – 11 Hz). EEG rhythms reflect physiological and mental processes. EEG signal also may contain spikes, transients and other waves and patterns which are associated with various disorders of the nervous system such as epilepsy. [55 p.37-38, 56]

The identification of epileptiform patterns of features is started with categorizing EEG segments into attenuation, transient and repetition types (Figure 8.). Attenuation is defined as either low voltage or a marked decrease in amplitude. The attenuations are categorized only into focal, hemispheric, bilateral, or generalized according to distribution. The distributions have subheadings based on the more detailed locations: frontal, temporal, parietal, central, occipital and anywhere. A transient is defined as an isolated wave or complex and it first also categorized by the distribution and characterized by location(s). Transients are further categorized into monophasic, diphasic and triphasic and whether it is a spike, sharp wave, or slow wave. Finally, the transients are characterized by polarity. Repetitions mean a recurring transient and they are also categorized by distribution and whether the repeating transient is monophasic or polyphasic. Repetitions are further characterized by location(s) and whether the pattern is regular or evolving. [53]

There are various physiological and psychological conditions that may have similar EEG patterns than epileptic discharges. These physiological and psychological conditions are listed in “*EEG in diagnostics of epilepsy and status epilepticus*” section. There are also physiological and non-physiological artefacts that may mimic epileptic discharges in EEG. These artefacts are respectively mentioned in “*EEG artefacts*” section of this thesis. These factors need to be taken also into account in EEG interpretation. [10, 50]

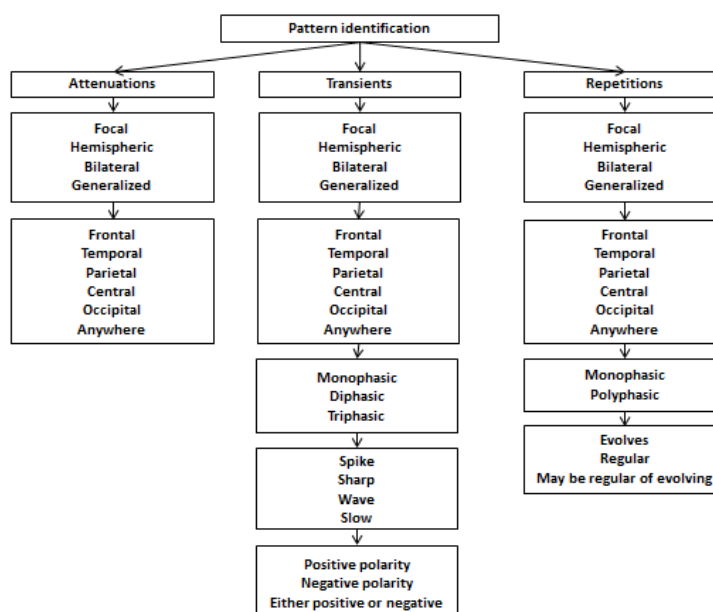


Figure 8. Pattern identification.

2.1.9. Normal EEG

Identification of epileptiform patterns is based on the development of criteria for normal EEG activity. It is essential to provide knowledge of normal waveform variations for each age group throughout the lifecycle from infants to elder. Brain waves can be divided into delta, theta, alpha, beta and additionally mu and gamma bands. Different frequency bands and their physiological relations are presented in Table 2. Normal EEG of adults consists of alpha and beta activity, but includes also other wave frequencies within a range of 0.5-70 Hz. The waves occur in clear rhythms or spontaneously separately, where waves in different frequencies are mixed in together. EEG is constantly dynamic and changes along with the state of physiology. Consequently, it can be interpreted from EEG is person awake, in a light sleep, in a sleep, nervous, cognitively active, having eyes open, having eyes closed etc. EEG cannot be only used to give evidence are patterns in EEG recording normal or abnormal, but with the relevance between EEG and condition of patient. EEG differs between individuals but for the same individual EEG is fairly similar between separate recordings but changes during aging. [42 p.5-7, 57 p.109-110, 57 p.127]

The alpha rhythm is a normal pattern that relates to gated levels of visual attention, perhaps as an active stand-by state [58]. The alpha rhythm is the predominant activity while a patient is mentally and physically relaxed but not in light sleep. It is common in wakeful, resting adults, especially in the occipital area. The alpha rhythm attenuates while opening eyes, but the reactivity differs among individuals. Frequency of alpha rhythm is 9-11 Hz in 70% of 20 years old individuals. One out of one hundred has slow 4-5 Hz alpha variant rhythm in the posterior. Fast alpha variant rhythm that occurs in 14-20 Hz frequencies is even rarer than slow alpha variant rhythm. Alpha variant rhythms react to external stimulus just like alpha rhythm. Over half of the adults have alpha rhythm with amplitude of between 20-50 μ V. 10% of adults have less than 20 μ V and it is even rarer to have over 100 μ V amplitude of alpha rhythm. Alpha rhythm occurs usually in the occipital

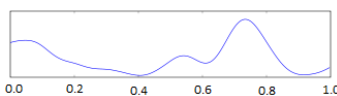
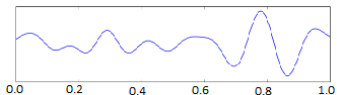
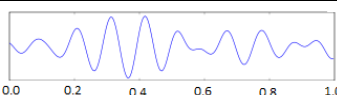
and posterior temporal areas. Sometimes alpha rhythm appears also in anterior temporal, central areas and even weakly in a couple of electrodes in frontal area. [55 p.38, 57 p.109-116]

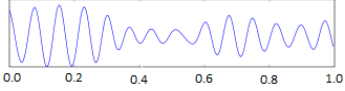
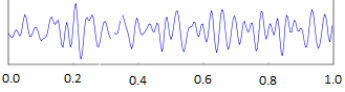
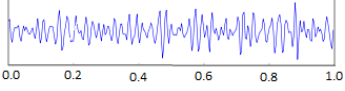
The alpha rhythm is replaced by slower rhythms in various stages of sleep. The theta is the frequency range from 4 Hz to 7 Hz. The theta waves are common in children in a wide area of the central temporal region of the brain. They may appear also in healthy adults at the beginning stages of sleep or during arousal. The presence of theta waves in wakeful adults indicates abnormality. One out of third of young adults has subdued theta activity at 6-7 Hz frequencies in frontal or frontal central areas. Theta activity is prominent during executing mental activation tasks, but it can be more clearly seen in light sleep because of attenuation of background rhythms. Bilateral or generalized polymorphic theta activity is the evidence of either normal drowsiness or of encephalopathy in adults. In infants, occurrence of theta activity is normal. [45 p.51-56, 55 p.34-38, 57 p.122-126, 59]

The delta is the frequency range from 0.5 Hz to 4 Hz. It tends to have the highest amplitude and the slowest waves compared to other EEG frequency bands. The delta waves appear in the deep-sleep stages in healthy adults. Normal polymorphic delta activity is a sign of non-rapid eye movement sleep. The frequency of polymorphic delta activity decreases with progression of sleep into its deeper sleep stages. Polymorphic delta activity is normal or abnormal finding depending on its features and circumstances. Abnormal polymorphic delta activity may indicate encephalopathy. [55 p. 37, 60]

High frequency beta waves (14 Hz to 30 Hz) are more commonly a sign of drowsiness or sleep onset, but it is present in full wakefulness in some individuals. More rarely, it is associated with anxiety and vigilance. The majority of beta activity seen in EEG in adults is occurring at 18-25 Hz frequencies. Less common are 14-16 Hz and over 35 Hz activity. Beta activity has an amplitude of less than 20 μ V. Beta activity occurs mostly in frontal, central and sometimes it is visible in occipital areas. Beta activity in occipital areas reacts to stimulus just like alpha rhythm and central beta as mu rhythm. Mu rhythm attenuates while planning movement or during movement and it is observable in C4 and C3 electrodes. Beta activity emphasizes during cognitive performance, in light sleep and in REM sleep. [55 p.37, 57 p.113, 61]

Table 1. Normal brain waves

Brain wave	Frequency band	Physiological relation	Graph
Delta	0.5-4 Hz	Deep sleep state, deep sleep, loss of bodily awareness	
Theta	4-8 Hz	Hypnoidal state, reduced consciousness, deep meditation, dreams, light sleep, REM sleep	
Alpha	8-13 Hz	Normal resting state, physically and mentally related, awake but drowsy	

Mu	8-13 Hz	Indicator of activation of motor neurons in resting state	
Beta	14-30 Hz	Normal waking state, awake, normal alert consciousness	
Gamma	30-100 Hz	Heightened perception	

2.1.10. Epileptiform abnormalities

Epileptiform abnormalities are seen as distinguishable transients from background activity in EEG signal [62]. The term epileptiform is defined as an EEG pattern that is associated with a relatively high risk of having seizures [51]. Epileptiform discharges appear in various morphologies (Appendix 9.) [63]. There are 6 criteria regarding to the characteristic morphology of epileptiform patterns and 4 of them must be fulfilled to define a specific pattern as an epileptiform pattern:

1. *“Di- or tri-phase waves with sharp or spiky morphology (i.e. pointed peak).*
2. *Different wave-duration than the ongoing background activity, either shorter or longer.*
3. *Asymmetry of the waveform: a sharply rising ascending phase and a more slowly decaying descending phase, or vice versa.*
4. *The transient is followed by an associated slow after-wave.*
5. *The background activity surrounding epileptiform discharges is distributed by the presence of the epileptiform discharges,*
6. *Distribution of the negative and positive potentials on the scalp suggests a source of the signal in the brain, corresponding to a radial, oblique or tangential orientation of the source. This is best assessed by inspecting voltage maps constructed using common-average reference” [56].*

Based on the descriptions of Vakkuri et al. [64] delta with spikes (DSP), rhythmic polyspikes (PSR), periodic epileptiform discharges (PED), and suppression with spikes (SSP) are regarded as epileptiform patterns. Figure 9. [65] illustrates these typical epileptiform patterns in EEG signals. DSP includes delta activity of any frequency with regular or irregular spikes. PSR are waveforms with more than two negative and positive deflections appearing at regular intervals, associated with slow wave or mixed frequency EEG activity between spike complexes. SSP means short episodes consisting mostly of single spike appearing during complete EEG suppression. [65]

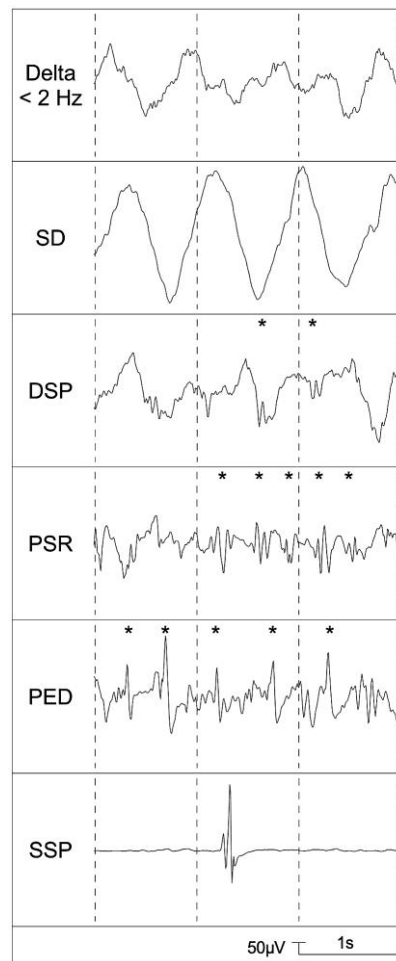


Figure 9. Epileptiform patterns [65]

Each of these epileptiform patterns consists of different combinations of spikes, sharp waves, polyspikes or slow-wave patterns. Sharp waves and spikes differ in their duration (70-200 ms vs. less than 70 ms), but in practice they reflect the same thing. Polyspikes are multiple spikes observed typically in rapid succession (typically at frequencies of 10 Hz or faster). Spike-and-slow-wave complex corresponds to a phenomena where a spike is instantly followed by a slow wave. Polyspike-and-slow-wave complex is similar to spike-and-slow-wave complex, but is associated with polyspikes followed by one or more slow waves. [62]

An epileptiform EEG pattern is split into preictal, ictal, interictal and postictal states (Figure 9.) [51, 66]. A preictal state means a period in EEG before a seizure occurs. A preictal state includes often aura symptoms in some epilepsy types, but sometimes appears without any symptoms. There has been evidence to possibly predict epileptic seizure before it occurs. Detection of the preictal state from EEG has been highly interesting field in epilepsy research [66]. The electrographic evolution of a seizure commonly includes one or more following features: change in frequency, amplitude, distribution, and/or waveform [67]. These features form at least five common EEG patterns at the start of the seizure: rhythmical evolving at theta, delta and alpha frequencies, rhythmic spiking, spike-waves or

electrodecremental. In some cases, there is no clear change in EEG. The epileptiform patterns and their characteristics are presented in Appendix 10. [51]

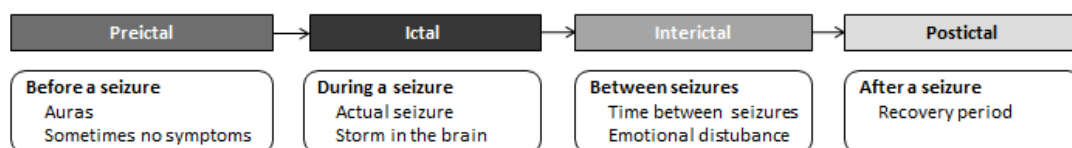


Figure 10. States of a seizure.

The ictal EEG pattern also called as a seizure pattern is defined as the EEG finding that is present during a seizure [51]. The ictal pattern consists of repetitive epileptiform EEG discharges at <2 c/s and/or characteristic with gradual change in frequency, amplitude, morphology and location, lasting at least several seconds (usually > 10 s) [56]. The ictal activity patterns are described in table 8. Ictal pattern typically includes rhythms or repetitive sharp waves regardless of a seizure type. There are two types of ictal patterns: ictal patterns for focal seizures and ictal patterns for generalized seizures [67].

A focal-onset has features related to frequency, amplitude, distribution, and waveform. Frequency of EEG pattern during a focal-onset may be increased or decreased several Hertz within any of the normal EEG frequency bands. Amplitude is usually increased, but often followed by attenuation at the beginning of the seizure as amplitude has evolved to its maximum. The distribution of evolution reflects the anatomic spread or the seizure across the cerebral hemisphere. The ictal rhythm of focal-onset seizure may lead to a generalized distribution. During generalization the ictal pattern has fast activity features of a generalized-onset seizure. The generalized-onset seizures have more similar pattern in their ictal and interictal EEG pattern than focal-onset seizures. [68 pp. 213-214]

Generalized-onset seizures can be characterized by one of three types: Generalized paroxysmal fast activity (GPFA), generalized spikes and slow-wave complexes (GSW) or an electro decrement. GPFA begins with low amplitude and fast frequency activity that often leads into postictal slowing and attenuation. GPFA evolves with increasing amplitude and a decreasing frequency. The GSW complex is most commonly triphasic with a negative initiating spike and negative completing slow wave. In some cases GSW may appear as polyphasic because of a polyspike initiation. GSW occurs repeatedly with duration of at least 3 to 5 seconds. Generalized-onset may be also in form of the electrodecremental pattern which means a sudden and generalized attenuation. In the electro decrement 20 to 40 Hz low voltage activity evolves with gradually increases in amplitude and decreases in frequency over the subsequent few seconds. The electro decrement typically evolves into GPFA which has similar waveform. Generalized-onset seizure typically ends as abrupt resolution without evolution, with evolution in waveform and frequency, or with disintegration of the ictal pattern. [68 p.234-235]

The interictal pattern means the EEG finding between seizures. During an interictal state a patient may have emotional disturbances. Interictal epileptiform discharges are sharp waves, spikes and polyspikes often followed by slow waves. Morphologies of epileptiform interictal activity are described in table 9. Often the boundaries of different states are unclear and arbitrary. Interictal-ictal boundary is blurred for many epileptiform discharges. Especially, the distinction between ictal

and interictal generalized IEDs are difficult because often the only difference in the waveform is duration. [51, 69]

Focal interictal epileptiform discharges (IEDs) have four characteristic features: a sharply countered component, electronegativity on the cerebral surface, disruption of the surrounding background activity, and a field that extends beyond one electrode. These features do not always appear at the time of IEDs and all of these features are not epileptiform, but still they are together a useful framework for considering IED waveforms. IEDs are a sign of epilepsy, but they are not infallible because they do not have full specificity. IEDs are present in a small percentage of individuals in the healthy population (0% to 2.6%). Morphologies of focal interictal epileptiform activity are presented in Appendix 9. [69]

Generalized interictal epileptiform discharges are not exactly generalized across the entire cerebral hemisphere. Generalized IEDs commonly appear in the midfrontal region and spread to the entire frontal region and some or most of the parietal region. Generalized IEDs often occur as spikes and sharp waves with after-going slow waves or as burst of successive spikes. The waveform of generalized IEDs varies less than the waveform of focal IEDs. Generalized IEDs are an indicator of generalized epilepsy syndromes and typically occurs across all forms of generalized epilepsies. [69]

Paroxysmal fast activity (PFA) is an epileptic pattern that is a specific type of bursting beta frequency range activity. Paroxysmal fast activity may occur either as focal or generalized epileptiform discharge. Paroxysmal fast activity can be also either interictal or ictal abnormality depending on its duration. PFA begins with a fast, regular or irregular rhythm that has usually greater or occasionally lower amplitude than the background activity. The abrupt change in amplitude and the presence of a beta frequency rhythm are key features of PFA. A waveform of PFA consists of repetition of monophasic waves with a sharp contour produced by the high frequency. [70]

Periodic epileptiform discharges (PEDs) are triphasic with a sharply contoured wave, followed by a slow wave, which have three types: periodic lateralized epileptiform discharges, bilateral independent periodic epileptiform discharges and bilateral periodic epileptiform discharges. Periodic lateralized epileptiform discharges (PLEDs) occur with a singular focus at any scalp electrode location. Whereas PLEDs have a singular focus, bilateral periodic epileptiform discharges (BiPEDs) are symmetrically and synchronously occurring periodic epileptiform discharges. When PLEDs have asynchrony and dissimilar appearance, they are termed as bilateral independent PLEDs (BIPLDs). In BIPLDs, discharges occur at more than one location in the cerebral hemisphere. [71]

The postictal pattern is defined as a recovery period after an epileptic seizure. The distinction between ictal to postictal EEG can be classified by rate or location. The Postictal is seen as attenuation or slowing in usually in the delta frequency range in EEG. In some cases, a patient enter sleep during recovery state. The postictal includes commonly cognitive disturbances, focal motor and sensory deficits, postictal headaches and migraines, and postictal automatisms. The postictal state often includes depression or rarely even psychosis. [51, 72-74]

Epileptiform patterns are characteristics of epilepsy, but their appearance in EEG does not always mean that a patient has epilepsy. Beside clear epileptiform pattern there are so called normal patterns that may also be a sign of epilepsy in particular circumstances. The benign epileptiform transients of sleep (BETS) are normal EEG

patterns which commonly occur during light sleep. Although some studies have found evidence that individual with epilepsy have higher rates of BETS than among those without epilepsy. Moreover, needle spikes may have also indirect relation to epilepsy. The burst suppression that is generally associated with coma and anaesthesia. In epilepsy context the burst suppression pattern indicates end stage of status epilepticus or severe epileptic encephalopathies of infancy. A transient with a vertex localization called as K complexes with spiky waveform may occur during arousal from NREM sleep in generalized epilepsies or focal epilepsies. [62, 75-78]

2.1.11. EEG artifacts

The analysis and interpretation of EEG are often made complex by the presence of a variety of artefacts [79]. The EEG artefacts are seen in EEG signal as changes in voltage that originate from source other than the brain [80]. An EEG artefact is defined as a physiological potential difference due to an extra cerebral sources present in EEG recordings or a modification of EEG caused by non-biological extra cerebral factors [56]. It is necessary to recognize physiological and non-biological artefacts, identify the source of artefacts and especially eliminate nonessential nonphysiological from EEG before analysis are performed [50, 81]. The sources of noise in EEG could be physiological, the instrumentation used, or the environment of the experiment [82 p.91]. Every type of EEG recording always contaminates artefacts [50]. It is impossible to remove all of the noise from EEG because some sources of noise are difficult to identify [81]. Beside removal, appearance of artefacts in EEG signal can be also reduced by attenuating or protecting against interfering signals [80].

Physiological artefacts include cardiac, pulse, respiratory, sweat, gloss kinetic, eye movement, muscle and movement artefacts [80]. The most common artefacts during EEG recording are eye movements, tongue movements, talking, chewing, other body movements and sweat artefacts [83]. The most common non-biological artefacts are electrode artefacts that include electrode pop, electrode contact, lead movement, perspiration and salt bridge [84]. The artefacts may negatively affect the manifestation of seizure pattern and therefore significantly influence the identification of epileptic seizure. The three most vital artefacts are muscle artefacts, eye movements/blinks and white noise. Physiological and non-biological EEG artefacts are presented in Appendix 11. [85]

Some artefacts mimic EDs and may be therefore misinterpreted as epileptiform pattern. Cardiac sources, myogenic sources, body and eye movements may result in artefacts that mimic abnormality in the EEG. Single-electrical artefact may mimic spike-and-waves. For instance, vertical eye blink might be interpreted as a sharp wave on EEG. Incorrect recognition of artefact as an ED can adversely affect treatment and lead to wrong diagnosis. Denoising EEG signal before interpretation and analysis is a challenge priority. Digital filters and montage manipulation are used to recognize what is beneath artefacts. The use of polygraphs is helpful for the recognition of artefacts and their sources [50, 79].

2.2. Deep learning

2.2.1. Overview

Artificial intelligence (AI) (Fig. 8) is a field of computer science of developing computers to learn from experience and understand concepts through its relation to simpler concepts. The hierarchy of concepts enables computers to learn by building simpler ones out of the complicated ones. Some algorithms of artificial intelligence rely on hard-coded knowledge using logical inference rules. This concept is known as the knowledge base approach. [86]

Machine learning is a subset of artificial intelligence techniques that has capability to extract patterns from raw data and learn from them through statistical methods. The representation of the data effects heavily on the performance of a simple machine learning algorithm. A simple machine learning algorithm requires selection of the right set of features for artificial intelligence tasks. In many tasks it is difficult to know which features should be extracted. One solution to this is a representation learning approach, also known as deep learning. The concept of representation learning is based on obtaining better algorithm performance with learning from representations rather than hand-design representations. [86]

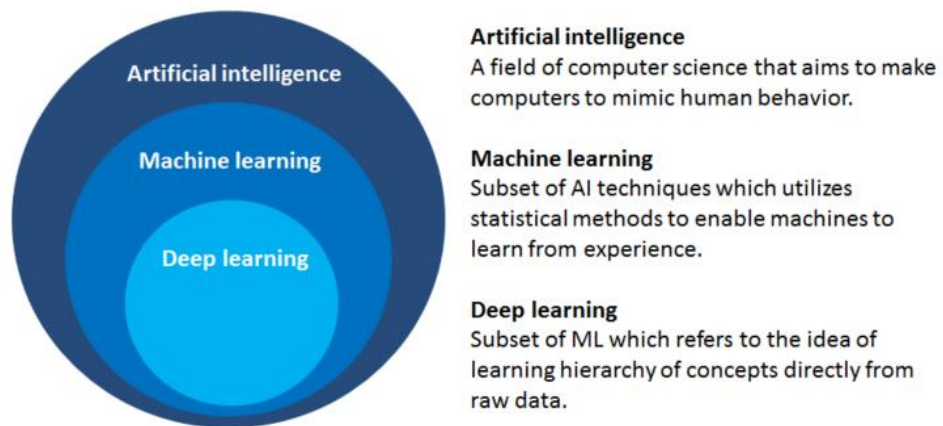


Figure 11. Venn diagram of the components of artificial intelligence.

Since 2006, deep learning has emerged as new, highly interesting area of machine learning research. Deep learning exploits multi-layer non-linear information processing for supervised or unsupervised feature extraction, transformation, pattern analysis and classification. Deep learning is a subset of machine learning that is based on algorithms for learning multi-level representations in order to model complex relationships among data. Learning the right representation for the data provides one perspective on deep learning. Another perspective is that depth enables deep learning neural network to learn complex multiple levels of representation and abstraction. Greater depth enables a neural network to execute more instructions in sequence. [86-88]

2.2.2. Artificial neural networks

Deep learning methods are based on artificial neural networks (ANNs) that have been motivated from its inception by the recognition capability similar to the human brain. ANNs are inspired by the human brain, but their computations are based on entirely different manners. The brain is nonlinear, a highly complex and parallel computer that performs certain computations via neurons for pattern recognition, perception and motor control. The artificial neural network can be either linear or nonlinear, currently do not reach the complexity with the same level than the human brain and has more limitations in parallel computing. [89] ANNs have the ability to learn by example through several iterations. They have the advantage that there is no need for deep knowledge of a field where they are applied. This has resulted in the rapid evolution of ANNs in various diverse scientific fields. [90].

In a big picture, a typical neural network consists of an input layer, multiple hidden layers and an output layer (Fig 10). The input layer accepts different type of input data such as biosignals, images, videos, speech, sounds, text, or numeric data. The hidden layers mostly consist of convolutional and pooling layers. The convolutional layers detect the local patterns and features in data from previous layers. The pooling layers semantically merge similar features into one. The output layer presents the classification or prediction results that are often controlled by activation functions such as softmax. [91, 92]

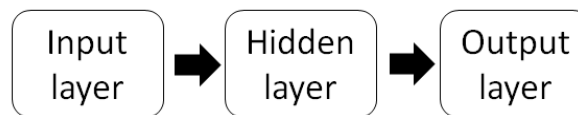


Figure 12. A typical block diagram of deep learning model.

In more details, ANNs consist of information-processing units, the neurons, and directed, weighted connections between those neurons [89, 93]. Each neuron has a limited computing capacity restricted by utilized activation function for combining input signals in order to calculate the output one. The output signal may be sent to other units along weighted connections. Weights excite or inhibit the signal being communicated. There are three basic elements of the neural model:

1. A set of *synapses*, or *connecting links*, which are characterized by a weight.
2. An *adder* sums the input signals, weighted by the respective synaptic weights of the neuron.
3. An *activation function* for limiting the amplitude of the output of a neuron by computing the weighted sum of input and biases, of which is used to decide if a neuron fires or not. [89-91]

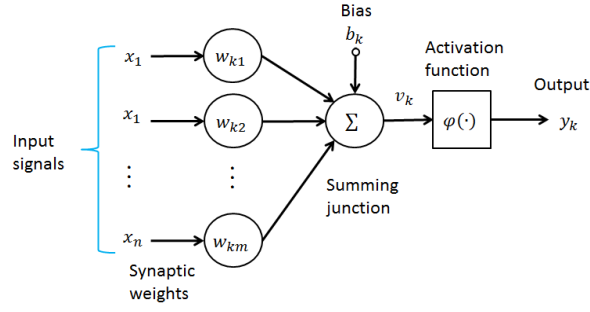


Figure 13. Nonlinear model of a neuron.

The output activity of neuron is expressed in equation (1):

$$y_k = \varphi \left(\sum_{j=1}^m w_{kj} x_j + b_k \right) \quad (1)$$

where x_1, x_2, \dots, x_n are the input signals, w_1, w_2, \dots, w_n are the respective synaptic weights of neuron, b_k is externally applied bias, $\varphi(\cdot)$ is the activation function and y_k is the output signal of the neuron [94]. The activation functions can be either linear or non-linear and they control the output of a neuron [91]. There are multiple different activation functions and their suitability varies among different machine learning tasks [95]. The proper selection of activation function improves results of neural network [91].

As mentioned earlier a deep neural network consists of input layer, hidden layers and output layer. All of the layers are formed by nodes, which's structure and functionality was presented in more details in top. A deep neural network has a number of components and hyperparameters that specify the configurations of a neural network. The components and hyperparameters are: network topology, loss function, weight initialization, batch size, learning rate, epochs and data preparation. These components and hyperparameters are reviewed in more details in “*Tuning deep neural networks*” section. [96]

2.2.3. Categorization of deep neural networks

Deep learning refers to various types of machine learning techniques and architectures that exploit hierarchical multi-layer non-linear information processing [95]. Deep neural networks can be broadly categorized into three major categories: deep networks for unsupervised or generative learning, deep networks for supervised learning and hybrid deep networks [95]. These approaches differ in a way how the weights of a neural network are trained [97].

Unsupervised or generative learning refers to no use of task specific supervision information in the learning process [95]. In unsupervised learning all the training samples are not labeled and the goal is to divide the samples into clusters based on the features of the samples [97]. Unsupervised learning networks are intended to detect high-order correlation of the observed data for pattern analysis when information about target class labels is available. Deep networks in unsupervised or

generative learning category can be either generative or non-generative but most of them are generative in nature, such as restricted Boltzmann machines (RBMs), deep belief network (DBNs), deep Boltzmann machines (DBMs) and generalized denoising and auto-encoders (Figure 10.). Recurrent neural network (RNN) can be considered as deep networks for unsupervised learning, but the model is also used for supervised learning. [95]

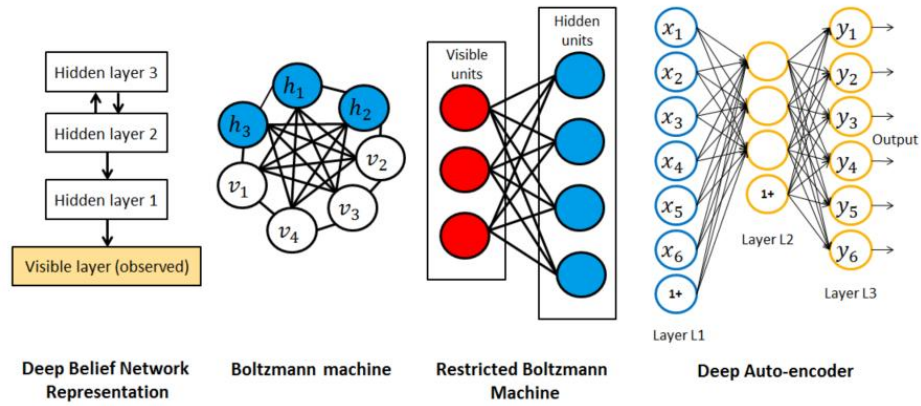


Figure 14. Deep Belief Network, Boltzmann machine, Restricted Boltzmann Machine and Deep Auto-encoder.

The most common approach of training is a supervised learning where all the training samples are labeled with the correct class [97]. Supervised learning networks are designed to directly provide discriminative power for pattern classification purposes. In supervised learning target label data are always available in direct or indirect forms. Most of the supervised architectures are shallow such as Hidden Markov Model (HMMs) and Conditional Random Fields (CRFs). [95]

Semi-supervised learning has characteristics of both supervised and unsupervised learning where only a small subset of the data is labeled [97]. Neural networks that use semi-supervised learning are called as hybrid networks. Hybrid networks are intended to discriminate data with the outcomes of generative or unsupervised deep networks. Discrimination in hybrid deep networks can be accomplished by optimization or/and regulation of supervised learning networks or with using discriminative criteria for supervised learning to estimate the parameters in any of the deep generative or unsupervised deep networks. [95]

Reinforcement learning means a training approach where the agent (a decision-maker) tries to maximize expected rewards in a particular environment over the long run. It interacts with through trials and errors to compute a behavior strategy with a given environment. Another common approach is fine-tuning, where previously trained weights are used as a starting point for new adjusted dataset or a new constraint. A technique where previously trained weights are adjusted with a new dataset is called as transfer learning. [97, 98]

2.2.4. Architectures

2.2.4.1. Deep feedforward networks

Deep feedforward networks (Fig 14.), also called as feedforward neural networks, or multilayer perceptrons (MLPs) are the first and simplest deep learning models [99]. A feedforward neural network has only direct connections from each neuron in one layer to the neurons of the next layer towards the output layer [100]. The calculations of feedforward network are based on approximating some function f^* . These models are called feedforward because data flows through function being evaluated from x without feeding the outputs back to itself. Feedforward neural networks typically compose together many different functions in a chain. A chain structure notation of three layer structured feedforward neural network for functions $f^{(1)}$, $f^{(2)}$ and $f^{(3)}$ is $f(x) = f^{(3)}(f^{(2)}(f^{(1)}(x)))$, where $f^{(1)}$ is the first layer, $f^{(2)}$ is the second layer and $f^{(3)}$ is the third layer of the network. The overall length of the chain indicates the depth of the model. [99]

Backpropagation is the most common learning techniques used in feedforward neural networks. Backpropagation is a gradient descent procedure to train the weights of the neural network with error function. The method calculates the gradient of the error function with respect to the neural network's weights. A typical problem with backpropagation is that the algorithm finds only a local minimum of the error function rather than a global minimum [101]. Moreover, in some cases a network may overfit the training data when there is only limited amount of data available resulting to a poor performance with testing data. [102, 103]

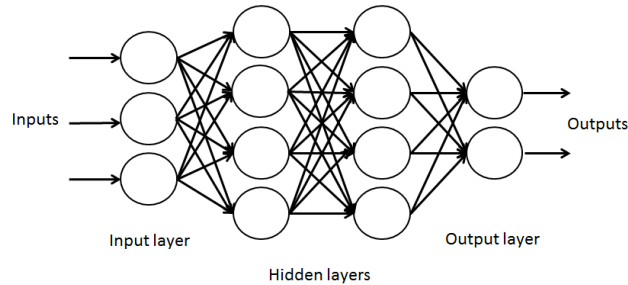


Figure 15. Example of feedforward network with an input layer, two hidden layers and an output layer.

Deep neural nets have improved significantly results of artificial neural networks, but they have had a serious problem with overfitting. Deep neural nets are also slow to train and use. Regularization is that can be used for addressing a problem of overfitting by adding weights of a network to a loss function. There are two different regularization techniques called as L1 and L2. In L1 regularization, the absolute weights are added to the loss function to make a model more generalized. In L2 regularization, the squared weights are added to the loss function. [96, 104]

Beside L1 and L2, there is dropout regularization technique. The basic idea of dropout is that the dropout layer randomly drops out units of the neural network during training by preventing units from co-adapting too much (Fig 15.). Dropout means removing temporarily units of the network along with all its incoming and outgoing connections. Srivastava, Hinton et al. has showed that dropout improves the

performance of neural networks. Dropout layers can be only applied to fully connected regions of neural network, but although they are not limited to feedforward neural networks. Data augmentation is also one technique to reduce overfitting. Data augmentation means a process where data are artificially enlarged using label-preserving transformations. [14, 104]

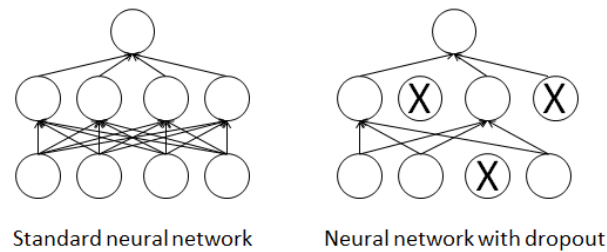


Figure 16. Standard neural network and neural network with dropout.

The problem of feedforward neural network, that has standard connections is that the model learns high-level features and this leads to vanishing gradient problem [16]. To overcome the vanishing gradient problem feedforward networks with shortcut connections has been developed. A neural network can have shortcut connections that skip one or more layers (Fig 16.). In feedforward networks these shortcut connections can be directed only towards the output layer. The word perceptron is most of the time used to describe a feedforward network with a shortcut connection. [100, 102]

Some studies have shown that shortcut connections achieve better results than the models that have standard connection [16, 105]. He, Zhang et al. has proved that residual networks are easier to optimize, and can gain accuracy from considerably increased depth of networks. The study has claimed that by skipping over layers the network avoids the problem of vanishing gradient. The residual network uses element-wise addition. Densenet connection enables a network to receive additional information from all preceding layers and passes on its own feature maps to all subsequent layers (Fig 16.). Since layers receive information from all preceding layers, the network requires fewer parameters and less computation to achieve better results than the models with standard and shortcut connections. [16, 17]

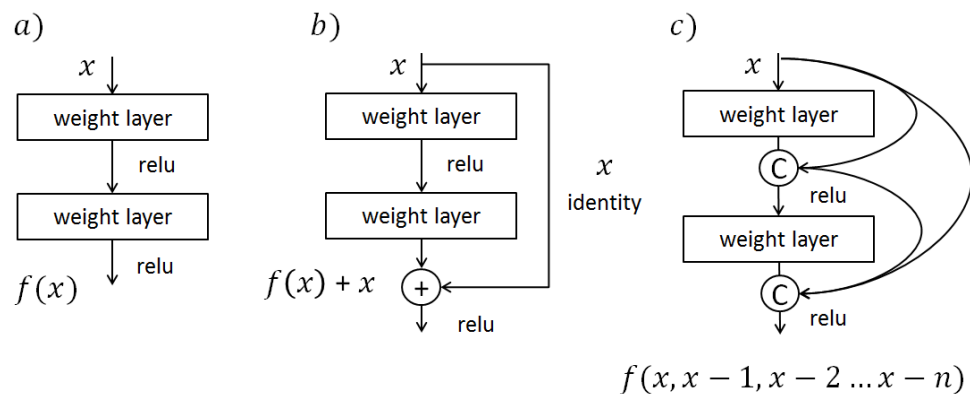


Figure 17. a) Standard connection, b) shortcut connection network with double layer skip and c) densenet connection.

2.2.4.2. Convolutional neural networks

Convolutional neural networks (CNN) are artificial neural networks that are specialized for processing grid-like topology [106]. They are biologically-inspired variants of Multi-Layer Perceptrons (MLPs) [107]. LeCun et al. introduced convolutional networks in the early 1990's [108]. A typical convolutional neural network model consists of several convolutions, pooling, and fully connected layers [92]. Each layer of CNN generates higher level abstraction of the input data called as a feature map [97]. Computations inside of CNNs are based on specialized kind of linear operations called as convolutions [92]. Goodfellow, Bengion et al. described CNNs as: "*Convolutional networks are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers.*" [106].

CNNs have demonstrated excellent performance at pattern recognition and classification [109]. CNNs improve idea of traditional machine learning systems by providing sparse interactions, parameter sharing and equivariant representation [106]. CNN is one of the neural networks that are appropriate for time series classification beside recurrent neural network (RNN) [110]. CNN is capable of learning deep features in recursive patterns and thus is good at inferring long term repetitive activities [110]. CNNs ensure some degree of shift and distortion variance due local receptive fields, shared weights and in some cases, spatial and temporal subsampling [13]. Local dependency, scale invariance and multivariate input support are benefits of CNN as compared to the other models [110, 111]. However, in some cases, local dependency may turn out to be disadvantageous because the size of convolutional kernels limits the captured range of dependencies between samples [112].

Data that can be fed to CNNs are for example 1-D grid of time series data, images as a 2-D grid of pixels or even more complex multidimensional grids such as 3D [106]. All CNNs follow the same approach no matter what the dimensionality of given input data is. The only key difference is how the feature detector also called as filter slides across the data. In 1D CNNs convolution calculations are performed in one dimensional, whereas in 2D in two and in 3D in three dimensional respectively. In 1D-CNNs computational complexity is significantly lower than in 2D or 3D CNNs. Recent studies have shown that 1D-CNNs are capable to learn challenging tasks related to 1D-signals. Often 2D CNNs require much deeper networks than 1D CNNs involving 1D-signals. 1D-CNNs are well-suited for low-cost and real-time applications due to their low computational requirements. [113]

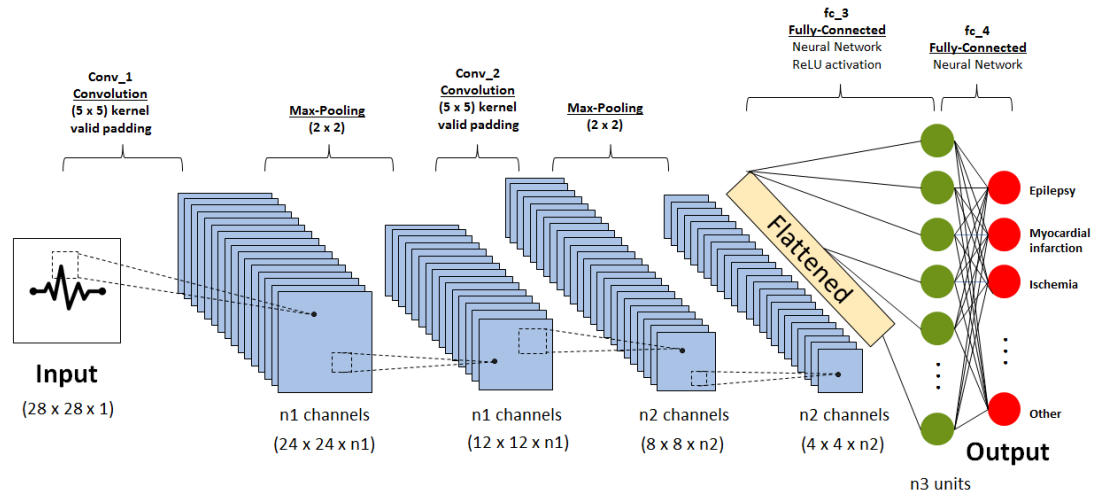


Figure 18. Example of CNN with two convolutional layers (Conv_1 and Conv_2), two max-pooling layers and two fully-connected layers (fc_3 and fc_4).

CNN consist of convolutional and fully connected layers as represented in fig 16 but there are also optional layers like nonlinear activation, pooling and normalization layers. A nonlinear activation function is often applied after each convolutional or fully-connected layer. The most commonly used activation function is rectified linear unit (ReLU) due its ability to enable fast training. Pooling layers reduce the dimensionality of feature maps. Reduction in dimensionality of feature maps enables the network to be robust and invariant to small shifts and distortions. Pooling layer can be configured based on the size of its receptive field. Strides of pooling layers determine the way how the filter convolves around the input volume. There are various types of pooling layers such as max pooling, average pooling and global average pooling. [106]

2.2.4.3. Recurrent and long short-term memory networks

Recurrent neural networks differ from feedforward neural networks in a way that they have at least one feedback loop. In a feedback loop, each neuron feeds their outputs back to the inputs of all the other neurons of a previous layer. [89] Recurrent neural networks are particularly useful when the solution depends not just on the current inputs, but on all previous inputs. During training, the recurrent neural network feeds data through a network including feedback loop from outputs to inputs until the values of the outputs do not change. It can be said that recurrent neural networks have a memory because the solution is based on the current input and the previous outputs. [114]

Recurrent neural networks like the Long Short-Term Memory network (LSTM) adds native support for input data comprised of a sequence of observations [115]. LSTM is capable to solve many time series tasks using fixed size time windows better than feedforward [115]. LSTM is capable to learn long-term correlations in sequence and thus is also able to accurately model complex multivariate sequences [116]. RNN and LSTM make use of the time-order relationship between readings and therefore they are capable to recognize short activities that have a natural order [110].

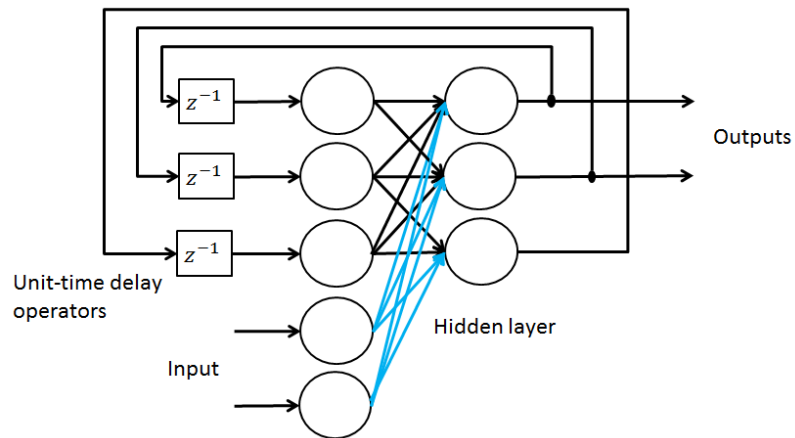


Figure 19. Example of recurrent neural network with hidden layer.

There are multiple different LSTM architectures. Vanilla LSTMs are the simplest LSTM models with memory cell of a single LSTM layer in the simple network structure [117]. Stacked LSTM networks contain multiple hidden LSTM layers stacked on top of each other into deep networks. Stacking enables the network to achieve deeper structure than Vanilla LSTM. CNN-LSTM is a model that consists of CNN model and LSTM model stacked on top of each other. CNN-LSTM model combines powers of CNN and LSTM models together. In CNN-LSTM model, CNN is used to learn features of samples, while the LSTM can be used to support sequence input and output type. Encoder-decoder LSTM means a LSTM model, which contains separate encoder LSTM, and decoder LSTM networks. Encoder LSTM maps a variable-length sequence to a fixed-length vector, while decoder LSTM network maps the vector back to a variable-length sequence. In bidirectional LSTM, the input sequence is learned both forward and backward, whereas in generative LSTM, the structure relationship of the input sequence is learned in a way that it can generate new sequences. The architectures are presented as block diagrams in figure 18. [118-122]

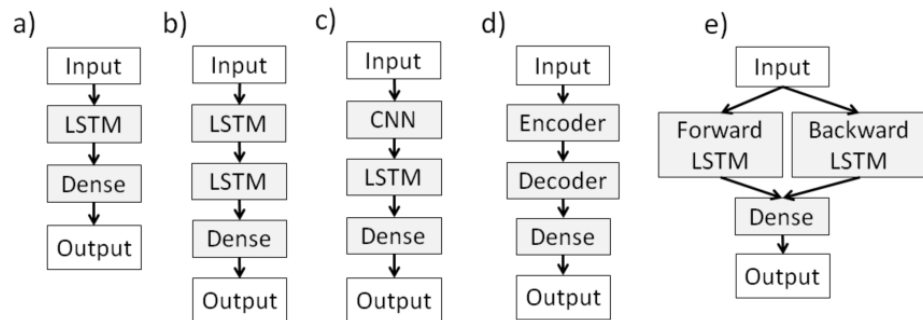


Figure 20. Block diagrams of LSTM architectures: a) Vanilla LSTM, b) stacked LSTM, c) CNN LSTM, d) Encoder-Decoder LSTM and e) bidirectional LSTM.

2.2.5. *Tuning deep neural networks*

Tuning deep neural networks means a process where different combinations of parameters and components of a network are tested to improve performance of a model. These parameters that define a model's holistic structure are called as hyperparameters. Hyperparameter are for example the number of neurons in a layer, the number of hidden layers, the activation function, the optimizer, the learning rate of the architecture, the number of epochs, batch size etc. Some of them have effect to other components and some components are more important than others. The choices of hyperparameters and components are critical for learning algorithm. However, there is no foolproof recipe to find optimal hyperparameters for a model because they are largely problem and data dependent. [96, 123]

The concept of network topology includes two components: the number of nodes in each hidden layer and the number of hidden layers. Network topology refers to the capacity of the network to fit a wide variety of functions. The capacity of the network controls whether a model is more likely to overfit or underfit. Thus, model with more layers and more nodes per layer has higher capacity than a model that has fewer of them. [124, 125]

Neural networks are trained using stochastic gradient descent that requires selection of loss function. Loss function is a measure that calculates the model error. The model error describes the performance of a model with specific weights. Loss function is the objective function that is being minimised, also called as the cost function, or error function. The selection of loss function relies on the choice of output activation function. Wrong selection of loss function may result to mislead in evaluation of model performance. [126-129]

Weight initialization means a procedure in which the weights of a model are initialized with small random values before training process [130]. There are multiple different methods to initialize the weights for a neural network like Xavier method, however yet there is no best weight initialization scheme investigated [131, 132]. Greedy Layer-wise pretraining and transfer learning can be used as initialization schemes [133, 134]. Greedy layer-wise pretraining is a technique for developing deeper neural networks by initializing weights in a region near a good local minimum [135]. Transfer learning refers to process where a model that is trained on one problem is used for another problem to improve generalization [134].

Batch size, also called as mini-batch size is the number of samples used to estimate the error before updating the model parameters. Batch size has effect on the speed and stability of the training process. Small batch size has shown to improve generalization and it needs less memory at one time during the training process. A good default value for batch size is 32 but it can be chosen between 1 and a few hundreds depending on settings and data. [136, 137]

Epochs means the number of passes through the training dataset that a model completes. It is highly important to stop training at the right time to improve generalization. Early stopping is a procedure where training is stopped at the time of smallest error with respect to the validation data set. Early stopping is perceived as one type of regularization method [138]. Criteria which stop training later on the average leads to improved generalization compared to faster stopping. However, the results of using later stopping criteria vary dramatically. Criteria which should be considered for early stopping are: no change, decrease, average change or an absolute change in metric over a given number of epochs. [139, 140]

Learning rate, also called as the step size, means the amount that each model parameter is updated per iteration [128]. It controls how quickly a model learns a problem. The learning rate is tightly related to the number of epochs and batch size. The learning rate is the single most important hyperparameter that should be tuned for a model that uses stochastic gradient descent. The range values that should be considered for the learning rate are less than 1.0 and greater than 10^{-6} . [124, 136]

There are various techniques for hyperparameter tuning to gain better model performance. Manual search means a technique to select manually by hand the best candidate value for the hyperparameters in a model. Good starting point is to use default values of a model and change the hyperparameters of a model based on the model performance. Grid search means a technique that tests grid-like combinations of the provided values of each hyperparameter. In this technique disadvantage is that the grid does not have always good hyperparameter combinations. Random search is a technique where hyperparameter combinations are selected randomly. This technique provides bigger hyperparameter space and thus much higher probability of selecting better hyperparameter combinations to improve performance of a model. Bayesian optimization is a technique that accelerates hyperparameter selection. This approach chooses hyperparameter combinations based on the previous evaluations and focuses on those areas of the hyperparameter space that possibly brings the most promising testing scores. [96, 141, 142]

2.2.1. Deep learning workflow in signal processing

Deep learning neural networks are especially used for image classification, but they also fit well for signal processing tasks. Neural networks are capable to process 1D signal, time-series data, or even text as input data. The main steps in the deep learning workflow are importing data, preprocessing data, configuring network architecture, training network and checking network accuracy. Configuration of network architecture is iterative, because the objective is to find the best network architecture that produces the highest accuracy for test data. Test data need to contain similar data than data in the deployment environment. [143]

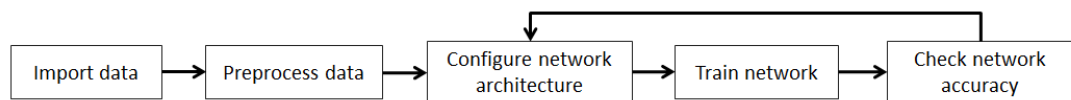


Figure 21. Deep learning workflow.

Typically deep learning does not require much preprocessing and feature extraction, so often raw image is passed into the network to get results. Ideal deep learning does not work with signals because most often raw signal data are noisy, variable, and smaller than image data. Thus, raw signal data require careful preprocessing and sometimes feature extraction to ensure the best possible results. Possible preprocessing tasks are for example, removing signals with constant values, normalizing training predictors, or denoising. Classical signal processing filters are still used in preprocessing but the challenges with these approaches are that they filter out also some information of the desired signal along with noise. Denoising

with neural networks has proved to be a good selection to enhance quality of signals, for example to increase the noise-to-signal ratio. [143]

Data scaling is recommended pre-processing step because unscaled inputs may cause a slow and unstable learning process. Input variables may have different units which may mean the variables have different scales. Input values should be small values in the range of 0-1 or standardized with a zero mean and a standard deviation of one. If the distribution of data has shape of a normal distribution, then it needs to be standardized, otherwise the data should be normalized. Moreover, the output variables must be scaled to match the scale of the activation function of the output layer of a neural network. [123, 144, 145]

Time series forecasting is one of the most challenging tasks in data mining [146]. Often time series data are highly imbalanced and the most interesting predictive samples are the least represented [147]. Concrete example is epilepsy data where recordings may have length of multiple hours, but epileptic seizure may last only a few seconds to several minutes [23 p. 15]. In these types of scenarios, learning algorithms bias the models towards the majority class leading to performance degradation [147]. There are various resampling strategies to solve this type of problem such as random under sampling, random oversampling and random under sampling with synthetic minority over-sampling technique [148-150]. Beside resampling techniques, another way to deal with imbalanced data is cost sensitive learning [151]. Resampling strategies and cost sensitive learning can significantly improve the predictive evaluation of the models [152].

To prepare signal data for neural networks the signal data are often transformed into spectrograms, the wavelets or Fourier's transforms. A spectrogram is a time-frequency transformation that represents raw signal data as a 2D image. Wavelet transformation is similar to spectrograms but includes time resolution for non-periodic signal components. Wavelets are also used for feature extraction called as wavelet scattering or invariant scattering convolutional networks. Short-time Fourier Transform (STFT) transforms the signal into a 2D signal in frequency domain. The technique allows breaking the signal into overlapping segments to enable see predictions while the signal is being acquired. [143]

3. RELATED WORK

Over three decades, research has attempted to solve the problem of EEG-based epileptic seizure detection to help clinicians in their work. Therefore, a huge number of research papers are published for identification of epileptic seizure. The published research papers can be categorized into three main classification problems. The first and the easiest one compared to other problems is to distinct classes: normal and ictal EEG patterns. The second problem is to distinct normal, inter-ictal and ictal EEG patterns. The third and most challenging is to address discrimination between classes: normal, ictal, inter-ictal, post-ictal and pre-ictal EEG patterns. This section only focuses on the literature related to two class classification problem to differentiate normal and ictal EEG patterns. [153, 154]

There are huge amount of different approaches and techniques how to develop an epileptic seizure detector. From machine learning point of view in 2005, Aarabi et al. [155] proposed an automatic epileptic seizure detection system using a set of representative EEG features. They extracted features from time domain, frequency domain, wavelet domain to feed together along with auto-regressive coefficients and cepstral features into a back-propagation neural network (BNN). This research setup achieved an average classification accuracy of 93.00%. In 2007, Polat et al. [156] achieved higher accuracy of 98.68% using fast Fourier transform (FFT) as an input for a decision tree classifier. In same year, Subasi et al. [157] achieved a comparable accuracy of 94.50% by using wavelet transform to derive EEG frequency bands as an input for the mixture of experts (ME) classifier. [155-157]

In 2008, Mirowski et al. [158] compared for the first time convolutional neural networks with SVM to predict epileptic seizure. Mirowski et al. concentrated on aggregated features that encode the relationship between pairs of EEG channels, such as cross-correlation, nonlinear interdependence, difference of Lyapunov exponent and wavelet analysis-based synchrony such as phase locking. They achieved a zero-false-alarm seizure prediction on 20 patients out of 21, compared to 11 only using SVM. [158] One year later, Chandaka et al. [159] used support vector machine (SVM) along with three statistical features as an input. Features were computed via cross correlation coefficients. The model achieved seizure detection accuracies of 95.96%. [158, 159]

Yuan et al. [160] used ELM and nonlinear dynamic features extracted from EEG in 2011. They used features such as approximate entropy (ApEN), Hurst exponent and scaling exponent obtained by detrended fluctuation analysis (DFA). ELM algorithm resulted satisfactory recognition accuracy of 96.5%. Khan et al. [161] used simple linear discriminant analysis (LDA) along with discrete wavelet transform (DWT). This study reached to an accuracy of 91.80%. In same year, Nicolaou et al. [162] used Permutation Entropy (PE) as an input for SVM. PE and SVM combination resulted accuracy of 93.80%. Zhou et al. [163] used lacunarity and Bayesian Linear Discriminant Analysis in intracranial EEG and achieved accuracy of 96.67%. In 2014, Kumar et al. [164] disband the EEG signals into delta, theta, alpha, beta and gamma bands via wavelet transform. Statistical and non-linear features were extracted from these bands and fed into a SVM classifier. Kumar et al. achieved an accuracy of 97.50% with this setup. In 2016, Song et al. [165] used SVM, along with weighted-permutation entropy and achieved an accuracy of 97.25%. In 2016, Bugeja et al. [166] used multilevel wavelet transform as a feature extractor and fed them into ELM classifier. [160-162, 165, 166]

In 2017, Hosseini et al. [167] used multimodal rs-fMRI and EEG with CNN classifier to localize an epileptogenic area. The result of this study was normal p-value $1.85e-14$ and p-seizure value $4.64e-27$. In the same year, Arharya et al. [168] used 13-layer deep 1-D convolutional neural network to detect normal, preictal and seizure classes. They normalized data with z-score normalization and set the sampling rate of EEG signal to 173.61 Hz. The technique achieved an accuracy, specificity, and sensitivity of 88.67%, 90.00% and 95.00%, respectively. [167, 168]

In next year 2018, Hussein et al. [169] used 1-D fully connected three layer LSTM network along with softmax classifier to detect epileptic seizure. The network included also time distributed dense layer and average pooling layer. They achieved convincing results of sensitivity, specificity, and accuracy of 100% each in two-class problems, three-class problem and five-class problem. Approaches of this study were evaluated over the University of Bonn. Recordings of Bonn EEG data set contain only one channel EEG and have no artefacts. Thus, Bonn dataset differs a lot compared to real life EEG data. In the same year, Liang et al. [170] tried a deep learning classifier to identify seizures through the EEG images. They justified their approach with a visual inspection to identify epileptiform abnormalities and achieved reasonably good result, with 84% for sensitivity, 99% for specificity, and 99% for accuracy. [169, 170]

The recent deep learning approaches developed in [5, 171-174] are based on neural network architectures such as CNN, RNN, LSTM and GRU. CNN is suitable for two or more dimensional data, but not suited for one dimensional sequence data. RNN suffers from the gradient vanishing. LSTM and GRU improved on RNN, but they both still have the gradient decay problem. In 2019, Yao et al. [175] proposed promising a variety of RNN called IndRNN to detect epileptic seizure. The study obtained reasonably good results, the average sensitivity, specificity, F1-score, precision and accuracy are 88.80%, 88.60%, 88.71%, 88.69%, and 88.70%, respectively on CHB-MIT dataset. This approach exceeded the LSTM approach [5] and the CNN approach [171] with an improvement of at least 4%.

4. METHODS

4.1. Designed system

The development of neural networks is typically time consuming task because of the large amount of hyperparameter combinations. It is convenient to use publicly available tools or build up a system that covers the whole deep learning workflow pipeline. This fact led to the development of a system that attempts to include the main core of the deep learning workflow from pre-processing to prediction visualization. The designed system contains python modules pre-process, hyperparameter search and evaluation, MySQL database for storing data and Jupyter notebook for visualizing predictions and evaluation metrics. The sequence diagram of the system is illustrated in Figure 20.

Pre-processing module contains pre-processing and segmentation functionality. Pre-processing module produces hdf5 file and saves it on disk for hyperparameter search. Hyperparameter search module loads data from hdf5 file in batches and performs hyperparameter optimization for selected model type such as VGG, VGG-LSTM, DenseNet 1D etc. Model scripts are located in hyperparameter search module. Hyperparameter search uses k-fold cross validation to evaluate model performance on a subset of the dataset. In the hyperparameter search the subset of dataset contains only recordings which included epileptic seizure events. Thus, recording that included only normal EEG data was left out in this study. The subset of dataset was used to decrease hyperparameter tuning time. The developed system uses MySQL database for storing evaluation metrics and model related information. The structure of the database is described in Figure 21.

After each experiment, all evaluation metrics are stored into database. The 5-folds cross validation was used to evaluate a model. If F1-score of test dataset of the evaluated model does not reach over given threshold (for example $F1\text{-score} > 0.75$) then hyperparameter search continues to a new experiment without looping the remaining folds. A model along with trained weights is stored in a folder named with the present experiment. The hyperparameter search module uses Bayesian optimization to optimize a network structure and configurations.

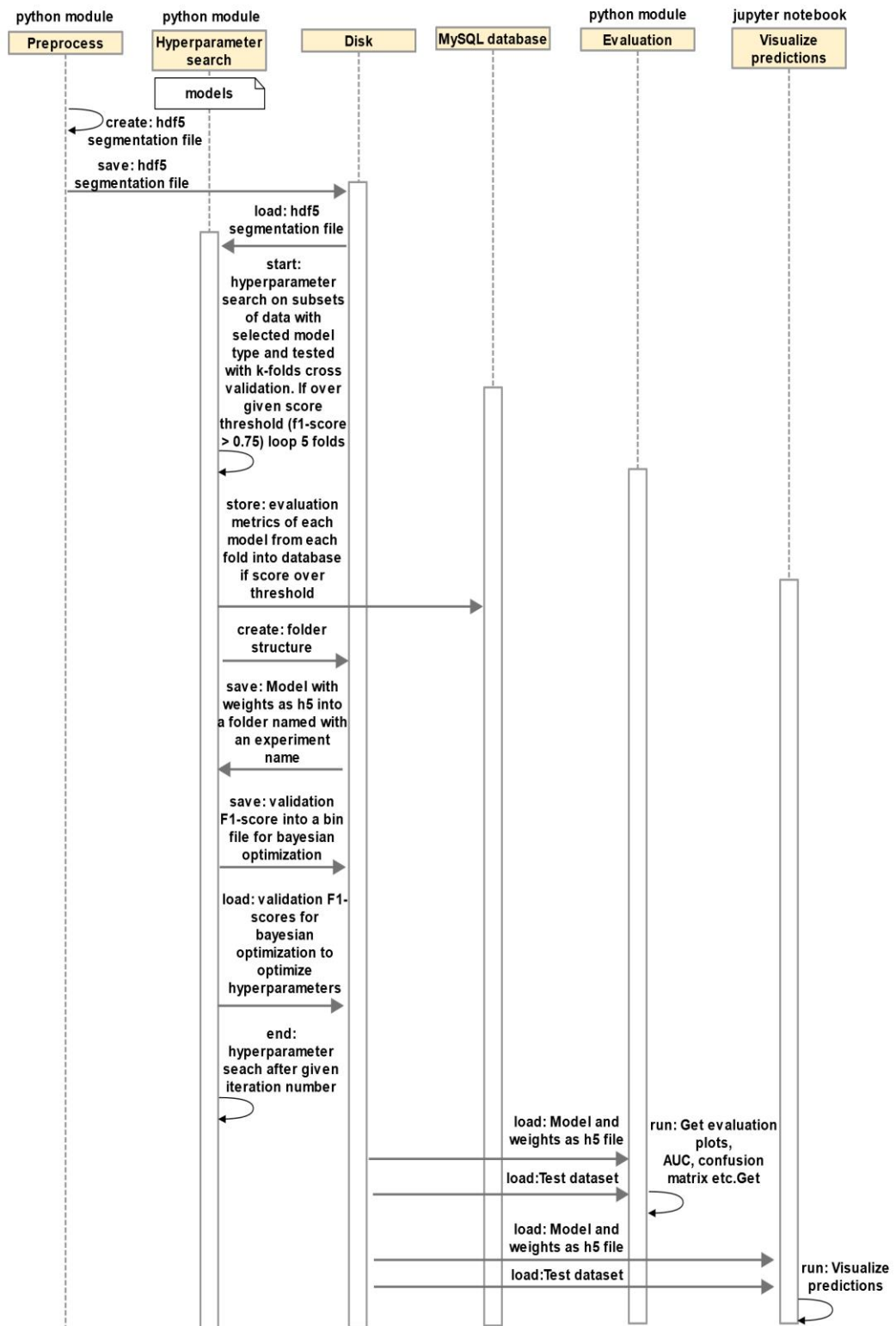


Figure 22. Sequence diagram.

The designed system saves metrics and other valuable data throughout the machine learning pipeline. The database structure of the system is illustrated in Figure 23. The *experiment* table is a parent table of the *fold* table. The *fold* table is a

parent table of *epoch_point*, *confusion_matrix*, *metric* and *model* tables. The column *experiment_id* is a primary key of the *experiment* table and creates a relationship with the child tables. Therefore, *experiment_id* column in each child table is a foreign key of the child tables. The *experiment* table contains information about single experiment. The child table *epoch_point* contains *accuracy* and *loss* of each epoch. The *epoch_point* table is used to visualize training and validation curves. The child table, *confusion_matrix* contains a confusion matrix of single fold calculated from test results. The table *metric* is a derivation table calculated from *confusion_matrix* table. The table *model* contains a model *file_name* that matches with a model file located in a disk described in the sequence diagram in Figure 22.

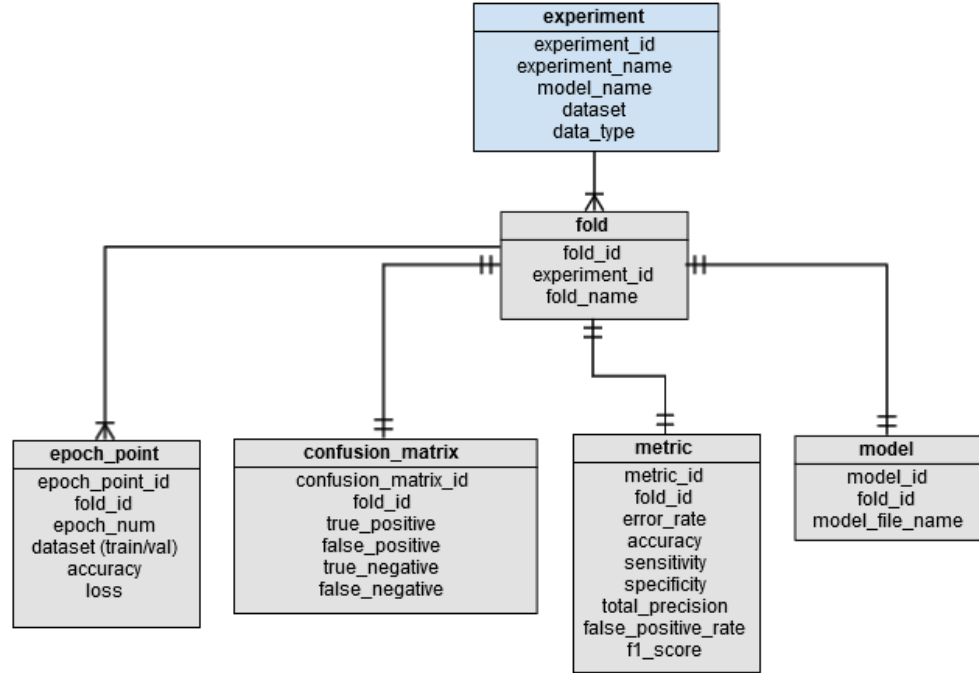


Figure 23. Database structure.

4.2. Input formats

In this study, EEG signal, 1D-FFT, 2D-FFT, 2D-FFT magnitude spectrum, 2D-FFT phase spectrum and magnitude spectrograms were used separately in different experiments as an input for a neural network. Raw signals have been widely used as an input for a neural network but often a signal is transformed into a frequency-domain. The EEG signal is a time-domain signal which can be transformed into frequency-domain signal through Fourier transform. The power of the Fourier transform has found wide use in diagnostics, speech and audio processing based on mechanical vibration [176, 177]. The discrete Fourier transform (DFT) is one type of the FFT and it converts a finite signal into a same-length complex valued frequency-domain signal. The DFT can be computed by:

$$f(m) = \frac{1}{n} \sum_{k=0}^{n-1} F(k) \exp(2\pi i \frac{mk}{n}) \quad (2)$$

for $m = 0, \dots, n - 1$ [178 p.400, 179]. When FFT is applied on the signal, the result is a set of complex numbers. The complex numbers contain a real part and an imaginary part. The magnitude and phase can be calculated by the real and imaginary parts of the complex numbers. The magnitude is similar to the amplitude of the signal, but does not contain the direction information thus it has only positive values. The magnitude of a variable is the measure of how far the quantity differs from zero. Respectively, the amplitude of a variable is the measure of how far and what direction the quantity differs from zero. Figure 23. In below illustrates the relations between the real and the imaginary parts of the complex numbers, and the magnitude and the phase. [180, 181]

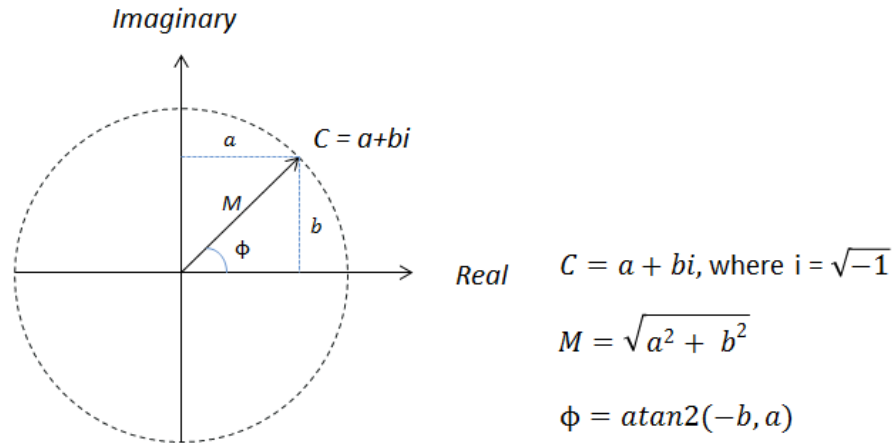


Figure 24. Illustration of relations between the complex numbers and the imaginary part, the real part, the magnitude and the phase. C = complex numbers, a = real part, b = imaginary part, i = imaginary unit = $\sqrt{-1}$, M = magnitude and ϕ = phase.

A spectrogram is a visual representation of the spectrum of frequencies of a signal varied within time generated by the fast Fourier transform (FFT). Spectrograms have been used in the analysis of complex sleep apnea to recognize oscillation of heart rate variability and detection of subtypes of epilepsy. Once the FFT is applied to short fixed windows of the signal in time, the result is known as the short-time Fourier transform (STFT) of the signal. SFFT is an algorithm that is used for converting a time-domain signal in fixed segments into an equivalent frequency-domain signal. The spectrogram is built by stacking a sequence of spectra together in time. Spectrogram has frequency along the vertical axis, time along the horizontal axis, and the amplitude is shown as a gray level. [178 p.400-402, 182 p. 478-481, 183, 184]

The Fourier transform can be performed in higher dimensions [185]. In this study, two-dimensional FFT was investigated as one input option for neural networks. The 2D discrete Fourier Transform (DFT) of f , denoted by $F(k, l)$, is given by:

$$F(k, l) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \exp(-2\pi i (\frac{x}{M}k + \frac{y}{N}l)) \quad (3)$$

for $k = 1, 2, 3, \dots, M-1$ and $l = 0, 1, 2, \dots, N-1$. The 2D FFT has been used for a long time in image processing. Significant solution that uses the 2D-FFT is MRI. 2D FFT is needed to reconstruct a 2D image in MRI. The first FFT transformation converts MR signals to frequency projection and the second FFT transformation converts frequency projections to 2D image [186]. In image processing the 2D FFT is used in image compression, sharpening, edge detection and smoothing. A two-dimensional fast Fourier transform is carried out on by executing a one-dimensional fast Fourier transform on all vertical lines of sample points, storing the first FFT transformation values at one or more specified positions in each vertical line in an internal buffer, and then executing a one-dimensional fast Fourier transform on each resulting horizontal line of transformed data. [185]

A 2D-FFT contains real part, imaginary part, magnitude and phase components just like a 1D-FFT. This is because performing a 2D-FFT is actually equivalent to performing two 1D transforms as:

1. Performing a 1D transform on each row of image $f(m, n)$ to get $F(m, l)$
2. Performing a 1D transform on each column of $F(m, l)$ to get $F(k, l)$.

A 2D-FFT provides shift and scale properties just like the 1D-FFT. There is a relationship that can be derived from shifting and image in one domain to the other. Shifting means rotation around the boundaries. Scale property means that as an object grows in the image, the corresponding features in the frequency domain will expand. The 2D-FFT provides one property that the 1D-FFT does not contain and that is rotation property. In the 2D-FFT, the content in the frequency domain is positioned based on the spatial location of the content in the space domain. This means that rotating the spatial domain contents rotates the frequency domain contents. [187]

4.3. Used CNN architectures

In this study, VGG, DenseNet 1D, VGG-LSTM and 3D convolutional neural network were tested, but only results from VGG and DenseNet 1D were reported. VGG is very deep convolutional network ConvNet architecture that is specialized for large-scale image recognition. It was developed by Karen Simonyan and Andrew Zisserman in ImageNet ILSVRC-2014. The aim of the VGG project was to investigate how the depth of the ConvNet affects the accuracy in the large-scale image recognition setting. It was investigated that the depth of neural networks is beneficial for classification accuracy. Original VGG contribution has 16-19 weight layers by means of very small (3x3) convolutional filters. The work showed that significant improvement on the prior art configurations can be achieved by very small filters with pushing the depth to 16-19 weight layers. The design of VGG

architecture was inspired by [14, 188]. VGG was used in the present work for feature extraction. [15]

Table 2. VGG-19 [15]

Layers	Output size	VGG 19
Block (1)	224×224	$[3 \times 3 \text{ conv}, 64] \times 2$
	112×112	$2 \times 2 \text{ max pooling}$
Block (2)	112×112	$[3 \times 3 \text{ conv}, 128] \times 2$
	56×56	$2 \times 2 \text{ max pooling}$
Block (3)	56×56	$[3 \times 3 \text{ conv}, 256] \times 3$
	28×28	$2 \times 2 \text{ max pooling}$
Block (4)	28×28	$[3 \times 3 \text{ conv}, 512] \times 3$
	14×14	$2 \times 2 \text{ max pooling}$
Block (5)	14×14	$[3 \times 3 \text{ conv}, 512] \times 3$
	7×7	$2 \times 2 \text{ max pooling}$
Fully connected layer	1×1	Dense 4096
Fully connected layer	1×1	Dense 4096
Fully connected layer	1×1	Dense 1000
Classification layer	1×1	Softmax

In 2017, DenseNet was introduced in Densely Connected Convolutional Networks paper. The paper received Best Paper Award at the IEEE Conference on Computer Vision and Pattern recognition (CVPR). The primary developer was Gao Huang. DenseNet achieved state-of-art performances across several highly competitive datasets. DenseNet required substantially fewer parameters and less computation to achieve higher results than ResNet. DenseNet needs fewer filters in each layer since each layer receives feature maps from all preceding layers. DenseNet handles the error signal better by propagating it to earlier layers more directly. Through propagation a network maintains low complexity features and a strong gradient flow. In this work DenseNet was used as an option for feature extractor from raw EEG signal beside VGG. [17]

Table 3. DenseNet-121 [17]

Layers	Output size	DenseNet-121
Convolution	112×112	$7 \times 7 \text{ conv, stride 2}$
Pooling	56×56	$3 \times 3 \text{ max pool, stride 2}$
Dense block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition layer (1)	56×56	$1 \times 1 \text{ conv}$
	28×28	$2 \times 2 \text{ average pool, stride 2}$
Dense block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition layer (2)	28×28	$1 \times 1 \text{ conv}$
	14×14	$2 \times 2 \text{ average pool, stride 2}$
Dense block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition layer (3)	14×14	$1 \times 1 \text{ conv}$

	7×7	2×2 average pool, stride 2
Dense block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Classification layer	1×1	7×7 global average pool

In this paper VGG-LSTM was tested but not reported in the paper. VGG-LSTM achieved poor results on the CHB-MIT dataset because of lack of time. General CNN-LSTM has been developed for video analysis. It is widely used in emotion and visual recognition from videos. [189, 190] Shahbazi et al. used CNN-LSTM with short-time Fourier transform (STFT) input for seizure prediction. The task of CNN is to extract high-level time frequency from the samples. The LSTM cell of the CNN-LSTM model uses the temporal trajectory of these features. [18]

3D convolutional neural network was also tested in this study but was not reported in the paper. 3D convolutional neural network achieved poor results mainly because of lack of time. In 2016, 3D convolutional neural network made great progress in dealing with various video analysis tasks. 3D convolutional neural network is capable to handle appearance and motion information simultaneously. The features with a linear classifier can achieve good results on different video analysis. [20]

4.4. Bayesian optimization

Bayesian optimization is an approach to automatically optimize objective function by finding a global minimum of the function. Bayesian optimization typically works by assuming the unknown function that was sampled from a Gaussian process. In machine learning Bayesian optimization is used for tuning hyperparameters. Hyperparameters control the rate of learning and the capacity of the underlying model. The Bayesian optimization approach should be considered in a situation where the number of tuning hyperparameters is high and the evaluation of a model takes a long time. [191, 192].

Bayesian optimization constructs a probabilistic model for $f(x)$ and then exploits this model to make decisions towards the global minimum of the function. The procedure relies on the information gathered from previous evaluations rather than local gradient and Hessian approximations. A prior over functions must be selected before applying Bayesian optimization. A prior could be for example accuracy, loss or etc. which will express assumption about the function being optimized. Secondly, acquisition functions must be determined to deploy the model to decide the next point to evaluate. [192]

4.5. Validation methods and evaluation metrics

Designed system used 5-fold or leave-one-out cross validation method. Used validation method depended on F1-score computed on test data. The system executed more folds if the F1-score exceeded given threshold which was 75%. 5-fold cross validation and leave-one-out cross validation are illustrated in fig 24. In both cross validations at first test data is isolated from training data for final evaluation. In 5-fold cross validation training data is split 5 times into validation and training data. The performance measure reported by 5-fold cross validation is the average of the values computed in the loop. Leave-one-out means that data is split into training and

validation dataset only once. The performance measure is therefore the measure computed in one loop. Leave-one-out approach may have a high bias if there is a limited amount of data. Thus, k-fold cross validation provide results in a less biased model compared to other methods. [193, 194]

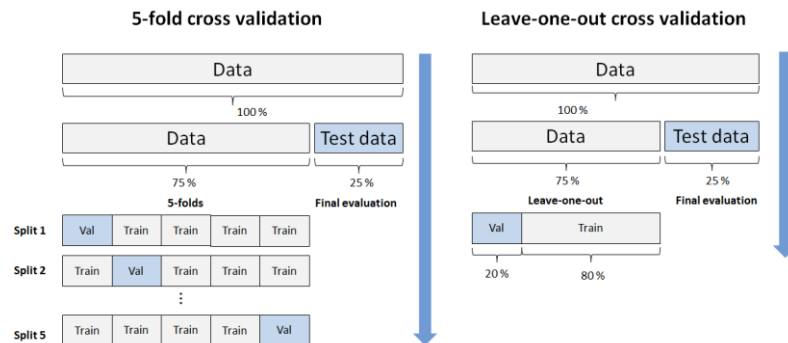


Figure 25. 5-fold cross validation in the left and leave-one-out cross validation on the right.

Prediction models have four results: true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) which can be presented by a confusion matrix. In Figure 26. is a confusion matrix which includes information about actual and predicted observations carried by a binary classification algorithm. The performance of classification algorithm is evaluated using the data in the confusion matrix. Results of confusion matrix can be described as the following:

- True positive (TP) is a number of positive cases that were correctly identified
- False positive (FP) is a number of negative cases that were incorrectly classified as positive
- True negative (TN) is a number of negative cases that were correctly classified
- False negative (FN) is a number of positive cases that were incorrectly classified as negative. [195]

		Actual values	
		Positive	Negative
Predicted values	Positive	TP	FP
	Negative	FN	TN

Figure 26. Confusion matrix.

The designed system saved multiple evaluation metrics into a database for post analysis. The evaluation of the model performance was performed on a testing dataset which was 25% of all data. During training accuracy and loss were computer from the training and validation datasets and saved into database simultaneously. Used loss function was binary cross entropy. Because of imbalanced data the accuracy of overall dataset was not valid evaluation metric for the epilepsy detection.

During training class weights were calculated and given to a model. The class weights affects only to loss function not to accuracy. [195]

Accuracy is not a valid evaluation metric for detection problem where data is imbalanced because accuracy stays high even if a model is able to detect all normal segments but not any rarely appearing segments. The formula of accuracy is described in Equation 4. The error rate is computed as the number of all incorrect predictions divided by the total number of the dataset. Simply it can be computed as $1 - \text{accuracy}$. Similar to accuracy error rate should not be used to evaluate model performance on imbalanced dataset. The formula of the error rate is described in Equation 5. [195, 196]

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} = \frac{TP + TN}{P + N} \quad (4)$$

$$\text{Error rate} = \frac{FP + FN}{TP + TN + FN + FP} = \frac{FP + FN}{P + N} = 1 - \text{accuracy} \quad (5)$$

In the evaluation phase precision, sensitivity, specificity, F1-score, false positive rate, error rate and accuracy were computed from the test dataset. Precision is a percentage of positive instances out of the total predicted positive instances. Dominator of true positive instances is the model prediction done as positive from the whole dataset. In other words: *“how much the model is right when it proposes to be right”*. The formula of precision is described in Equation 6. [195, 196]

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

Sensitivity also called as recall or true positive rate describes a percentage of positive instances out of the total actual positive instances. In the present study, sensitivity word is used. In sensitivity the dominator of true positives is the actual number of positive instances present in the dataset. In other words: *“how much extra right ones, the model missed when it predicted the right ones”*. The formula of sensitivity is described in Equation 7. [195, 196]

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{TP}{P} \quad (7)$$

Specificity also known as the true negative rate is a percentage of negative instances out of the total actual negative instances. In this study, specificity word is used in this context. In specificity the dominator of true negatives is the actual number of negative instances present in the dataset. It is similar to sensitivity, but specificity regards negative instances instead of positive ones. Simply put: *“how many healthy patients were not having epilepsy and were told they don’t have epilepsy”*. The formula of specificity is described in Equation 8. [195, 196]

$$\text{Specificity} = \frac{TN}{TP + FP} = \frac{TN}{N} \quad (8)$$

F1-score is the harmonic mean of precision and sensitivity. F1-score is high if both precision and sensitivity are high and if precision or sensitivity goes down then F1-score also goes down. Otherwise speaking F1-score is high if the positive predicted are actually positives and a model does not miss out on the positives and predict them negative. In some application precision or sensitivity needs to be higher than another, thus in this type of cases F1-score cannot be the measure to follow. The formula of F1-score is described in Equation 9. [195, 196]

$$F1 - score = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (9)$$

False Negative rate (FNR) or Miss Rate is computed as the number of incorrect negative predictions divided by the total number of actual positives. False Positive Rate (FPR) or Fall-out describes the probability of false alarms. It is computed as the number of incorrect predictions divided by the total number of actual negatives. The formulas of the False Positive Rate and True Positive Rate are described in Equation 10 and Equation 11 respectively. [195, 196]

$$\text{False Negative Rate} = \frac{FN}{FN + TP} \quad (10)$$

$$\text{False Positive Rate} = \frac{FP}{FP + TN} \quad (11)$$

In situations, where data is imbalanced and both precision and sensitivity/recall needs to be taken into account it is recommended to use precision-recall curve. The precision-recall curve is the curve between precision and recall of various threshold values. A receiver operating characteristics curve (ROC curve) is a plot that illustrates performance measurement for binary classification tasks. The x-axis of the ROC curve indicates the false positive rate and the y-axis indicates the true positive rate. [197]

ROC curve is appropriate measurement only when the observations are balanced between each class, whereas precision-recall curve is appropriate for imbalance data. Davis et al. used highly imbalanced data in their study and the authors noted that the true positive rate should be used with precision instead of with the false positive rate to evaluate performance. AUC is the area under the curve. It can be computed from both ROC and precision-recall curve. AUC-PR is a more suitable measure for imbalanced data than AUC-ROC since it is derived from precision and recall instead False Positive Rate and True Positive Rate. [197]

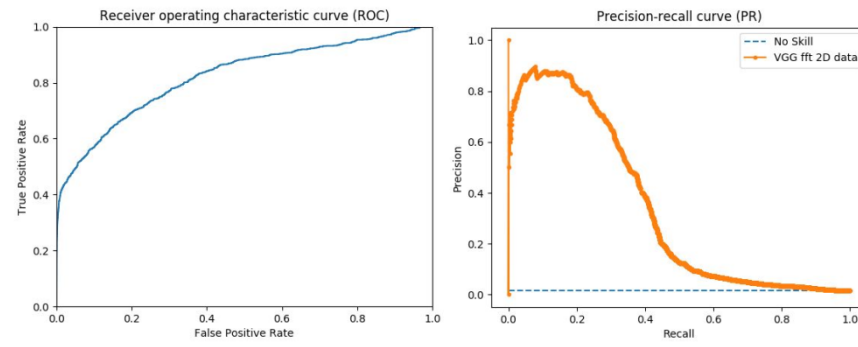


Figure 27. Receiver operating characteristics curve on the left and precision-recall curve on right. AUC is the area under the curve.

5. IMPLEMENTATION

5.1. Workflow

5.1.1. Data preparation

In the present study, an open-source scalp EEG database, provided by the Massachusetts Institute of Technology (CHB-MIT) was used. The recordings of CHB-MIT database were collected from 23 patients diagnosed with epilepsy. The database included recordings from 17 females (age from ~1.5 to 19 years) and five males (age from 3 to 22 years) (Appendix 12). Case chb21 was obtained 1.5 years after case chb01 from the same female subject. Case chb24 was added later to the collection without personal information. All subjects were requested to stop medicine related to epilepsy one week before data collection. A sampling frequency of the data was 256Hz with 16-bit resolution and the recordings were in European Data Format (EDF). The recordings included 198 seizures. The beginning and end of each seizure were marked in the .seizure annotation files.

Data preparation is the process of cleaning and transforming data for machine learning or analytics. In CHB-MIT database, EEG montage types varied between recordings and some of them included polygraph channels such as ECG, VNS, LOC-ROC, LUE-RAE, EKG1-EKG2, EKG1-CHIN. The majority of the recordings were recorded in a longitudinal montage, but a few of the recordings were in a common reference montage or had only the direct signals from each electrode. All the data were transferred to the international 10-20 system of electrode placement in bipolar longitudinal montage (Figure 30.) by selecting particular channels and then calculating signal derivatives. Some recordings had duplicate channels, empty channels and channels that do not belong to the 10-20 system. The transformation was needed to make the montages uniform in the collection.

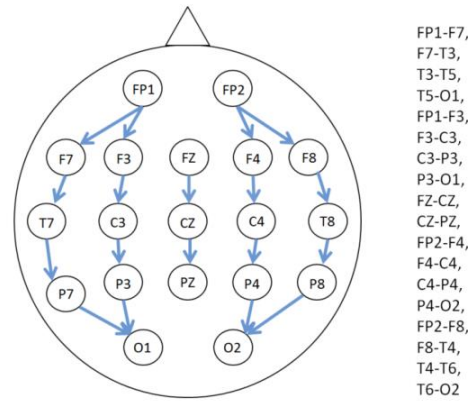


Figure 28. The 10-20 system of electrode placement in bipolar longitudinal montage.

In machine learning, data needs to be scaled into a common scale, without distorting differences in the ranges of values. Data normalization scales variables to a specific range, while standardization transforms data to have a mean of zero and a standard deviation of 1. This standardization is called as z-score standardization.

Both of these techniques were experimented in the present study. Identification of outliers in EEG signals for machine learning is important difficult task. Most of the EEG recordings were contaminated by power line noise at 60 Hz for the CHB-MIT dataset. In the frequency domain, it is convenient to effectively remove the power line noise by excluding components of 57–63 Hz and 117–123 Hz for a power line frequency of 60 Hz. The DC component (at 0 Hz) was also removed in some experiments. Based on this fact, effect of simple digital filters (high pass and low pass) and standardization algorithm that is robust to outliers was examined.

One important parameter in data preparation is a segment length of the input sample. In this context, it is vital to keep the length of the segment long enough. Short segments enable an algorithm to detect more closely the start and the end of a seizure but in other hand, a small segment might not have enough features for an algorithm to detection (Figure 32.). Segment length of 1 second, 2 seconds, 5 seconds, 10 seconds and 1 minute were experimented. Eventually, a length of the segment was selected based on a literature regarding to the duration of a seizure. A seizure may last from seconds to minutes or even hours depending on seizure type. For example, atonic seizure may last only 1 to 2 seconds, whereas auras or focal onset impaired awareness seizures may last up to 2 minutes. Some of the recordings did not split equally into given segment length. Too short segments were discarded from the data.

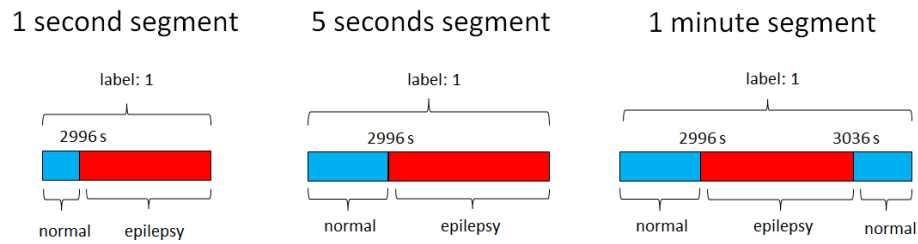


Figure 29. 1 second, 5 second and 1 minute segments labeled as 1 to indicate epilepsy segment. Seizure starts at 2996 seconds and ends at 3036. Smaller segment length gives more accurate start point of epileptic seizure, but has a less partition of an epileptic seizure segment than longer segment.

5.1.2. Feature engineering

Choosing a suitable representation of the input data is a major part of the machine learning process and effects on a model selection. In the present study, magnitude spectrogram was the first representation of the input data (Figure 30.). The transformation transforms the EEG signal from time-domain signal to the frequency-domain representation of the signal. Selection of spectrograms influenced on data scaling process. Before EEG signals were transformed to magnitude spectrograms, EEG signals were standardized with standardization algorithm that is robust to outliers. After spectrogram transformation, spectrograms were scaled with a logarithmic function with a small constant. A small constant was used to prevent logarithmic function to get zero values. Magnitude spectrograms as representation of the input data were used only in the first experimental group.

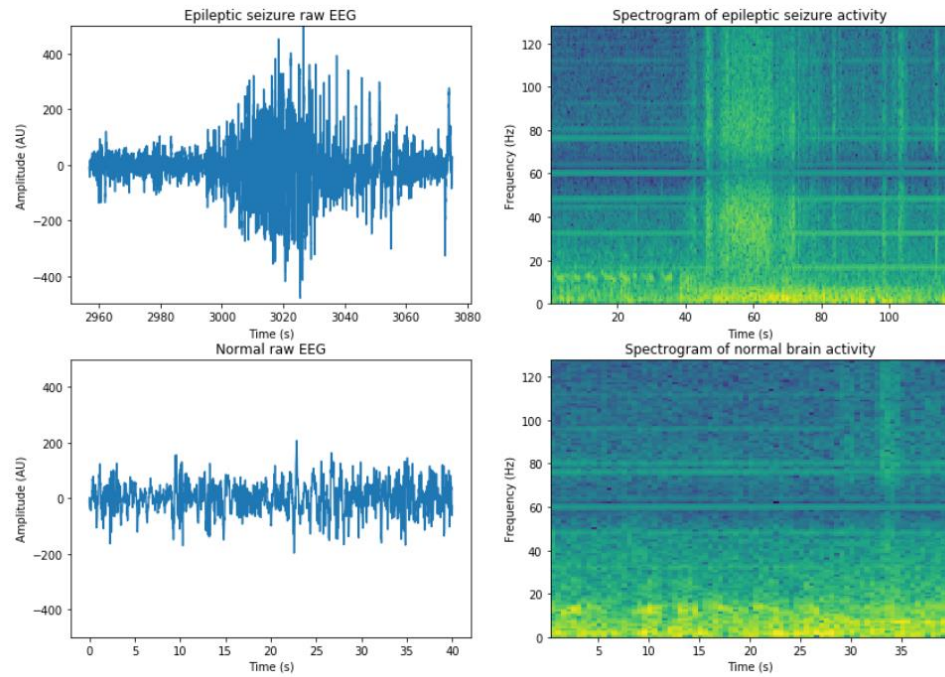


Figure 30. On the top left epileptiform activity in EEG signal, on the top right epileptiform activity of the same epileptiform activity in magnitude spectrogram, on the bottom left normal brain activity in EEG signal and on the bottom right the same normal brain activity in magnitude spectrogram (case chb01_03).

EEG signals in time-domain representation was one candidate representation of the input data (Figure 29.). Time-domain EEG signals were examined in each experimental group testing all normalization techniques presented in this study. The effect of the standardization algorithm that is robust to outliers was examined in the first experiment. In the next experiment bandpass and low pass filters along with min-max normalization was examined. Finally, in the last experiment the effect of bandpass and low pass filters along with z-score normalization was investigated. Selection of data scaling technique depends on the data. The purpose of going through all these experiments was to find a right scaling technique for the dataset.

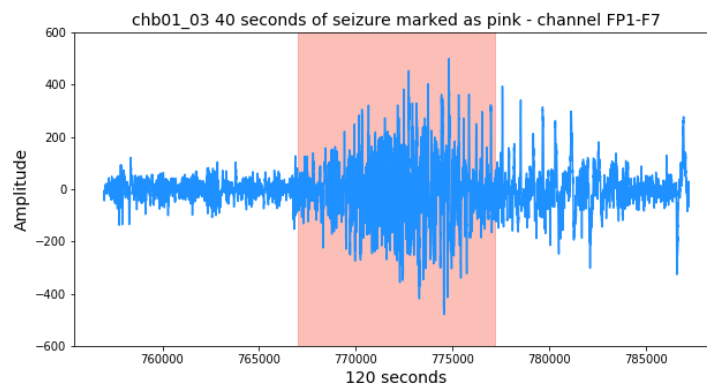


Figure 31. 120 seconds of epileptiform activity in channel FP1-F7 of case chb01_03.

A 1D-FFT was another representation candidate of the input data. The DFT needed data segmentation before the transformation to frequency-domain because the transformation disposes time information. A 1D-FFT was used as representation of the input data only in one experiment. In this experiment a bandpass and a low pass filtering and min-max scaling was utilized before the DFT transformation. The results of a 1D-FFT were compared with the results of a 2D-FFT.

A 2D-FFT was the most experimental representation candidate of the input data of the other candidates for a detection of an epileptic seizure (Figure 31.). The purpose of examining a 2D-FFT was to investigate does spatial information improve the detection results. A 2D-FFT has spatial shift and rotation property that are commonly used in image classification tasks. Similar to 1D-FFT preprocessing, data segmentation was needed to do before the transformation into frequency-domain signals. In the first experiment a bandpass and a low pass filtering and min-max scaling was applied before the DFT transformation in 2D. In the last experiments a bandpass and a low pass filtering and z-score scaling was used before the transformation. In the last experiments, it was investigated how a magnitude and a phase of a 2D-FFT effect on detection results.

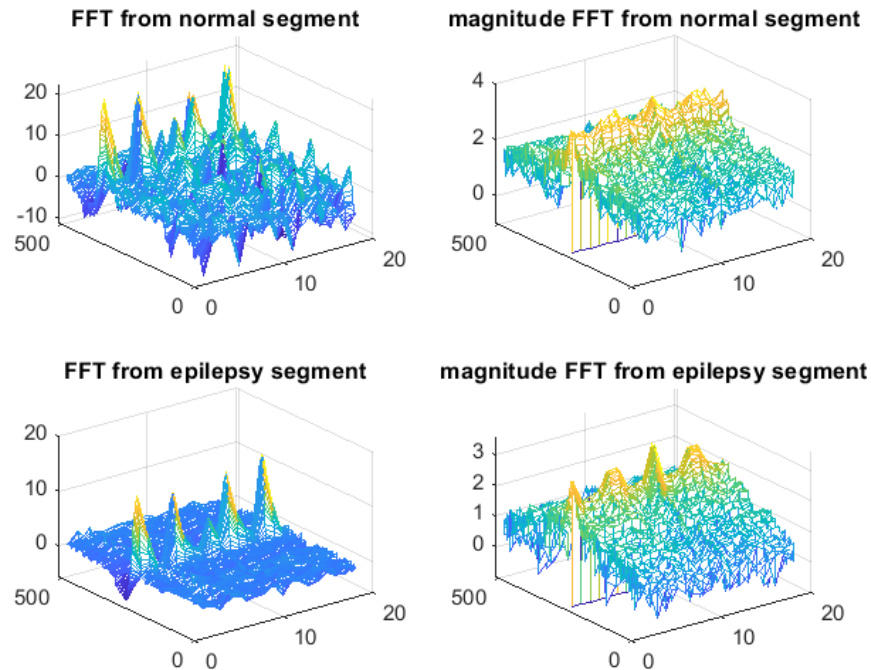


Figure 32. 2D FFT of a normal brain activity and an epileptiform brain activity (2 seconds length segment).

5.1.1. Model selection

A selection of representation of the input data has an effect on a model selection. Choosing magnitude spectrograms as one representation of the input data, an intuitive model selection was VGG model architecture. The VGG model is 2D convolutional neural network that is used for image classification. The input shape of VGG must be (*height, width, channels*). A *height* and *width* is the size of an image

and *channels* are color channels of an image. In the present study, one segment produced 18 spectrograms. A number of spectrograms were equal to a number of EEG signals. In the present study, VGG model handled spectrogram size similarly to image size, but a number of spectrograms were handled similarly to image color channels.

Using grayscale images in image classification a number of channels is one. Therefore, this approach enabled to use time-domain EEG signals as one representation of the input data of the VGG model. In this context, height was the number of EEG signals; a width was a length of a segment and channels was handled similarly to a number of color channels in a grayscale image. The same approach was used in experiments where FFT variants were a representation of the input data.

In the present study, the DenseNet 1D model was used only along with time-domain EEG signal data. An input shape of DenseNet 1D must be (width, channels). Thus, this fact led to conclusion that DenseNet 1D could be used along with time-domain EEG signal data and FFT variants. The decision to use DenseNet 1D only for time-domain EEG signals was based on the slowness of training. The feature maps produced by the DenseNet 1D expanded too large causing slowness in the training process. Hence, examination of DenseNet 1D remained just at a satisfactory level and caused to leave out experiments in which DenseNet 1D could have been used with all FFT variants.

5.1.2. *Hyperparameter tuning*

The detection results of DenseNet 1D model were compared with the results of VGG model in experiments in which time-domain EEG signals were used as a representation of the input data. The hyperparameters that were tuned for DenseNet 1D were a number of blocks, a bottleneck size, dilations, a growth rate and a number of epochs in each experiment. A number of blocks in this context refer to a one-dimensional array that consists of the number of DenseNet composition layers in each block of DenseNet 1D. Typical DenseNet block has four blocks and one block contains multiple basic DenseNet composition layers followed by a transition layer or a classification layer. [17] The bottleneck size gives filter size to the bottleneck layers in DenseNet block. True set dilations parameter adds space between each cell in filters. The growth rate means the number of feature maps of one layer. The learning rate is a hyperparameter that controls the rate of speed at which a model learns. The number of epochs is the number how many times algorithm goes through a training dataset.

VGG was used in the experiments where an input data type was spectrograms, 1D-FFT, 2D-FFT, 2D-FFT magnitude, 2D-FFT phase or raw EEG signal. Tuned hyperparameters of VGG in each experiment were a number of blocks, a number of convolutional layers in one block, a number of filters in one convolutional layer, a number of dense layers, a size of dense layers, a dropout rate and a number of epochs. The number of blocks in VGG refers to block which contains one or multiple convolutional layers, followed by max-pooling layer [15]. The structural logic of VGG remained the same even number of blocks varied. A batch size was tuned only in the experimental group 1 and a learning rate was not tuned in the experimental group 1. The hyperparameters and their values of DenseNet 1D and VGG are described in Table 6.

Table 4. Model, tuned hyperparameters and their values

Model	Hyperparameters	Values
DenseNet 1D	a batch size (only in the experimental group 1)	[32, 64 ,128]
	a number of blocks = [Block 1, Block 2, Block 3, Block 4]	Block 1 = [4, 6] Block 2 = [4, 6, 8, 10, 12] Block 3 = [4, 6, 8, 10, 12] Block 4 = [4, 6, 8, 10, 12]
	a bottleneck size	[2, 4, 6]
	dilations	[False, True]
	a growth rate	[4, 8, 16, 32]
	a learning rate (not in the experimental group 1)	[0.0001, 0.001, 0.01, 0.1]
	a number of epochs	[20, 40, 60, 80]
VGG	a number of blocks	[1, 2, 3, 4, 5, 6]
	a number of convolutional layers in one block	[1, 2, 3]
	a number of filters in convolutional layers	[4, 8, 16, 32]
	a number of dense layers	[1, 2, 3]
	a size of dense layers	[128, 256, 512, 1024]
	a dropout rate	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]
	a learning rate (not in the experimental group 1)	[0.0001, 0.001, 0.01, 0.1]
	a batch size (only in the experimental group 1)	[32, 64 ,128]
	a number of epochs	[20, 40, 60, 80]

5.2. Experiments

The experiments of the present study were grouped into three experimental groups based on executed pre-processing steps. One experiment of a specific experimental group followed the same pre-processing steps than other experiments within the experimental group. The objective of composing experimental groups was to organize experiments and to identify similarities between the experiments.

The hyperparameters of the DenseNet 1D and VGG were optimized by maximizing a validation F1-score in each experiment. The input parameters of the objective function for VGG are described in section *Hyperparameter tuning* in the present study. Because of data imbalance class weights were computed and fed to a classifier to penalize a loss function during training. The leave-one-out procedure was used to validate trained models if the model did not exceed a given threshold value for validation F1-score. If a model exceeded a given threshold (75%) then 5 folds cross validation was utilized.

5.2.1. Experimental group 1

In the experimental group 1, EEG signal was standardized by robustScalar of scikit-learn python library with the range between the 1st quartile (25th quantile) and the 3rd quartile (75th quantile). RobustScalar method removes the median and scales the data according to the quantile range using statistics that are robust to outliers. The data were centered before scaling. After data scaling, EEG signal was segmented into equal segments of 2 seconds.

Each experiment of the experimental group 1 had the same data split procedure. The data was split randomly into training, validation and testing datasets. The procedure is called as subject dependent split. In the subject dependent split samples from a single patient can be in training, validation and testing datasets. In the experimental group 1, 20% of the data was taken for testing, 20% for validation and 60% for training. The experiment 1 was the only one experiment within the experimental group 1 that exceeded 75% F1-score threshold. Thus, 5 folds cross validation was used only for this experiment and the leave-one-out cross validation was used for others.

In the experiment 1, EEG segments of 2 seconds were transformed into magnitude spectrograms. One EEG segment produced 18 magnitude spectrograms after transformation. The problem with spectrogram representations is that its values possess a large dynamic range. As a result, small values may be dominated by large values. Logarithmic compression was used to balance out this decadency by reducing the difference between large and small values and make the distribution more Gaussian. Logarithmic compression function is $f(x) = \log(x + c)$, where x is an input signal and c is a small constant. In the experiment 1, $1e-9$ was used as a small constant in logarithmic compression. In the experiment 1, the magnitude spectrograms and VGG model were investigated for epileptic seizure detection. The purpose of the experiment was to investigate is VGG capable to detect epileptic seizure patterns from magnitude spectrograms.

In the experiments 2 and 3, EEG signal segments were used as a representation of the input data. In the experiment 2, the DenseNet 1D and EEG signal were investigated and in the experiment 3, VGG with EEG signal as input representation were examined. The objective of the experiment 2 and 3 was to compare performance of 1D and 2D convolutional neural networks for EEG signal. The block diagram of the experimental group 1 is illustrated in Figure 33.

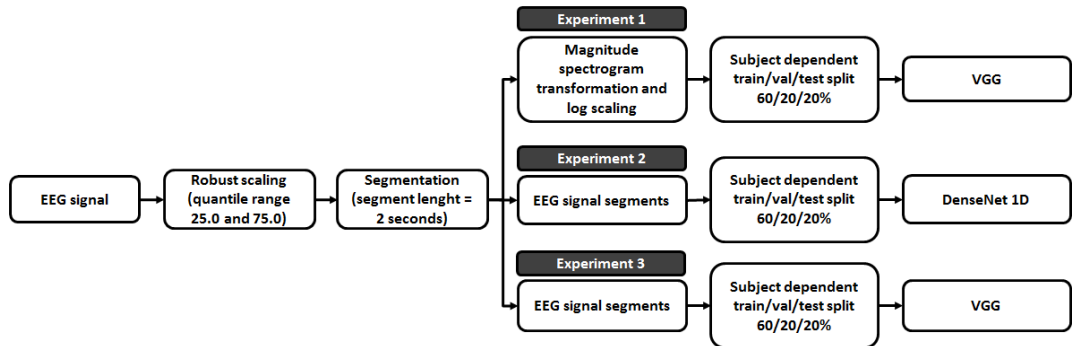


Figure 33. The block diagram of the experimental group 1.

5.2.2. Experimental group 2

In the experimental group 2, EEG signals were filtered by bandpass and low-pass filters to remove the powerline noise. The powerline noise was removed by excluding components in the frequency ranges of 57-63 Hz and 117-123 Hz for a power line frequency of 60 Hz. Moreover, DC component at 0 Hz was removed. The EEG data was normalized by min-max normalization to scale values between 0 and 1. In the figure 32, spectrograms visualize filtering and standardization steps of EEG signals. Spectrograms are used only for visualization in this context. After data scaling, EEG signal was segmented into equal segments of 2 seconds.

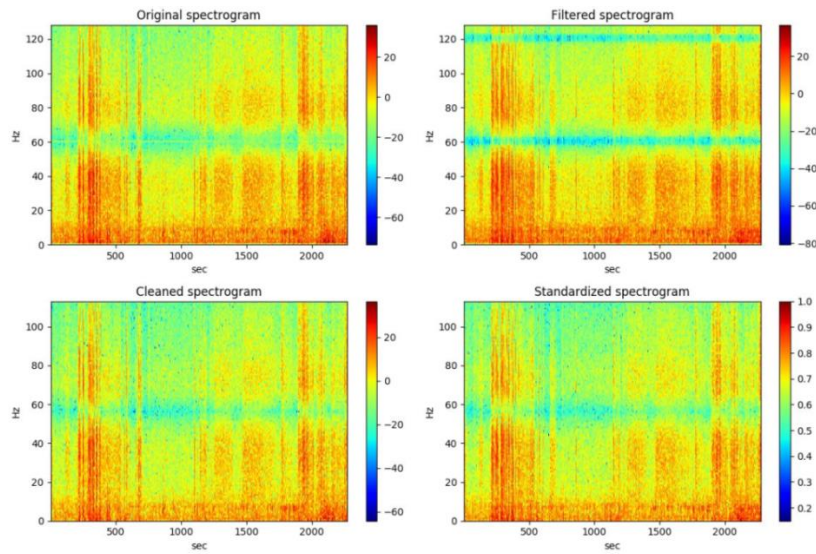


Figure 34. Pre-processing steps: original spectrogram, filtered with bandpass and high pass filter, cleaned and standardized.

Each experiment of the experimental group 2 had the same data split procedure. In the first split, subject dependent train/test split was used. 25% of the data was taken for testing and 75% of the data for training. At the second split, subject dependent train/val split was used. 20% of the data was used for validation and 80% for training. The reason for using subject dependent train/val split rather than subject dependent was a lack of data. The assumption was that a model is not able to generalize properly if there is not enough data. In the experimental group 2, the leave-one-out cross validation was used to validate trained models since validation F1-score of the models did not exceed the 70% threshold.

In the experiment 1, EEG segments of 2 seconds were fed into VGG and in the experiment 2, EEG segment with DenseNet 1D were investigated. The results of these two experiments were compared to each other. The objective of the comparison was to confirm findings regarding to using 2D convolutional model architecture rather than 1D convolutional model for multi-channel signal.

In the experiment 3, EEG segments of 2 seconds were transformed into 1D-FFT whereas in the experiment 4, EEG segments were transformed into 2D-FFT. In both experiments FFT variants were representations of the input data for VGG model. The purpose of the experiment 3 and 4 were to investigate could 2D-FFT give better

results than 1D-FFT regarding to spatial shift and rotation properties. The block diagram of the experimental group 2 is illustrated in Figure 35.

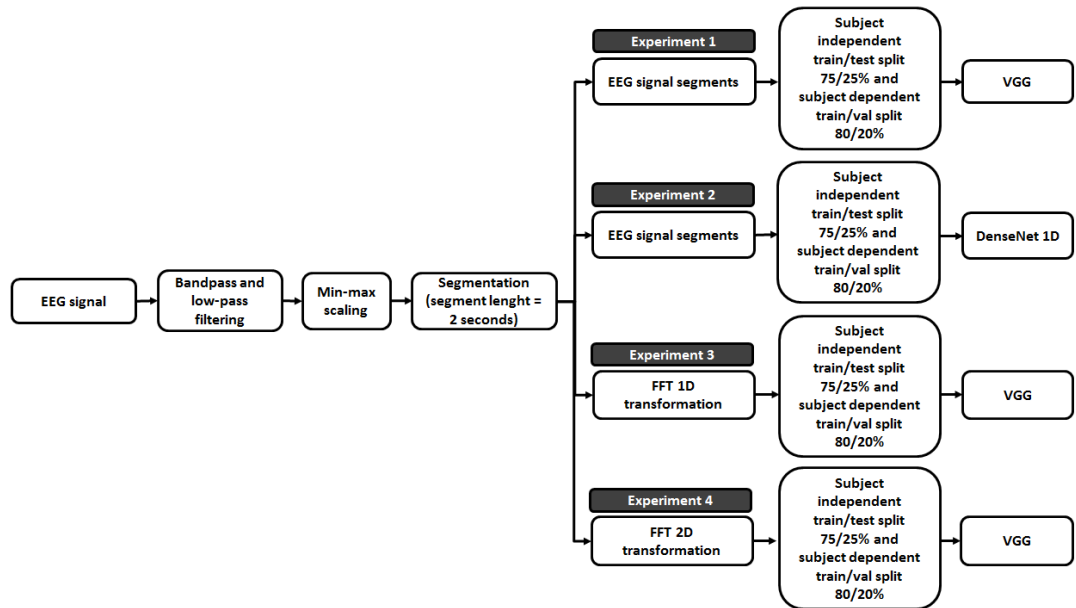


Figure 35. The block diagram of the experimental group 2.

5.2.3. Experimental group 3

In the experimental group 3, EEG signals were filtered by bandpass and low-pass filters to remove the powerline noise. The powerline noise was removed by excluding components in the frequency ranges of 57-63 Hz and 117-123 Hz for a power line frequency of 60 Hz. Moreover, DC component at 0 Hz was removed. The EEG data were standardized by z-score normalization to scale values, relative to the sample mean and standard deviation. After data scaling, EEG signal was segmented into equal segments of 2 seconds.

Each experiment of the experimental group 3 had the same data split procedure. In the first split, subject dependent train/test split was used. 25% of the data were taken for testing and 75% of the data for training. In the second split, subject dependent train/val split was used. 20% of the data was used for validation and 80% for training. Similar to experimental group 2, the reason for using subject dependent train/val split rather than subject dependent was a lack of data. In the experimental group 3, the leave-one-out cross validation was used to validate trained models since validation F1-score of the models did not exceed the 70% threshold.

In the experiment 1, EEG segments with VGG model were investigated. The objective of the experiment 1 was compare effect of z-score normalization with min-max normalization that was examined in the experimental group 2. The purpose of the experiment 2 was to investigate 2D-FFT with DenseNet 1D and compare the results with the results of the experiment 4 of the experimental group 2 where VGG model were used with the same input representation. In experiment 3 and 4, the objective was to investigate how dominant is the magnitude component as compared to phase component in the detection of epileptic seizure. The block diagram of the experimental group 3 is illustrated in Figure 36.

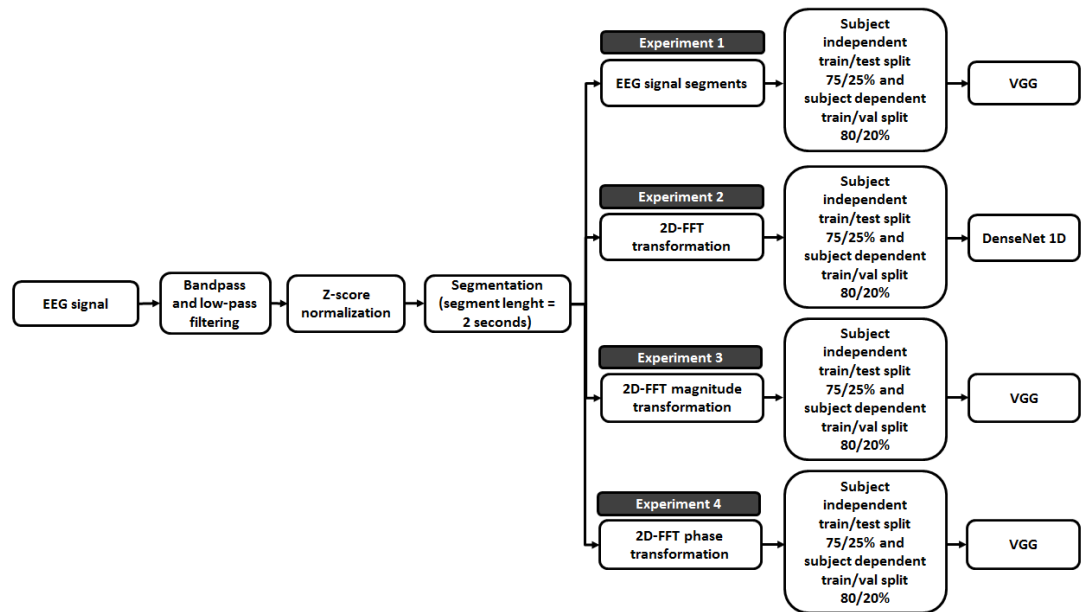


Figure 36. The block diagram of the experimental group 3.

6. RESULTS

6.1. Experimental group 1

The results of the experimental group 1 are represented in the Table 10 and Table 11. The Table 10 contains the results of the experiment 1. The best model in the experiment 1 was evaluated with subject dependent data. VGG model exceeded validation f1-score threshold (70%) therefore a model was tested with 5 subsets of the data. In the experiment 1, 1453 hyperparameter iterations were executed.

Table 5. The results of the experiment 1. Iterations = 1453, ACC = accuracy, F1 = F1-score, PREC = precision, SENS = sensitivity, SPEC = specificity, FPR = false positive rate and ER = error rate

Fold	ACC	F1	PREC	SENS	SPEC	FPR	ER
Fold 1	0.989	0.813	0.937	0.719	0.998	0.0017	0.01147
Fold 2	0.987	0.783	0.863	0.716	0.996	0.0041	0.01383
Fold 3	0.986	0.782	0.830	0.739	0.995	0.0054	0.01431
Fold 4	0.987	0.793	0.866	0.731	0.996	0.0041	0.01327
Fold 5	0.986	0.775	0.851	0.712	0.995	0.0045	0.01434

The Table 11 contains the results of the experiment 2 and 3. The best models were evaluated with subject dependent data and leave-one-out cross validation method. The experimental group 1 is illustrated in Figure 37.

Table 6. The results of the experiment 2 and 3 of the experimental group 1. EXP = experiment, ITER = number of hyperparameter iterations, ACC = accuracy, F1 = F1-score, PREC = precision, SENS = sensitivity, SPEC = specificity, FPR = false positive rate and ER = error rate

EXP #	ITER	ACC	F1	PREC	SENS	SPEC	FPR	ER
Exp 2	34	0.980	0.675	0.790	0.589	0.995	0.0055	0.01931
Exp 3	32	0.984	0.754	0.841	0.684	0.995	0.0047	0.01566

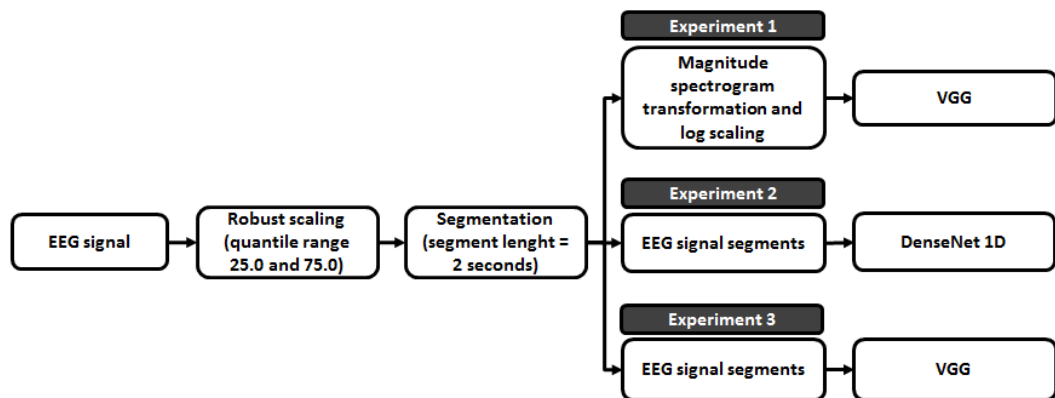


Figure 37. The experimental group 1 (subject dependent).

6.2. Experimental group 2

The results of the experimental group 2 are represented in the Table 12. In the final evaluation, the best models were evaluated with subject dependent data and leave-one-out cross validation method. The experimental group 2 is illustrated in Figure 38.

Table 7. The results of the experimental group 2. EXP = experiment, ITER = number of hyperparameter iterations, ACC = accuracy, F1 = F1-score, PREC = precision, SENS = sensitivity, SPEC = specificity, FPR = false positive rate and ER = error rate

EXP #	ITER	ACC	F1	PREC	SENS	SPEC	FPR	ER
Exp 1	13	0.982	0.362	0.492	0.287	0.995	0.0046	0.01545
Exp 2	57	0.985	0.358	0.490	0.282	0.996	0.0046	0.01550
Exp 3	177	0.986	0.388	0.634	0.280	0.997	0.0025	0.01350
Exp 4	70	0.984	0.368	0.470	0.303	0.995	0.0053	0.01589

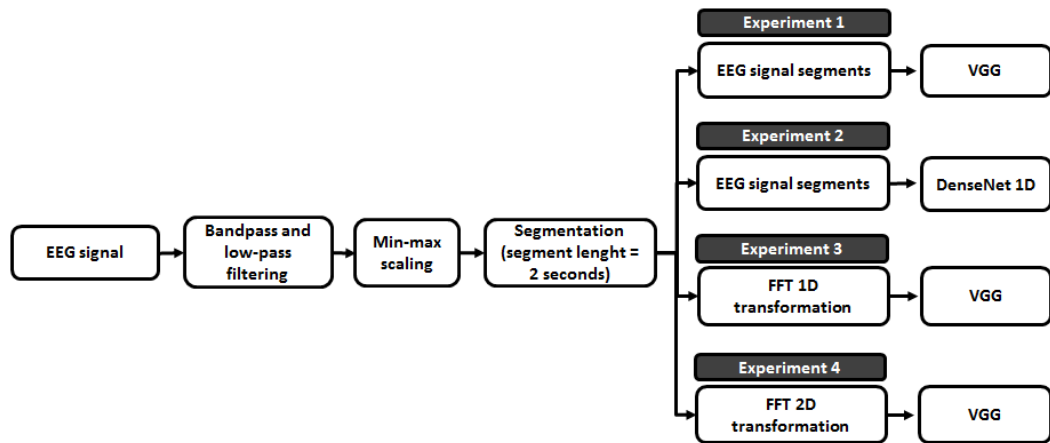


Figure 38. The experimental group 2 (subject independent).

6.3. Experimental group 3

The results of the experimental group 3 are represented in the Table 13. In the final evaluation, the best models were evaluated with subject dependent data and leave-one-out cross validation method. The experimental group 3 is illustrated in Figure 39.

Table 8. The results of the experiment 4. EXP = experiment, ITER = number of hyperparameter iterations, ACC = accuracy, F1 = F1-score, PREC = precision, SENS = sensitivity, SPEC = specificity, FPR = false positive rate and ER = error rate

EXP #	ITER	ACC	F1	PREC	SENS	SPEC	FPR	ER
Exp 1	43	0.987	0.485	0.606	0.405	0.996	0.0041	0.01315
Exp 2	33	0.988	0.527	0.675	0.432	0.997	0.0032	0.01189
Exp 3	56	0.988	0.523	0.697	0.418	0.997	0.0028	0.01692
Exp 4	38	0.980	0.105	0.179	0.075	0.995	0.0053	0.01942

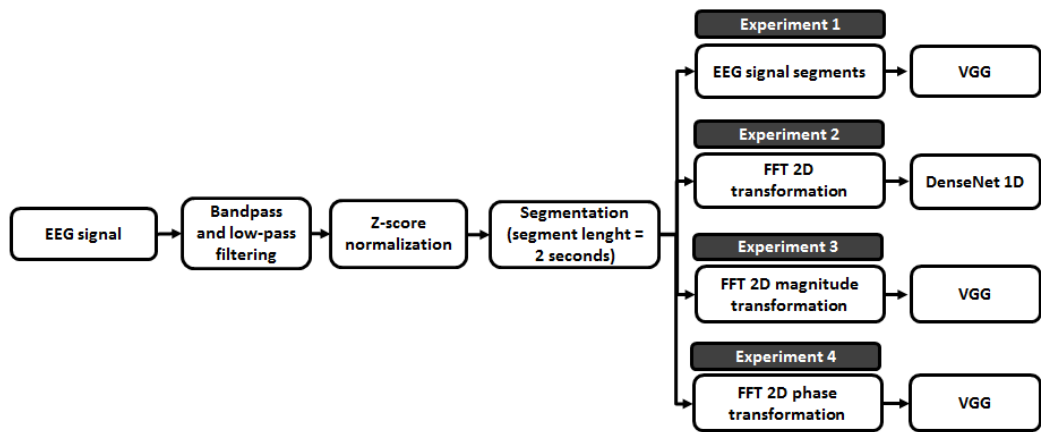


Figure 39. The experimental group 3 (subject independent).

7. DISCUSSION

In this study, the subject dependent experiments gave reasonable good results with magnitude spectrograms as a representation of input data for VGG model. The magnitude spectrograms could be used for personalized epileptic seizure detection. For developing subject dependent model, there was not enough data for used deep learning approaches in this study. The segment length of the samples was too small to classify individually the samples to normal and epileptic brain activity samples with subject dependent data.

One solution to address the noticed issue is to combine VGG model with recurrent model architecture or use totally different model architecture. VGG-LSTM and CNN 3D are these type of neural networks that were under investigation during this thesis, but unfortunately the time run out to make them work properly. Another possible solution is to implement post-processing algorithm to fill out error caps between correctly predicted segments. However, this approach has many problems and is more likely quick fix for a model that doesn't work as expected. The third possible solution could be to use multiple different representations of input data together as an input data for a neural network. All of these proposed solutions should be investigated in future studies.

Zhou et al. [198] used the CNN model to detect epileptic seizures based on EEG signals. They used time and frequency domain signals separately in different experiments. They concluded that frequency domain better achieved better results than time domain signals. They achieved with frequency domain signal satisfactory results with Freiburg and CHB-MIT datasets on three-class problem. The results of the present study agree with the results of Zhou et al. The results of all experimental groups show that frequency-domain signals give better results than EEG signals as a representation of input data for a neural network.

Weixia Liang et al. [170] and Ramy Hussein et al. [153] have proved that CNN-LSTM and LSTM neural network architectures are good model candidates for epileptic seizure detection. In this context need to point out that Liang et al. used image based EEG samples and Hussein et al. used intracranial EEG data. Images are the most studied representation of input data for neural networks, therefor image based EEG samples are easier to apply for existing neural network architectures. Image based EEG signals are interesting topic and should be investigated in future studies.

In the study of Ramy Hussein et al., intracranial EEG offers more accurate EEG data than scalp EEG. However, there are only a limited amount of use cases where intracranial EEG data are available. One rare use case is, when epilepsy is severe and epileptogenic needs localization for a surgery. In this thesis, the requirement was that the data is scalp EEG data so there could be more existing use cases for a developed solution.

Kostas et al. used LSTM with time domain, frequency domain, correlation and graph theory features from 5-s long EEG segments [199]. This study indicates that there still might be need for a manual feature extraction in some levels with EEG data. Kostas et al. used subject dependent data for evaluation. Therefore, the results of the present study in the experiment 1 and 2 are somewhat comparable with results of Kosta et al.'s study. However, not all of the normal data was used for evaluation in the present study due to a lack of time.

Njedly et al. used multimodal feature fusion for seizure forecasting in canines [200]. This study showed that multimodal feature fusion with EEG signal and FFT data representations of input data together might be a good candidate for epileptic seizure detection. The study was conducted with EEG data from canines, therefore there is no guarantee it is suitable for human EEG data. A similar approach might be worth to try with different combinations of representation of input data that were explored in this present study. Moreover, CNN-LSTM, LSTM and CNN 3D have stood out to be good model candidates along with multimodal feature fusion-based models for further studies.

The study of Njedly et al. inspired to investigate different representation of input data along with the EEG-signal. As mentioned earlier magnitude spectrograms appeared to be the most powerful representation of input data as compared to EEG signal in the subject dependent experiments. The decision to use spectrograms as a representation of input data led into interesting idea of handling spectrograms as a video. This way there is a possibility to use the same model architectures than in a video classification. This idea was good, but it was not conducted because a lack of time. In the future studies, this approach should be investigated.

The comparison of 1D-FFT and 2D-FFT as a representation of input data was interesting. However, the results show that using 2D-FFT as a representation of input data does not improve the results. The finding indicates that shift and rotation properties are not valuable information for epileptic seizure detection with these deep learning approaches used in this study. However, this approach could be more relevant in the distinction between epilepsy types from EEG signal, or in localization of epileptogenic.

The magnitude of 2D-FFT was compared to the 2D-FFT phase spectrum to confirm the importance of the magnitude spectrum in the detection of epileptic seizure. The difference of the results was significant, but at the same time expected. During epileptic seizure, the magnitude of FFT often increases, but phase spectrum varies. The phase of FFT is not useful without the magnitude spectrum in detection of epileptic seizure. Instead, the magnitude of FFT can be used without the phase spectrum, but still the magnitude of FFT cannot be the only representation of input data for epileptic seizure detection.

In 2018, Shahbazi et al. proposed CNN-LSTM neural network with STFT input to classify preictal and interictal states [18]. Shahbazi et al. used also CHB-MIT dataset. The study claimed there is still a gap in the literature for a reliable and generalizable method that can be used in practice to predict epileptic seizure. They believed that there are two main reasons: 1) A lack of reliable EEG data. Preictal stage is accepted fact, which unlike its onset is not recognizable by eyes from EEG signal. Therefore, there is no global consensus about how to define and identify it, making it hard to develop a practical prediction approach. 2) The complexity and variability of the patterns of the preictal state. This leads to a conclusion that the preictal state is different among different patients and even between different seizures of the same patient. A lack of reliable data appeared to be a big challenge also in the present study. Another prominent challenge was the complexity and variability of the patterns in EEG signal between patients and even between epileptic seizures of a single patient. [18]

It is important to consider other approaches for epileptic seizure detection beside EEG signal-based solutions. Rambopal et al. has listed [201] seizure detection systems invented before the year 2014. In the review of Rambopal et al., there are 17

EEG based seizure detection systems, but they all are stated to be specialized in specific epilepsy types such as an absence seizure, or detection algorithms that uses handmade features. Rambopal et al. listed five detection systems that are based on intracranial EEG signals. As mentioned earlier, intracranial EEG is not suitable for a common usage in epileptic seizure treatment since it is only used in research of severe epilepsy. In the review, there are also over 20 solutions that are based on ECG, motion or video data and only a few of these might be suitable for all epilepsy types because they are limited to one specific epilepsy type.

The review of Rambopal et al. indicates how difficult is to build epileptic seizure detection system because of complexity of epilepsy disorders. Epilepsy types have different characteristics and their appearance differ from each other in EEG signal. Another challenge in this context is that in 2017 ILAE announced a new classification for epileptic seizure types. A publication of the announcement underlines the fact that the classification of epilepsy has not been fully established.

It is obvious that the detection of epileptic seizure is a difficult task to do even by current deep learning approaches. The major challenge in the epileptic seizure detection is a stochastic nature of EEG signals. It is hard to determine is there enough data available for a neural network to identify normal and epileptic seizure features. The amount of data in the CHB-MIT dataset is limited, and it does not contain information about epileptic seizure types. Regarding to literature review, epilepsy is not a singular disease entity, but a variety of disorders caused by different brain dysfunctions. The characteristics of epileptic seizure patterns in EEG signal may vary even for a single person. Therefore, it is vital to have enough EEG data for training and testing a model, to ensure that a model has a higher possibility to work in a real life.

Another challenge is an appearance of artefacts in EEG signal. In a real life, there is variety of artefact that a model needs to be able to handle so that they do not have an effect on epileptic seizure detection. It has been claimed that a deep learning approaches are robust to noise. However, it needs to be pointed out that if there is so much noise and artefacts that a neurologist is not able to distinguish epileptic seizure and the normal EEG signal from each other, then a model cannot do it either. The EEG signal contains always noise and variety of artefact. This is a challenge that needs to be solved at first, before building an automatic epileptic seizure detection system based on EEG signals.

Currently, one major challenge is limited amount of available deep learning architectures. Most of the deep learning architectures are developed for image classification, but not for signals. The research in a field of applying deep learning models for signals is novel as compared to the research of a deep learning with images. Detection and classification from time series data often require features from previous or next segments to improve a detection score. This finding led into a conclusion that without recurrent model or overlapping sliding window segmentation it might be impossible to detect epileptic seizure based on the features from single segments. As mentioned earlier, multimodal feature fusion-based model with a variety of representations of input data might be one possible solution. Transfer learning was also considered to overcome this challenge. The idea was to use pre-trained VGG model with LSTM model. Investigation of a stacked VGG-LSTM model was excluded from the thesis because a lack of time.

Deep learning is claimed to be handy because expert knowledge is not needed for a specific domain but on the other hand, that is a serious problem. A data scientist

needs to be able evaluate the correctness of a single prediction of a model from a visual representation of input data. This was a huge challenge during the thesis because the interpretation of EEG signals is difficult. Difficulty of interpretation of EEG signals highlighted the importance of feature engineering also in deep learning.

8. CONCLUSION

In the present thesis, deep learning approach for epileptic seizure detection from EEG signal was addressed. The study included three experimental groups based on data processing pipeline steps. Subject dependent experiment gave reasonable good results with magnitude spectrograms as representation of input data of VGG model. The same approach was examined with subject independent data but the results were poor and were therefore left out from the results section. Thus, it can be concluded that VGG was not able to generalize features from segments of 2 seconds when it was trained with subject dependent data.

Regarding to the results, the magnitude spectrograms proved to be better representation of input data than EEG signal for VGG model. However it need to be noted that there were 1453 hyperparameter iterations for VGG model with magnitude spectrograms versus 34 hyperparameter iterations for VGG and DenseNet models with EEG signal data. Therefore, in the experiment where magnitude spectrograms where used had much more time to optimize validation F1-score.

The performance of VGG and DenseNet 1D with EEG signal as representation of the input data were compared in all experimental groups. The results show that 2D convolutional neural network also can be used for multi-channel signals in the same way as 1D convolutional neural network in this study. The objective of this study case was to investigate spatial information essential for epileptic seizure detection. The results with this small dataset show that spatial information is not essential. However, the size of the dataset advocates that this conclusion cannot be fully made.

The essentiality of the spatial information of epileptic brain activity in EEG signal was also investigated with 2D-FFT transformation. The results of 2D-FFT are in the same level than the results of 1D-FFT. This study case supports the previous conclusion that spatial information of epileptic brain activity is not essential for epileptic seizure detection. However, it needs to be noted again that the dataset was too small for this conclusion.

The most interesting research question of the thesis was that is magnitude spectrum more essential than phase spectrum of FFT for epileptic seizure detection. This research question was addressed with magnitude and phase spectrums extracted from 2D-FFT. The results show that magnitude spectrum is more essential than phase spectrum for epileptic seizure detection from EEG signal. VGG model achieved better results with 2D-FFT magnitude spectrum than with 2D-FFT phase spectrum, or with original 2D-FFT.

In the pre-processing phase, different normalization techniques were examined for data normalization. Regarding to the results of the experiments, the z-score normalization can be suggested as the most promising normalization approach for the dataset in this study. The z-score normalization enabled better results than min-max normalization. The reason is that z-score normalization handles outlier better than min-max normalization. Moreover, it was convenient to remove the power line noise by bandpass and low pass filters from the data.

In final words, the scope of this study turned out to be enormous along the way, due to a variety of concepts that were difficult to embrace in a short time such as interpretation of EEG signals, feature engineering and deep learning methods. The study did not produce a model design for epileptic seizure detection, but it covers a

variety of pain points that need to be taken into account while developing an automated EEG-based epileptic seizure detection system. The study also brought up various deep learning approaches that might be the key factors for designing a real life epileptic seizure detection system in a near future.

9. REFERENCES

- [1] Kälviäinen, R. (2016) 1. epilepsia on muutakin kuin kohtaus. In Kälviäinen, M. Järvisoutu-Hulkkonen, T. Keränen & H. Rantala (Eds.), *Epilepsia* (pp. 8-10). Printon, Tallinna: Kustannus Oy Duodecim.
- [2] Reynolds, E. H. (2000) The ILAE/IBE/WHO global campaign against epilepsy: Bringing epilepsy "out of the shadows". *Epilepsy & Behavior* : E&B, 1(4), S3-S8. DOI:S1525-5050(00)90104-5.
- [3] Roivainen, R. & Kallela, M. (2015) *Neurologia, Epilepsia*. In Kauma, I. & Kotila, L (eds.) *Therapia fennica* (10. laitos ed.). Helsinki: Kandidaattikustannus.
- [4] Pirkanmaan sairaanhoitopiiri. (2016) Aikuisen epilepsia hoitoketjut. URL: https://www.terveysportti.fi/dtk/ltk/koti?p_haku=epilepsia. Accessed 28.2.2019.
- [5] Hussein, R., Palangi, H., Ward, R., & Wang, Z. J. (2018a) Epileptic seizure detection: A deep learning approach.
- [6] Shoeb, A. (2010) Application of machine learning to epileptic seizure detection.
- [7] Mirowski, P. (2009) Classification of patterns of EEG synchronization for seizure prediction. *Clinical Neurophysiology*, 120(11), 1927-1940. DOI:10.1016/j.clinph.2009.09.002
- [8] Meier, R. (2007) Detecting epileptic seizures in long-term human EEG: A new approach to automatic online and real-time detection and classification of polymorphic seizure patterns. *Epilepsia*, 48, 33.
- [9] Gardner, A. B. (2006) One-class novelty detection for seizure analysis from intracranial EEG. *Journal of Machine Learning Research*, 7, 1025-1044.
- [10] Fisch, B. J., & Fisch, B. J. (2010) Patient evaluation and selection for routine and invasive epilepsy monitoring. *Epilepsy and intensive care monitoring: Principles and practice*. New York: Demos Medical.
- [11] Fisher, R. S., van Emde Boas, W., Blume, W., Elger, C., Genton, P., Lee, P., & Engel, J., Jr. (2005) Epileptic seizures and epilepsy: Definitions proposed by the international league against epilepsy (ILAE) and the international bureau for epilepsy (IBE). *Epilepsia*, 46(4), 470-472. DOI:EPI66104
- [12] Fisher, R. S., Cross, J. H., French, J. A., Higurashi, N., Hirsch, E., Jansen, F. E., Lagae L., Moshé, S. L., Peltola J., Perez, E. R., Scheffer, I. E., Zuberi, S. M. (2017) Operational classification of seizure types by the

international league against epilepsy: Position paper of the ILAE commission for classification and terminology. *Epilepsia*, 58(4), 522-530.

- [13] LeCun, Y., & Bengio, Y. (1995) Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, 3361(10), 1995.
- [14] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012) Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097-1105.
- [15] Simonyan, K., & Zisserman, A. (2014) Very deep convolutional networks for large-scale image recognition.
- [16] He, K., Zhang, X., Ren, S., & Sun, J. (2016) Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778.
- [17] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017) Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700-4708.
- [18] Shahbazi, M. (2018) A generalizable model for seizure prediction based on deep learning using cnn-lstm architecture. DOI:10.1109/GlobalSIP.2018.8646505
- [19] Shuiwang Ji. (2013) 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), 221-231. DOI:10.1109/TPAMI.2012.59
- [20] Tran, D. (2015) Learning spatiotemporal features with 3D convolutional networks. DOI:10.1109/ICCV.2015.510.
- [21] Stafstrom, C. E., & Carmant, L. (2015) Seizures and epilepsy: An overview for neuroscientists. *Cold Spring Harbor Perspectives in Medicine*, 5(6), a022426.
- [22] Gastaut, H. (1973) Dictionary of epilepsy part 1: Definitions. Geneva: World Health Organization. URL: <https://apps.who.int/iris/bitstream/handle/10665/62903/Dictionary-of-Epilepsy-Part1-1973-eng.pdf?sequence=1&isAllowed=y>. Accessed 3.3.2019.
- [23] Kälviäinen, R., & Erikson, K. (2016a) 2. Mitä epilepsia on? In R. Kälviäinen, M. Järvisoutu-Hulkkonen, T. Keränen & H. Rantala (Eds.), *Epilepsia* (pp. 12-19). Printon, Tallinna: Kustannus Oy Duodecim.
- [24] Scheffer, I. E., Berkovic, S., Capovilla, G., Connolly, M. B., French, J., Guilhoto, L., Hirsch, E., Jain, S., Mathern, G. W., Moshé, S. L., Nordli, D. R., Perucca, E., Tomson, T., Wiebe, S., Zhang, Y., Zuberi, S. M. (2017)

ILAE classification of the epilepsies: Position paper of the ILAE commission for classification and terminology. *Epilepsia*, 58(4), 512-521.

- [25] Scheffer, I. E., Berkovic, S., Capovilla, G., Connolly, M. B., French, J., Guilhoto, L., Hirsch, E., Jain, S., Mathern, G. W., Moshé, S. L., Nordli, D. R., Perucca, E., Tomson, T., Wiebe, S., Zhang, Y. & Zuberi, S. M. (2017) ILAE classification of the epilepsies: Position paper of the ILAE commission for classification and terminology. *Epilepsia*, 58(4), 512-521. DOI:10.1111/epi.13709.
- [26] Centeno, M., & Carmichael, D. W. (2014) Network connectivity in epilepsy: Resting state fMRI and EEG-fMRI contributions. *Frontiers in Neurology*, 5, 93. DOI:10.3389/fneur.2014.00093.
- [27] Bancaud, J., Henriksen, O., Rubio-Donnadieu, F., Seino, M., & Dreifuss, F. E. (1981) Proposal for revised clinical and electroencephalographic classification of epileptic seizures. from the commission on classification and terminology of the international league against epilepsy. *Epilepsia*, 22(4), 489-501.
- [28] Kälviäinen, R., & Erikson, K. (2016b) Epilepsian diagnosointi. *Epilepsia* (pp. 22-30). Tallinna: Kustannus Oy Duodecim.
- [29] Fisher, R. S., Cross, J. H., D'souza, C., French, J. A., Haut, S. R., Higurashi, N., Hirsch, E., Jansen, F. E., Lagae, L., Moshé, S. L., Peltola, J., Roulet, P. E., Scheffer, I. E., Schulze- Bonhage, A., Somerville, E., Sperling, M., Yacubian, E. M. & Zuberi, S. M. (2017) Instruction manual for the ILAE 2017 operational classification of seizure types. *Epilepsia*, 58(4), 531-542.
- [30] Kiriakopoulos, E., & Shafer, P. O. (2017a) Types of seizures. URL: <https://www.epilepsy.com/learn/types-seizures>. Accessed 13.2.2019.
- [31] Van Wagenen, W. P., & Herren, R. Y. (1940) Surgical division of commissural pathways in the corpus callosum: Relation to spread of an epileptic attack. *Archives of Neurology & Psychiatry*, 44(4), 740-759.
- [32] Blume, H. (1997) Corpus callosotomy. URL: <https://www.epilepsy.com/learn/professionals/diagnosis-treatment/surgery/corpus-callosotomy>. Accessed 12.2.2019.
- [33] Britton, J. W., Frey, L. C., Hopp, J. L., Korb, P., Koubeissi, M. Z., Lievens, W. E., Pestana-Knight E. M., St. Louis, E. K. (2016) *Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants*.
- [34] Suomalaisen Lääkäriseuran Duodecimin ja Suomen Neurologinen Yhdistys ry:n asettama työryhmä. (2014) Epilepsiat (aikuiset). URL: <http://www.kaypahoito.fi/web/kh/suosituksset/suositus?id=hoi50072>. Accessed 18.2.2019.

- [35] Trinka, E., Cock, H., Hesdorffer, D., Rossetti, A. O., Scheffer, I. E., Shinnar, S., Shorvon, S., Lowenstein, D. H. (2015) A definition and classification of status epilepticus—Report of the ILAE task force on classification of status epilepticus. *Epilepsia*, 56(10), 1515-1523.
- [36] Suomalaisen Lääkäriseuran Duodecimin, Suomen Lastenneurologinen Yhdistys ry:n ja Suomen Neurologinen Yhdistys ry:n asettama työryhmä. (2016) Käypä hoito -suositus: Epileptinen kohtaus (pitkittynyt; status epilepticus). URL: <http://www.kaypahoito.fi/web/kh/suosituksset/suositus?id=hoi50030>. Accessed 5.2.2019.
- [37] Karppinen, A., Ansakorpi, H., & Liisanantti, J. (2016) Status epilepticus—tunnista nopeasti, hoida tehokkaasti. Finnanest. URL: http://www.finnanest.fi/files/karppinen_ansakorpi_liisanantti_status_epilepticus.pdf. Accessed 26.3.2019.
- [38] Tatum, W. (2014a) The EEG in status epilepticus. *Handbook of EEG interpretation*, second edition (pp. 155-188) Demos Medical.
- [39] Sculier, C., Gaínza-Lein, M., Sánchez Fernández, I., & Loddenkemper, T. (2018) Long-term outcomes of status epilepticus: A critical assessment. *Epilepsia*, 59, 155-169.
- [40] Suomalaisen Lääkäriseuran Duodecimin, Suomen Lastenneurologinen Yhdistys ry:n ja Suomen Neurologinen Yhdistys ry:n asettama työryhmä. (2015) Epileptinen kohtaus (pitkittynyt; status epilepticus). URL: <http://www.kaypahoito.fi/web/kh/suosituksset/suositus?id=hoi50030>. Accessed 5.2.2019.
- [41] Mervaala, E., & Jäntti, V. (2006) Status epilepticus ja hoidon seuranta. In J. Partanen, B. Falck, J. Hasan, V. Jäntti, T. Salmi & U. Tolonen (Eds.), *Kliininen neurofysiologi* (pp. 199-205). Helsinki: Duodecim.
- [42] Tatum, W. O. (2008) Normal EEG. *Handbook of EEG interpretation* (pp. 1-50). New York: Demos Medical.
- [43] Vanhatalo, S., Heinonen, H., Kallio, M., Mervaala, E., & Lauronen, L. (2018) EEG:N perusta. In E. Mervaala (Ed.), *Kliininen neurofysiologia*. Helsinki: Duodecim. DOI:knf00801
- [44] Britton, J.W., Frey, L.C., Hopp, J.L., Korb, P., Koubeissi, M.Z., Lievens, W.E., Pestana-Knight, E.M. & St, E.L. (2016a) Introduction. *Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants* (pp. 4-7) American Epilepsy Society, Chicago.
- [45] Huttunen, J., Tolonen, U., & Partanen, J. (2006) EEG:N fysiologiaa ja patofysiologiaa. In J. Partanen, B. Falck, J. Hasan, V. Jäntti, T. Salmi & U. Tolonen (Eds.), *Kliininen neurofysiologia* (pp. 50-64). Helsinki: Duodecim.

- [46] Vanhatalo, S., Lauronen, L., Heinonen, H., Kallio, M., & Mervaala, E. (2018) EEG:N rekisteröinti. In E. Mervaala (Ed.), *Klininen neurofysiologia*. Duodecim.
- [47] Britton, J. W., Frey, L. C., Hopp, J. L., Korb, P., Koubeissi, M. Z., Lievens, W. E., Pestana-Knight E. M., St, E. L. (2016b) *Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants* American Epilepsy Society, Chicago.
- [48] Koivu, M., Eskola, H., & Tolonen, U. (2006) EEG:N rekisteröinti, aktivaatiot ja lausunto. In J. Partanen, B. Falck, J. Hasan, V. Jäntti, T. Salmi & U. Tolonen (Eds.), *Klininen neurofysiologia* (pp. 65-83). Helsinki: Duodecim.
- [49] Tolonen, U., & Partanen, J. (2006) 12 EEG-tutkimuksen klininen käyttö: Aiheet ja EEG-häiriön löydöstyypit. In J. Partanen, B. Falck, J. Hasan, V. Jäntti, T. Salmi & U. Tolonen (Eds.), *Klininen neurofysiologia* (pp. 144-154). Helsinki: Duodecim.
- [50] Tatum, W. O., IV, & Tatum, W. O. (2017) *Artifact and ambulatory EEG. Ambulatory EEG monitoring* (pp. 41-74). New York: Demos Medical.
- [51] Fisher, R. S., & Scharfman, H. E. (2014) How can we identify ictal and interictal abnormal activity? *Issues in clinical epileptology: A view from the bench* (pp. 3-23) Springer.
- [52] Beniczky, S., Aurlen, H., Brøgger, J. C., Hirsch, L. J., Schomer, D. L., Trinka, E., Pressler, R. M., Wennberg, R., Visser, G. H., Eisermann, M., Diehl, B., Lesser, R. P., Kaplan, P. W., Nguyen The Tich, S., Lee, J. W., Martins-Da-Silva, A., Stefan, H., Neufeld, M. (2017) Standardized computer-based organized reporting of EEG: SCORE–Second version. *Clinical Neurophysiology*, 128(11), 2334-2346.
- [53] Stern, J. M., & Engel, J. (2013a) Identifying EEG patterns by their features. *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [54] Stern, J. M., & Engel, J. (2013b) Introduction to EEG interpretation. *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [55] Rangayyan, R. M. (2015a) The electroencephalogram (EEG). In T. Samad (Ed.), *Biomedical signal analysis* (pp. 34-40). New Jersey: John Wiley & Sons. DOI:10.1002/9781119068129
- [56] Kane, N., Acharya, J., Beniczky, S., Caboclo, L., Finnigan, S., Kaplan, P. W., Hiroshi S., Ronit P., van Putten, M. J. (2017) A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. revision 2017. *Clinical Neurophysiology Practice*, 2, 170.

- [57] Tolonen, U., & Lehtinen, I. (2006) Aikuisen normaali EEG. In J. Partanen, B. Falck, J. Hasan, V. Jäntti, T. Salmi & U. Tolonen (Eds.), *Klininen neurofysiologia* (pp. 109-128). Helsinki: Duodecim.
- [58] Stern, J. M., & Engel, J. (2013c) Alpha activity. In J. Goolsby (Ed.), *Atlas of EEG patterns* (2nd ed.,). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [59] Stern, J. M., & Engel, J. (2013d) Theta activity. In J. Goolsby (Ed.), *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [60] Stern, J. M., & Engel, J. (2013e) Delta activity. In J. Goolsby (Ed.), *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [61] Stern, J. M., & Engel, J. (2013f) Beta activity. In J. Goolsby (Ed.), *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [62] Lauronen, L., Vanhatalo, S., Heinonen, H., Kallio, M., & Mervaala, E. (2019) *Poikkeava EEG. Klininen neurofysiologia* (1st ed). Helsinki: Kustannus Oy Duodecim.
- [63] Tatum, W. (2014b) Epileptiform discharges. *Handbook of EEG interpretation*, second edition (pp. 81-95) Demos Medical.
- [64] Vakkuri, A., Yli- Hankala, A., Särkelä, M., Lindgren, L., Mennander, S., Korttila, K., Saarnivaara, L. & Jäntti, V. (2001). Sevoflurane mask induction of anaesthesia is associated with epileptiform EEG in children. *Acta Anaesthesiologica Scandinavica*, 45(7), 805-811.
- [65] Kreuzer, I. (2014) In Alexander Osthaus W. (Ed.), Typical examples for the EEG patterns delta (B. Schultz Trans.). DOI:10.1371/journal.pone.0089191.g001
- [66] Litt, B. (2002) Prediction of epileptic seizures. *Lancet Neurology*, 1(1), 22-30.
- [67] Stern, J. M., & Engel, J. (2013g) Ictal epileptiform patterns. *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [68] Stern, J. M., & Engel, J. (2013h) Ictal epileptiform patterns. *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [69] Stern, J. M., & Engel, J. (2013i) Interictal epileptiform patterns. In J. Goolsby (Ed.), *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.

- [70] Stern, J. M., & Engel, J. (2013j) Paroxysmal fast activity. In J. Goolsby (Ed.), *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [71] Stern, J. M., & Engel, J. (2013k) Periodic epileptiform discharges. In J. Goolsby (Ed.), *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [72] Fisher, R. S., & Schachter, S. C. (2000) The postictal state: A neglected entity in the management of epilepsy. *Epilepsy & Behavior*, 1(1), 52-59.
- [73] Abood, W., & Bandyopadhyay, S. (2019) Seizure, postictal state. *StatPearls*. Treasure Island (FL): StatPearls Publishing LLC. DOI:NBK526004
- [74] Kaibara, M. (1988) The postictal electroencephalogram. *Electroencephalography and Clinical Neurophysiology*, 70(2), 99-104. DOI:10.1016/0013-4694(88)90109-5
- [75] Stern, J. M., & Engel, J. (2013l) Benign epileptiform transients of sleep. In J. Goolsby (Ed.), *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [76] Stern, J. M., & Engel, J. (2013m) Needle spikes. In J. Goolsby (Ed.), *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [77] Stern, J. M., & Engel, J. (2013n) Burst-suppression pattern. In J. Goolsby (Ed.), *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [78] Stern, J. M., & Engel, J. (2013o) K complexes. In J. Goolsby (Ed.), *Atlas of EEG patterns* (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [79] Gélisse, P. (2016) Artifacts. *Atlas of electroencephalography : Activation procedures and artifacts* (pp. 343) John Libbey Eurotext.
- [80] Hakalax, N., Sainio, K., & Tolonen, U. (2006) EEG:N artefaktit ja valvonta. In J. Partanen, B. Falck, J. Hasan, V. Jäntti, T. Salmi & U. Tolonen (Eds.), *Kliininen neurofysiologia* (pp. 98-108). Helsinki: Duodecim.
- [81] Cohen, M. X. (2014) EEG artifacts: Their detection, influence, and removal. *Analyzing neural time series data: Theory and practice* (pp. 87-96) The MIT Press.
- [82] Rangayyan, R. M. (2015b) Filtering removal of artifacts. In T. Samad (Ed.), *Biomedical signal analysis* (Second edition ed., pp. 91-100). Hoboken (N.J.): Wiley.

- [83] St. Louis, E. K., & Frey, L. C. (2016) Appendix 4. common artifacts during EEG recording. *Electroencephalography* (pp. 84-92) American Epilepsy Society.
- [84] Stern, J. M., & Engel, J. (2013p) Artifacts. In *Atlas of EEG patterns* (Ed.), (2nd ed). Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- [85] Delorme, A., Sejnowski, T., & Makeig, S. (2007) Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, 34(4), 1443-1449.
- [86] Goodfellow, I., Bengion, Y., & Courville, A. (2016) Introduction. *Deep learning* (pp. 1-8) MIT Press. URL: <https://www.deeplearningbook.org/contents/intro.html>. Accessed 25.4.2019.
- [87] Deng, L., & Dong, Y. (2014) Definitions and background. *Deep learning: Methods and applications* (pp. 198-204) Now Publishers. URL: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/DeepLearning-NowPublishing-Vol7-SIG-039.pdf> Accessed 10.5.2019.
- [88] Bengio, Y. (2009) Learning deep architectures for AI. *Foundations and Trends® in Machine Learning*, 2(1), 1-127.
- [89] Haykin, S. (2009) Introduction. *Neural networks and learning machines* (3rd ed., pp. 1-46) URL: https://cours.etsmtl.ca/sys843/REFS/Books/ebook_Haykin09.pdf. Accessed 12.5.2019.
- [90] Diamantaras, K., Duch, W., Iliadis, L. S., & Kanade, T. (2010a) Preface. In T. Kanade, K. Diamantaras, W. Duch & L. S. Iliadis (Eds.), *Artificial neural networks – ICANN 2010* (pp. 5-7). Thessaloniki, Greece: Springer Berlin Heidelberg.
- [91] Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018) Activation functions: Comparison of trends in practice and research for deep learning.
- [92] LeCun, Y., Bengio, Y., & Hinton, G. (2015) Deep learning. *Nature*, 521(7553), 436.
- [93] Kriesel, D. (2005a) Components of artificial neural networks. A brief introduction to deep learning (pp. 33-38)
- [94] Diamantaras, K., Duch, W., Iliadis, L. S., & Kanade, T. (2010b) The use of feed forward neural network for recognizing characters of dactyl alphabe. In K. Diamantaras, W. Duch, L. S. Iliadis & T. Kanade (Eds.), *Artificial neural networks – ICANN 2010* (pp. 114-117). Thessaloniki, Greece: Springer Berlin Heidelberg.

- [95] Deng, L., & Yu, D. (2014) Three classes of deep learning networks. Deep learning: Methods and applications (pp. 214-229) Now Publishers, Inc.
- [96] Moolayil, J. (2019) Chapter 5 - tuning and deploying deep neural networks. Learn keras for deep neural networks: A fast-track approach to modern deep learning with python. Place of publication not identified: Apress.
- [97] Sze, V., Chen, Y., Yang, T., & Emer, J. S. (2017) Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE, 105(12), 2295-2329.
- [98] Agostinelli, F., Hocquet, G., Singh, S., & Baldi, P. (2018) From reinforcement learning to deep reinforcement learning: An overview. Braverman readings in machine learning. key ideas from inception to current state (pp. 298-328) Springer.
- [99] Goodfellow, I., Bengio, Y., & Courville, A. (2016a) Deep feedforward networks. Deep learning (pp. 164-223). MIT Press: URL: <http://www.deeplearningbook.org/contents/mlp.html>. Accessed 22.4.2019.
- [100] Kriesel, D. (2007) Network topologies. Brief introduction to neural networks (pp. 39-42) URL: <http://www.dkriesel.com>. Accessed 22.4.2019.
- [101] Gori, M., & Tesi, A. (1992) On the problem of local minima in backpropagation. IEEE Transactions on Pattern Analysis & Machine Intelligence, (1), 76-86.
- [102] Kriesel, D. (2005b) The perceptron, backpropagation and its variants. A brief introduction to neural networks (pp. 71-106)
- [103] Caruana, R., Lawrence, S., & Giles, C. L. (2001) Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. Advances in Neural Information Processing Systems, 402-408.
- [104] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014) Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), 1929-1958.
- [105] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017) Aggregated residual transformations for deep neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1492-1500.
- [106] Goodfellow, I., Bengio, J., & Courville, A. (2016b) Convolutional networks. Deep learning. MIT Press. URL: <http://www.deeplearningbook.org>. Accessed 20.5.2019.
- [107] LISA Lab, University of Montreal. (2015) Convolutional neural networks (LeNET). Deep learning tutorial (pp. 50-64) URL: <http://deeplearning.net/tutorial/deeplearning.pdf>. Accessed 21.5.2019.

- [108] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989) Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4), 541-551.
- [109] Zeiler, M. D., & Fergus, R. (2014) Visualizing and understanding convolutional networks. *European Conference on Computer Vision*, 818-833.
- [110] Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019) Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119, 3-11.
- [111] Yang, J., Nguyen, M., San, P., Li, X. & Krishnaswamy, S. (2015). Deep convolutional neural networks on multichannel time series for human activity recognition. *IJCAI International Joint Conference on Artificial Intelligence, 2015-*, pp. 3995-4001.
- [112] Murad, A., & Pyun, J. (2017) Deep recurrent neural networks for human activity recognition. *Sensors*, 17(11), 2556.
- [113] Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2019) 1D convolutional neural networks and applications: A survey.
- [114] Yadav, N., Yadav, A., & Kumar, M. (2015) Recurrent neural networks. An introduction to neural network methods for differential equations. Dordrecht: Springer.
- [115] Gers, F. A., Eck, D., & Schmidhuber, J. (2002) Applying LSTM to time series predictable through time-window approaches. *Neural nets WIRN vietri-01* (pp. 193-200) Springer.
- [116] Malhotra, P., Vig, L., Shroff, G. & Agarwal, P. (2015). *Long Short Term Memory networks for anomaly detection in time series*.
- [117] Hochreiter, S., & Schmidhuber, J. (1997) Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- [118] Schuster, M., & Paliwal, K. K. (1997) Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), 2673-2681.
- [119] Graves, A. (2013) Generating sequences with recurrent neural networks.
- [120] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation.
- [121] Donahue, J. (2017) Long-term recurrent convolutional networks for visual recognition and description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 677-691. DOI:10.1109/TPAMI.2016.2599174

- [122] Pascanu, R., Gulcehre, C., Cho, K., & Bengio, Y. (2013) How to construct deep recurrent neural networks.
- [123] Lecun, Y. (1998) Efficient backprop. *Neural Networks: Tricks of the Trade*, 1524, 9-50.
- [124] Goodfellow, I., Bengio, Y., & Courville, A. (2016c) Selecting hyperparameters. *Deep learning* (pp. 428) MIT press.
- [125] Goodfellow, I., Bengio, Y., & Courville, A. (2016d) Capacity, overfitting and underfitting. *Deep learning* (pp. 111-112) MIT press.
- [126] Janocha, K., & Czarnecki, W. M. (2017) On loss functions for deep neural networks in classification.
- [127] Goodfellow, I., Bengio, Y., & Courville, A. (2016e) Gradient-based learning. *Deep learning* (pp. 181) MIT press.
- [128] Goodfellow, I., Bengio, Y., & Courville, A. (2016f) Gradient-based optimization. *Deep learning* (pp. 86) MIT press.
- [129] Goodfellow, I., Bengio, Y., & Courville, A. (2016g) Stochastic gradient descent. *Deep learning* (pp. 147) MIT press.
- [130] Martens, J. (2010) Deep learning via hessian-free optimization. *International Conference on Machine Learning*, 27 735-742.
- [131] Glorot, X. & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research*, 9, pp. 249-256.
- [132] He, K., Zhang, X., Ren, S., & Sun, J. (2015) Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer Vision*, 1026-1034.
- [133] Goodfellow, I., Bengio, Y., & Courville, A. (2016h) Greedy layer-wise unsupervised pretraining. *Deep learning* (pp. 526-529) MIT press.
- [134] Goodfellow, I., Bengio, Y., & Courville, A. (2016i) Transfer learning domain adaptation. *Deep learning* (pp. 536) MIT press.
- [135] Bengio, Y., Lamblin, P., Popovici, D. & Larochelle, H. (2007). Greedy layer-wise training of deep networks. *Advances in Neural Information Processing Systems*, pp. 153-160.
- [136] Bengio, Y. (2012) Practical recommendations for gradient-based training of deep architectures. *Neural networks: Tricks of the trade* (pp. 437-478) Springer.

- [137] Masters, D. & Luschi, C. (2018). *Revisiting Small Batch Training for Deep Neural Networks*.
- [138] Zhang, C., Recht, B., Bengio, S., Hardt, M. & Vinyals, O. (2019). *Understanding deep learning requires rethinking generalization*.
- [139] Prechelt, L. (2012). Early stopping - But when? *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7700, pp. 53-67. DOI:10.1007/978-3-642-35289-8-5
- [140] Bishop, C. M. (2006) Early stopping. *Pattern recognition and machine learning* (pp. 259) springer.
- [141] Klein, A. (2016) Fast bayesian optimization of machine learning hyperparameters on large datasets.
- [142] Shahriari, B. (2016) Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1), 148-175. DOI:10.1109/JPROC.2015.2494218
- [143] MathWorks Inc. (2019) Deep learning for signal processing with MATLAB. MathWorks Inc. URL: <https://se.mathworks.com/campaigns/offers/deep-learning-for-signal-processing-white-paper.html>. Accessed 5.6.2019.
- [144] Bishop, C. M. (1995a) Pre-processing and post-processing. *Neural networks for pattern recognition* (pp. 296) Oxford university press.
- [145] Bishop, C. M. (1995b) Input normalization and encoding. *Neural networks for pattern recognition* (pp. 298) Oxford university press.
- [146] Yang, Q. (2006) 10 challenging problems in data mining research. *International Journal of Information Technology & Decision Making*, 5(4), 597-604. DOI:10.1142/S0219622006002258
- [147] Chawla, N. (2004) Editorial: Special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, 6(1), 1-6. DOI:10.1145/1007730.1007733
- [148] Japkowicz, N. (2000) The class imbalance problem: Significance and strategies. *Proc. of the Int'l Conf. on Artificial Intelligence*,
- [149] Kubat, M., & Matwin, S. (1997) Addressing the curse of imbalanced training sets: One-sided selection. *International Conference on Machine Learning*, 97 179-186.
- [150] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002) SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.

- [151] Wang, S., Liu, W., Wu, J., Cao, L., Meng, Q., & Kennedy, P. J. (2016) Training deep neural networks on imbalanced data sets. *International Joint Conference on Neural Networks (IJCNN)*, 4368-4374.
- [152] Moniz, N., Branco, P., & Torgo, L. (2017) Resampling strategies for imbalanced time series forecasting. *International Journal of Data Science and Analytics*, 3(3), 161-181.
- [153] Hussein, R., Palangi, H., Ward, R., & Wang, Z. J. (2018b) Epileptic seizure detection: A deep learning approach.
- [154] Sharmila, A. (2018) Epilepsy detection from EEG signals: A review. *Journal of Medical Engineering & Technology*, 42(5), 368-380. DOI:10.1080/03091902.2018.1513576
- [155] Aarabi, A. (2006) Automated neonatal seizure detection: A multistage classification system through feature selection based on relevance and redundancy analysis. *Clinical Neurophysiology*, 117(2), 328-340. DOI:10.1016/j.clinph.2005.10.006
- [156] Polat, K., & Güneş, S. (2007) Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast fourier transform. *Applied Mathematics and Computation*, 187(2), 1017-1026.
- [157] Subasi, A. (2007) EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, 32(4), 1084-1093. DOI:10.1016/j.eswa.2006.02.005
- [158] Mirowski, P. W. (2008) Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. DOI:10.1109/MLSP.2008.4685487
- [159] Chandaka, S., Chatterjee, A., & Munshi, S. (2009) Cross-correlation aided support vector machine classifier for classification of EEG signals. *Expert Systems with Applications*, 36(2), 1329-1336.
- [160] Yuan, Q. (2011) Epileptic EEG classification based on extreme learning machine and nonlinear features. *Epilepsy Research*, 96(1-2), 29-38. DOI:10.1016/j.eplepsyres.2011.04.013
- [161] Khan, Y. U. (2012) Automated seizure detection in scalp EEG using multiple wavelet scales. DOI:10.1109/ISPCC.2012.6224361
- [162] Nicolaou, N. (2012) Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. *Expert Systems with Applications*, 39(1), 202-209. DOI:10.1016/j.eswa.2011.07.008
- [163] Weidong Zhou. (2013) Epileptic seizure detection using lacunarity and bayesian linear discriminant analysis in intracranial EEG. *IEEE Transactions on Biomedical Engineering*, 60(12), 3375-3381. DOI:10.1109/TBME.2013.2254486

- [164] Kumar, A. (2014) Machine learning approach for epileptic seizure detection using wavelet analysis of EEG signals. DOI:10.1109/MedCom.2014.7006043
- [165] Song, Z., Wang, J., Cai, L., Deng, B. & Qin, Y. (2016) *Epileptic seizure detection of electroencephalogram based on weighted-permutation entropy*. DOI:10.1109/WCICA.2016.7578764
- [166] Bugeja, S., Garg, L. & Audu, E. (2016) A novel method of EEG data acquisition, feature extraction and feature space creation for early detection of epileptic seizures. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2016-*, pp. 837-840. DOI:10.1109/EMBC.2016.7590831
- [167] Hosseini, M., Tran, T. X., Pompili, D., Elisevich, K. & Soltanian-Zadeh, H. (2017) *Deep Learning with Edge Computing for Localization of Epileptogenicity Using Multimodal rs-fMRI and EEG Big Data*. DOI:10.1109/ICAC.2017.41
- [168] Acharya, U. R. (2018a) Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in Biology and Medicine*, 100, 270-278. DOI:10.1016/j.compbimed.2017.09.017
- [169] Hussein, R., Palangi, H., Ward, R., & Wang, Z. J. (2018c) Epileptic seizure detection: A deep learning approach.
- [170] Liang, W., Pei, H., Cai, Q. & Wang, Y. Scalp EEG epileptogenic zone recognition and localization based on long-term recurrent convolutional network. *Neurocomputing*. DOI:10.1016/j.neucom.2018.10.108
- [171] Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H. & Adeli, H. (2018) Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in Biology and Medicine*, 100, pp. 270-278. DOI:10.1016/j.compbimed.2017.09.017
- [172] Thodoroff, P., Pineau, J. & Lim, A. (2016) *Learning Robust Features using Deep Learning for Automatic Seizure Detection*.
- [173] Ansari, A. H. (2018) Neonatal seizure detection using deep convolutional neural networks.
- [174] Vidyaratne, L., Glandon, A., Alam, M., & Iftekharruddin, K. M. (2016) Deep recurrent neural network for seizure detection. 2016 International Joint Conference on Neural Networks (IJCNN), 1202-1207.
- [175] Yao, X., Cheng, Q., & Zhang, G. (2019) A novel independent RNN approach to classification of seizures against non-seizures.

- [176] Léonard, F. (2007) Phase spectrogram and frequency spectrogram as new diagnostic tools. *Mechanical Systems and Signal Processing*, 21(1), 125-137.
- [177] Badshah, A. M., Ahmad, J., Rahim, N., & Baik, S. W. (2017) Speech emotion recognition from spectrograms with deep convolutional neural network. 2017 International Conference on Platform Technology and Service (PlatCon), 1-5.
- [178] Rangayyan, R. M. (2015c) The short -time fourier transform. *Biomedical signal analysis : A case-study approach* (2nd ed., pp. 400-402). Hoboken (N.J.): IEEE Press.
- [179] Allen, J. B. (1977) A unified approach to short-time fourier analysis and synthesis. *Proceedings of the IEEE*, 65(11), 1558-1564. DOI:10.1109/PROC.1977.10770
- [180] Randall, R. B. (1987) 2.1.1 complex notations. *Frequency analysis* (3rd ed, pp. 16-18). Naerum: Brüel & Kjaer.
- [181] Arnold, M. (2003) Fast fourier transforms using the complex logarithmic number system. *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, 33(3), 325-335. DOI:10.1023/A:1022236132192
- [182] Rangayyan, R. M. (2015d) The short-time Fourier transform. *Biomedical signal analysis* (Second edition ed., pp. 478-481). Hoboken (N.J.): Wiley.
- [183] Rao, H. (2013) Complex sleep apnea. *Current Treatment Options in Neurology*, 15(6), 677-691. DOI:10.1007/s11940-013-0260-7
- [184] Goenka, A. (2018) Comparative sensitivity of quantitative EEG (QEEG) spectrograms for detecting seizure subtypes. *Seizure: European Journal of Epilepsy*, 55, 70-75. DOI:10.1016/j.seizure.2018.01.008
- [185] James, J. F. (2011) Two-dimensional fourier transform. *A student's guide to fourier transforms : With applications in physics and engineering* (3rd ed, pp. 97-100). Cambridge; New York: Cambridge University Press.
- [186] Rosen, B., & Wald, L. (2006) MR image encoding. URL: http://mriquestions.com/uploads/3/4/5/7/34572113/imageencoding_mit_courseware_wald.pdf. Accessed 20.6.2019.
- [187] Jayaraman, S., Esakkirajan, S., & Veerakumar, T. (2009) Properties of 2D-DFT. *Digital image processing* (pp. 159-173) Tata McGraw Hill Education Private Limited.
- [188] Cireşan, D., Meier, U., Masci, J., Gambardella, L. & Schmidhuber, J. (2011) Flexible, high performance convolutional neural networks for image classification. *IJCAI International Joint Conference on Artificial*

Intelligence, pp. 1237-1242. DOI:10.5591/978-1-57735-516-8/IJCAI11-210

- [189] Donahue, J., Hendricks, L. A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K. & Darrell, T. (2017). Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), pp. 677-691. DOI:10.1109/TPAMI.2016.2599174.
- [190] Ebrahimi Kahou, S., Michalski, V., Konda, K., Memisevic, R. & Pal, C. (2015) *Recurrent Neural Networks for Emotion Recognition in Video*. DOI:10.1145/2818346.2830596.
- [191] Frazier, P. I. (2018). *A Tutorial on Bayesian Optimization*.
- [192] Snoek, J., Larochelle, H., & Adams, R. P. (2012) Practical bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 2951-2959.
- [193] Geisser, S. (1975) The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70(350), 320-328.
- [194] Stone, M. (1974) Cross- validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2), 111-133.
- [195] Armah, G. K. (2014) A deep analysis of the precision formula for imbalanced class distribution. *International Journal of Machine Learning and Computing*, 4(5), 417-422. DOI:10.7763/IJMLC.2014.V4.447
- [196] NCSS Statistical Software.Binary diagnostic tests – single sample. NCSS Statistical Software. URL: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Binary_Diagnostic_Tests-Single_Sample.pdf. Accessed 24.2.2020.
- [197] Davis, J. & Goadrich, M. (2006) *The relationship between Precision-Recall and ROC curves*.
- [198] Zhou, M. (2018) Epileptic seizure detection based on EEG signals and CNN.(convolutional neural network)(electroencephalography)(report). *Frontiers in Neuroinformatics*, 12 DOI:10.3389/fninf.2018.00095
- [199] Tsiouris, K M. (2018) A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. *Computers in Biology and Medicine*, 99, 24-37. DOI:10.1016/j.combiomed.2018.05.019
- [200] Nejedly, P., Kremen, V., Sladky, V., Nasser, M., Guragain, H., Klimes, P., Cimbalnik, J., Varatharajah, Y., Brinkmann, B. H., Worrell, G. A. (2019) Deep-learning for seizure forecasting in canines with epilepsy. *Journal of Neural Engineering*, 16(3), p. 036031. DOI:10.1088/1741-2552/ab172d

- [201] Ramgopal, S. (2014) Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. *Epilepsy & Behavior*, 37, 291-307. DOI:10.1016/j.yebeh.2014.06.023
- [202] Kiriakopoulos, E., & Shafer, P. O. (2017b) Atonic seizures. URL: <https://www.epilepsy.com/learn/types-seizures/tonic-clonic-seizures>. Accessed 16.2.2019.
- [203] Kiriakopoulos, E., & Shafer, P. O. (2017c) Tonic-clonic seizures. URL: <https://www.epilepsy.com/learn/types-seizures/tonic-clonic-seizures>. Accessed 16.2.2019.
- [204] Kiriakopoulos, E., & Shafer, P. O. (2017d) Clonic seizures. URL: <https://www.epilepsy.com/learn/types-seizures/clonic-seizures>. Accessed 16.2.2019.
- [205] Kiriakopoulos, E., & Shafer, P. O. (2017e) Tonic seizures. URL: <https://www.epilepsy.com/learn/types-seizures/tonic-seizures>. Accessed 16.2.2019.
- [206] Hernandez, A. (2016) Doose syndrome. URL: <https://www.epilepsy.com/learn/types-epilepsy-syndromes/doose-syndrome>. Accessed 16.2.2019.
- [207] Blume, W. T., Lüders, H. O., Mizrahi, E., Tassinari, C., van Emde Boas, W., & Engel Jr, E., Jerome. (2001) Glossary of descriptive terminology for ictal semiology: Report of the ILAE task force on classification and terminology. *Epilepsia*, 42(9), 1212-1218.
- [208] Kiriakopoulos, E., & Shafer, P. O. (2017f) Absence seizures. URL: <https://www.epilepsy.com/learn/types-seizures/absence-seizures>. Accessed 16.2.2019.
- [209] Kiriakopoulos, E., & Shafer, P. O. (2017g) Atypical absence seizures. URL: <https://www.epilepsy.com/learn/types-seizures/atypical-absence-seizures>. Accessed 16.2.2019.
- [210] Devinsky, O., & Sirven, J. I. (2013) Myoclonic seizures. URL: <https://www.epilepsy.com/learn/types-seizures/myoclonic-seizures>. Accessed 16.2.2019.
- [211] Hernandez, A. W. (2015) Epilepsy with myoclonic-absences. URL: <https://www.epilepsy.com/learn/types-epilepsy-syndromes/epilepsy-myoclonic-absences>. Accessed 16.2.2019.
- [212] Hernandez, A. (2018) Epilepsy with eyelid myoclonias (jeavons syndrome). URL: <https://www.epilepsy.com/learn/types-epilepsy-syndromes/epilepsy-eyelid-myoclonias-jeavons-syndrome>. Accessed 16.2.2019.
- [213] Kiriakopoulos, E., & Shafer, P. O. (2017h) Focal onset aware seizures (simple partial seizures). URL: <https://www.epilepsy.com/learn/types->

- [seizures/focal-onset-aware-seizures-aka-simple-partial-seizures](#). Accessed 16.2.2019.
- [214] Kiriakopoulos, E., & Shafer, P. O. (2017i) Focal onset impaired awareness seizures (complex partial seizures). URL: <https://www.epilepsy.com/learn/types-seizures/focal-onset-impaired-awareness-seizures-aka-complex-partial-seizures>. Accessed 16.2.2019.
- [215] Kiriakopoulos, E., & Shafer, P. O. (2017j) Focal to bilateral tonic-clonic seizures (secondarily generalized seizures). URL: <https://www.epilepsy.com/learn/types-seizures/focal-bilateral-tonic-clonic-seizures-aka-secondarily-generalized-seizures>. Accessed 16.2.2019.
- [216] Holmes, G. L. (2004) Infantile spasms. URL: <https://www.epilepsy.com/learn/professionals/about-epilepsy-seizures/overview-epilepsy-syndromes/infantile-spasms>. Accessed 17.2.2019.
- [217] Kerrigan, J. F., & Iyengar, S. (2017) Gelastic and dacrystic seizures. URL: <https://www.epilepsy.com/learn/types-seizures/gelastic-and-dacrystic-seizures>. Accessed 17.2.2019.
- [218] Shafer, P. O. (2016) New terms for seizure classification. URL: <https://www.epilepsy.com/learn/types-seizures/new-terms-seizure-classification>. Accessed 17.2.2019.
- [219] American Epilepsy Society. (2016) New guideline for treatment of prolonged seizures in children and adults. URL: https://www.aesnet.org/about_aes/press_releases/guidelines2016. Accessed 17.2.2019.

10. APPENDICES

- Appendix 1. Generalized motor seizures
- Appendix 2. Generalized non-motor seizures
- Appendix 3. Focal seizure based on awareness of consciousness
- Appendix 4. Focal motor seizures
- Appendix 5. Focal non-motor seizures
- Appendix 6. Characteristics and EEG features of different types of status epilepticus
- Appendix 7. Treatment chart of tonic-clonic status epilepticus
- Appendix 8. EEG registration methods
- Appendix 9. Epileptiform activities
- Appendix 10. Common sources of artefacts
- Appendix 11. Epileptiform patterns
- Appendix 12. Description of the CHB-MIT dataset

Appendix 1. Generalized motor seizures

Table 9. Generalized Motor Seizures [12, 23, 33, 202-207]

Generalized motor seizure type	Characteristics
1. Tonic-clonic	Sudden loss of consciousness and falling
	Shriek caused by spasm of laryngeal muscles
	Breathing stops and skin color turns to blue
	Foam coming out of mouth
	Tonic-clonic has two phases: tonic contraction phase with duration of 20 seconds followed by bilateral clonic contraction of somatic muscles phase with duration of 1-2 minutes (lasting more than 5 minutes is a medical emergency)
	Attacks are followed by sleep state
2. Clonic	Rhythmical jerking of the body or the body parts, regularly repetitive, involves the same muscle groups
3. Tonic	Tonic seizures involve tonic stiffening and cocontraction of agonist and antagonist musculature
	Tonic seizures are brief with duration of 15 seconds but may escalate into tonic status epilepticus
4. Myoclonic	Involves sudden brief jerks or twisting of limbs, face or axial musculature, usually preserved consciousness
5. Myoclonic-tonic-clonic	Involves sudden brief jerks or twisting of limbs, face or axial musculature, usually preserved consciousness
	Tonic seizures involve tonic stiffening and cocontraction of agonist and antagonist musculature
	Rhythmical jerking of the body or the body parts
6. Myoclonic-atonic	Involves sudden brief jerks or twisting of limbs, face or axial musculature, usually preserved consciousness
	Involves sudden loss of muscle tone, with variable severity from head nods/drops to complete loss of axial posture with falling and injury
7. Atonic	Involves sudden loss of muscle tone, with variable severity from head nods/drops to complete loss of axial posture with falling and injury

	Lasting 1 to 2 seconds
8. Epileptic spasms	Autonomic dysfunction symptoms such as pallor, flushing, sweating, pupillary dilatation, lacrimation, changes in respiratory and heart rate
	Brief head nods, whereas other seizures consist of violent flexion of the trunk, arms, and legs
	Frequently occurs in clusters

Appendix 2. Generalized non-motor seizures

Table 10. Generalized non-motor seizures (Absence) [12, 23, 33, 207-212]

Generalized non-motor seizure type	Characteristics
1. Typical absence	Absence seizures are the most common seizures and involve impaired awareness or consciousness for 5 to 10 seconds followed by immediate recovery
	Lacks a prodromal aura or postictal state
	Sometimes involve eye blinking, twitching of eyelids, corners of mouth or limbs
	Sometimes involve short automatisms such as pupil dilation, pallor, flushing, sweating, salivating, piloerection, urinary incontinence
2. Atypical absence	Unusual or not typical
	Staring but able to respond
	Eye blinking, chewing movements, lip smacking or jerking movements
	Begins and ends gradually
	Usually last 5 to 30 seconds most often more than 10 seconds
3. Myoclonic absence	Significant and continuous rhythmic shock-like jerks of a muscle or group of muscles
	Typical absence symptoms
4. Eyelid myoclonic	Brief and repetitive myoclonic jerks of the eyelids
	Brief absence
	Typically lasts less than 6 seconds

Appendix 3. Focal seizures based on awareness of consciousness

Table 11. Focal Seizures based on awareness of consciousness [23 p.13-14, 33, 207, 213-215]

Focal seizure type	Characteristics
1. Focal Onset Aware Seizure (Aura)	Remains alert and able to interact
	Lasting seconds to less than 2 minutes
	Motor symptoms: spasms of limbs, symptoms related to posture of view, head and body, vocalization or stop talking
	Sense symptoms: sensing, sight or hearing sensory, smell, taste, dizziness or imbalance
	Autonomic symptoms: heart pumping, blush, paleness, sweating etc.
	Psychological symptoms: memory disturbance, fear, anxiousness, euphoria, illusion
2. Focal Onset Impaired Awareness Seizures	Not aware of surroundings
	Lasting 1 to 2 minutes
	Attack may start with aura and lead to loss of consciousness or it can immediately start with loss of consciousness
	Seizures may involve automatisms
	Confusion and feeling tired for 5 to 15 minutes or longer after seizure
3. Focal to Bilateral Tonic Clonic Seizure	Focal seizure that turns into Bilateral tonic clonic seizure (starts from one area of the brain, then spreads to both sides of the brain)

Appendix 4. Focal motor seizures

Table 12. Focal motor seizures [12, 202, 204, 205, 207, 210, 216]

Focal motor seizure type	Characteristics
1. Automatisms	Coordinated, repetitive, motor movements
2. Atonic	Involves sudden loss of muscle tone, with variable severity from head nods/drops to complete loss of axial posture with falling and injury
	Lasts 1 to 2 seconds
3. Clonic	Rhythmical jerking of the body or the body parts, regularly repetitive, involves the same muscle groups
4. Epileptic spasms (Infantile spasms)	Autonomic dysfunction symptoms such as pallor, flushing, sweating, pupillary dilatation, lacrimation, changes in respiratory and heart rate
	Brief head nods, a sudden flexion, extension, or mixed extension-flexion of predominantly proximal and truncal muscles
	Frequently occurs in clusters
5. Hyperkinetic	Involving predominantly proximal limb or axial muscles in irregular sequential ballistic movements and increases rate of ongoing movement
	Often occurs during sleep and repeatedly, with intervals of a few seconds
6. Myoclonic	Involuntary shock-like jerks of a muscle or group of muscles
7. Tonic	Sustained increase in muscle contraction
	Lasts a few seconds to minutes

Appendix 5. Focal non-motor seizures

Table 13. Focal non-motor seizures [12, 217, 218]

Focal non-motor seizure type	Characteristics
1. Autonomic	Change in heart rate, breathing, color of skin, pupil dilation, pallor, flushing, sweating, salivating, piloerection
2. Behavior arrest	Blank stare, stop talking, stop moving
3. Cognitive	Confusion, slowed thinking, problems with talking and understanding
4. Emotional	Sudden fear, dread, anxiety, pleasure
5. Sensory	Change in hearing, vision, taste, feeling of numbness, tingling or pain

Appendix 6. Characteristics and EEG features of different types of status epilepticus

Table 14. Characteristics and EEG features of different types of status epilepticus [36]

Type of status epilepticus	Characteristics	EEG features
1. Tonic-clonic status epilepticus	Tonic, clonic or tonic-clonic seizures	Generalized epileptiform discharges
2. Absence status epilepticus	Absence seizure that lasts from hours to days with possible automatisms or myoclonic twitch	Continuous generalized 2-3 Hz multi spike slow wave discharges
3. Myoclonic status epilepticus	Prolonged myoclonic seizures without impaired consciousness	Single and polyspike complex or sometimes without any EEG correlation
4. Focal status epilepticus with impaired consciousness	Begins with series of separate focal absence seizures with automatisms, stopping of functions and speech impediment	Cyclic epileptiform discharges
5. Focal status epilepticus without impaired consciousness	Motor, somatosensory or aphasia symptoms	Focal epileptiform discharges or sometimes without any EEG correlation if discharge activity is in small area

Table 15. Operational dimensions. t_1 indicates time point where the seizure should be regarded as continuous seizure activity and t_2 indicates the time point where the risk of long-term consequences start to increase

Type of status epilepticus	t_1	t_2
Tonic-clonic status epilepticus	5 min	30 min
Focal status epilepticus with impaired consciousness	10 min	60 min
Absence status epilepticus	10-15 min	Unknown

Appendix 7. Treatment chart of tonic-clonic status epilepticus

Table 16. Treatment chart of tonic-clonic status epilepticus [36, 37, 219]

Time	Phase	Condition	Treatment	Clinical examination
5 min	Stabilization phase, first aid outside of hospital	Premonitoring stage		Blood sugar quick measurement
5-10 min	Initial therapy phase, treatment outside of hospital	Early status epilepticus	Antiepileptic drugs: benzodiazepines	SpO_2 , blood pressure, ECG, Blood sugar, a basic blood count, concentration of antiepileptic in blood, functions of liver and kidneys
30-60 min	Second therapy phase	Established status epilepticus	Antiepileptic drugs: fosphenytoin, valproate, levetiracetam	Thorax x-ray, computer tomography, brain and cerebrospinal fluid, EEG
Over 60 min	Third therapy phase in intensive care unit	Refractory status epilepticus	Antiepileptic drugs: propofol, midazolam, thiopental, ketamine Other: General anaesthesia	Continuous EEG monitoring for observing termination of epileptiform activity and deepness level of anesthesia
Over 24 hours		Super-refractory status epilepticus		

Appendix 8. EEG registration methods

Table 17. EEG registration methods [43]

EEG paradigm	Characteristics	Target
Basic EEG	<ul style="list-style-type: none"> • 19-26 electrodes placed on scalp's skin. • Takes 20 – 60 minutes • Often includes polygraph channels and video. 	<ul style="list-style-type: none"> • Basic measurement tool
Ambulatory EEG	<ul style="list-style-type: none"> • Several days lasting EEG recording outside of hospital. • No video. 	<ul style="list-style-type: none"> • Investigation of uncertain seizures
EEG monitoring	<ul style="list-style-type: none"> • Long lasting (hours, days or weeks) brain monitoring in which number of electrodes are limited • Sometimes includes polygraph and video • Often includes aEEG 	<ul style="list-style-type: none"> • Brain monitoring in intensive care and monitoring treatment of status epilepticus
Video-EEG monitoring	<ul style="list-style-type: none"> • EEG registration in epilepsy unit • Long lasting (days) • Includes polygraph and video 	<ul style="list-style-type: none"> • Diagnostics of epilepsy and preparation for surgery
Dense array EEG	<ul style="list-style-type: none"> • Big number (maximum of 256) of electrodes with various polygraph channels 	<ul style="list-style-type: none"> • Ultrafast frequencies are used for investigation of location of epileptogenic
Full-band EEG	<ul style="list-style-type: none"> • Measured from 0 (infraslow) to hundreds (ultrafast) of Hertz 	<ul style="list-style-type: none"> • High frequencies for investigation of location of epileptogenic and low only for research
Invasive EEG	<ul style="list-style-type: none"> • Electrodes are placed inside of a scalp 	<ul style="list-style-type: none"> • Investigation of location of epileptogenic and preparation for surgery
EEG-EP / ERP	<ul style="list-style-type: none"> • EEG and action potentials are measured simultaneously • Action potentials are calculated from EEG signals via averaging 	<ul style="list-style-type: none"> • Often used with sensory action potentials (EEG-SEP) • After brain injury or resuscitation

Appendix 9. Epileptiform activities

Table 18. Epileptiform ictal activity [52]

Name	Activity
Epileptiform ictal activity	No observable change Obscured by artifacts Polyspikes Fast spike activity or repetitive spikes Low voltage fast activity Polysharp-waves Spike-and slow-waves Rhythmic activity Slow wave of large amplitude Irregular delta or theta activity Burst-suppression pattern Electrodecrement change DC-shift High frequency oscillation (HFO) Disappearance of ongoing activity Other ictal EEG pattern

Table 19. Morphology of epileptiform interictal activity [52]

Name	Morphology
Epileptiform interictal activity	Spike Spike-and-slow-wave Runs of rapid spikes Polyspikes Polyspikes-and-slow-wave Sharp-wave Slow sharp-wave High frequency oscillation (HFO) Hypsarrhythmia - classic Hypsarrhythmia - modified

Table 20. More detailed morphologies of generalized and focal interictal epileptiform discharges [69]

Name	Morphology
Focal interictal epileptiform discharges	Spikes Sharps Sharp waves Polyspikes Multiple spike complexes Spike and (slow) waves Sharp and (slow) waves Polyspike and (slow) waves Rolandic discharges Multifocal independent spike discharges (MISD) Independent multifocal spike discharges (IMSD) Second bilateral synchrony
Generalized interictal epileptiform discharges	3 per second spike (and slow) wave complexes Atypical spike and slow waves Slow spike and waves Petit mal variant Fast spike and waves Polyspike and (slow) wave complexes Multiple spike complexes Spike and dome complexes Dart and dome complexes Hypsarrhythmia Modified Hypsarrhythmia

Appendix 10. Epileptiform patterns

Table 21. EEG epileptiform patterns [68 pp. 214-215, 69, 71, 75]

EEG epileptiform patterns	Subcategories
Ictal epileptiform patterns	Ictal patterns for focal-onset seizures
	Ictal patterns for generalized-onset seizures
Interictal epileptiform patterns	Focal interictal epileptiform discharges
	Generalized interictal epileptiform discharges
Periodic epileptiform discharges	Periodic lateralized epileptiform discharges (PLEDs)
	Bilateral independent periodic epileptiform discharges (BIPLEDs)
	Bilateral periodic epileptiform discharges (BiPEDs)
Benign epileptiform transients of sleep	None

Appendix 11. Common sources of artifacts

Table 22. Common sources of EEG artifacts [52]

Biological artifacts	Non-biological artifacts
Eye blinks	50 or 60 Hz
Eye movements	Induction or high frequency
Nystagmus	Dialysis
Chewing artifact	Artificial ventilation artifact
Sucking artifact	Electrode pops
Glossokinetic artifact	Salt bridge artifact
Rocking or patting artifact	Other artifact
Movement artifact	
Respiration artifact	
Pulse artifact	
ECG artifact	
Sweat artifact	
EMG artifact	

Appendix 12. Description of the CHB-MIT dataset

Table 23. Description of the CHB-MIT dataset

Case	Gender	Age (years)	# of seizures	Duration of recordings (hh:mm:ss)
chb01	F	11	7	40:33:08
chb02	M	11	3	35:15:59
chb03	F	14	7	38:00:06
chb04	M	22	4	156:03:54
chb05	F	7	5	39:00:10
chb06	F	1.5	10	66:44:06
chb07	F	14.4	3	67:03:08
chb08	M	3.5	5	20:00:23
chb09	F	10	4	67:52:18
chb10	M	3	7	50:01:24
chb11	F	12	3	34:47:37
chb12	F	2	27	20:41:40
chb13	F	3	12	33:00:00
chb14	F	9	8	26:00:00
chb15	M	16	20	40:00:36
chb16	F	7	10	19:00:00
chb17	F	12	3	21:00:24
chb18	F	18	6	35:38:05
chb19	F	19	3	29:55:46
chb20	F	6	8	27:36:06
chb21	F	13	4	32:49:49
chb22	F	9	3	31:00:11
chb23	F	6	7	26:33:30
chb24	-	-	16	21:17:47
Total			185	979:56:07