Optimization of spike and wave discharge prediction

Björn Budde

Münster - 2020

Biologie

Dissertationsthema

Optimization of spike and wave discharge prediction

Inaugural-Dissertation zur Erlangung des Doktorgrades
der Naturwissenschaften im Fachbereich Biologie
der Mathematisch-Naturwissenschaftlichen Fakultät
der Westfälischen Wilhelms-Universität Münster

vorgelegt von

Björn Budde
aus

Paderborn, Deutschland

Münster – 2020

Dekanin: Prof. Dr. Susanne Fetzner

Erster Gutachter: PD Dr. Thomas Seidenbecher

Zweiter Gutachter: Prof. Dr. Wolf-Michael Weber

Tag der mündlichen Prüfung: 18.02.2021

Tag der Promotion: 16.04.2021

List of contents

List of abbreviations

∑	sum
2ch	prediction based on two channels
3ch	prediction based on three channels
AI	artificial intelligence
ANN	artificial neural network
ANOVA	analysis of variance
ATN	anterior nucleus
BCI	brain-computer interface
CAE	Childhood absence epilepsy
CC	two cortical regions
CCC	three cortical regions
CCT	two cortical and one thalamic region
Cm	centromedian thalamus
CNN	convolutional neural network
cRTN	caudal thalamic reticular nucleus
CT	one cortical and one thalamic region
CTT	one cortical and two thalamic regions
ctx	cortex
DBS	deep brain stimulation
ECoG	electrocorticography
EEG	electroencephalogram
FFT	fast fourier transformation
f_t	forget gate
GABA	gamma-aminobutyric acid
GAERS	genetic absence epileptic rats from Strasbourg
GCCC	three cortical regions in GAERS
GHB	gamma-hydroxybutyric acid
g_t	main gate
HCN	hyperpolarization-activated cyclic nucleotide-gated channel
HP	high pass
HVS	high voltage spindle
i	channel
I_H	cation current
ILAE	International League against Epilepsy
IPSP	inhibitory postsynaptic potentials
I_T	low threshold calcium currents
i_t	input gate
LFP	local field potentials
LP	low pass
LSTM	long short-term memory
MCCC	three cortical regions located in the motorcortex

NEC	non-epileptic control
O_t	output gate
Po	posterior thalamic nucleus
ReLU	rectified linear unit
RF	random forest
RNN	recurrent neural networks
rRTN	rostral thalamic reticular nucleus
RTN	thalamic reticular nucleus
S.E.M.	standard error of the mean
SVM	support vector machine
SWD	spike and wave discharge
t	time
Tanh	tangens hyperbolicus
TT	two thalamic regions
TTT	three thalamic regions
VPM	ventral-postero-medial nucleus
W	wavelet energy
WAG/Rij	Wistar Albino Glaxo Rijswijk
WHO	World Health Organization
Δ	delta
ϕ	activation function

1. Introduction

1.1. Electroencephalogram and brain research

The electroencephalogram (EEG), developed and introduced by Hans Berger 1924 (Berger, 1929, 1939), is a diagnostic tool for the supervision of brain activity. Electrodes, either placed on the scalp of a person or directly implanted into the brain, are used to record signals of summed electrical (field potential) activity. Therefore, it is a multifunctional tool suited to investigate states of the patient's brain activity during clinical applications such as neurology, anesthesiology or intensive care. EEG is also used in a wide variety of research areas including neuroscience, biomedical engineering and neural engineering (Chen, 2014; Motamedi-Fakhr *et al.*, 2014; Lüttjohann *et al.*, 2016; He *et al.*, 2018; Li *et al.*, 2020). Specialists are required to decipher specific patterns belonging either to a pathological or physiological background. The automation of this process would make the application of EEG less dependent on professionals, quicker and reproducible to a higher degree. The structured form of EEG signals enables an automatized learning algorithm called Machine learning be used as a tool to process these data. Up to this point, several Machine learning algorithms are used for the differentiation and detection of EEG signals. These signals could correspond to the intention of a person to move his or her arm, thought processes or the detection of epileptic seizures (Aljalal *et al.*, 2020; Siddiqui *et al.*, 2020). Typically, EEG data is preprocessed in some ways to reduce the information load, thus increasing the pattern recognition speed. There are numerous techniques to process EEG signals. One example, also applied in this thesis, is the wavelet transformation (see Figure 1). With the application of wavelet transformation, time-series signals like EEG can be divided into their frequency components that appear in the analyzed time interval. In contrast to the standard Fast Fourier Transformation (FFT), which also decomposes the EEG into its

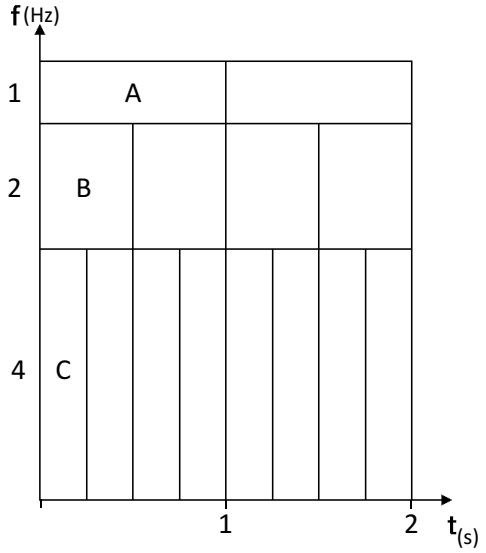


***Figure 1: Wavelet transformation output of EEG**. The signal is split into time (t) and frequency (f) components. A wavelet transformation of 2 seconds would be split into A (1 Hz), B (2 Hz) and C (4 Hz).*

frequency content, the wavelet analysis does not use a fixed duration of the analysis window, but adjusts the duration of the window to the frequency being analyzed. Consequently, the wavelet analysis is regarded as more suitable for the analysis of rapidly fluctuating signals like the EEG (Ende *et al.*, 1998). Spectral components of EEG signals are generally grouped into five main bands including: delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-32 Hz) and gamma (32-100 Hz). Each category is correlated to a specific vigilance state (Figure 2); e.g. delta waves are mostly detected during deep sleep and are paired with loss of awareness (Abhang *et al.*, 2016). However, during sleep several other components possessing a higher frequency content like short lasting sleep spindles with a frequency between 8-15 Hz or K-Complexes are present as well (Cash *et al.*, 2009). Activities like thinking or other cognitive tasks consist of faster frequency components (Figure 2).

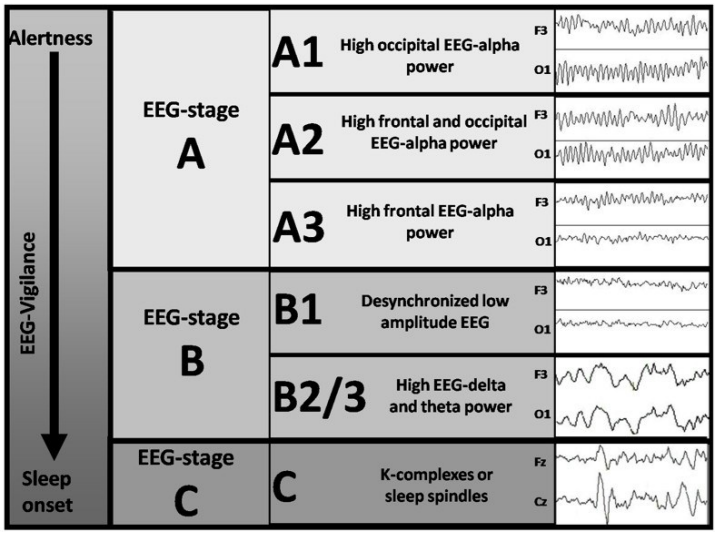


Figure 2: EEG of vigilance stages. *The change in vigilance from alertness to sleep onset (Top to bottom) is depicted in the left column. EEG is divided into six different stages of vigilance (A1-C) with its corresponding 2 second EEG samples frontal (F3) and occipital (O1) (Brainclinics, 2019).*

Small alterations in this intricate crosstalk can impair the accurate functioning of the brain. One of the neurological diseases, which originates from misfiring of neurons, is known as epilepsy. Still, EEG is the gold standard technique for the diagnosis of epilepsy and it is frequently used in research investigating the underlying pathology of epilepsy (Coenen, 1995; Kessler & McGinnis, 2019; Johnson, 2019).

EEG signals are limited to the spatial resolution in close proximity to the recording site of pyramidal cells (Murakami & Okada, 2006). To investigate communication between brain regions also relevant for the study of epilepsy, several connectivity analyses can be applied to the recorded EEG. In this thesis, the product of the wavelet power, which is

assessed in EEG traces recorded at different brain locations, is used as a measure for synchrony between the areas.

1.2. Epilepsy and seizures

Epilepsy is one of the oldest known neurological disorders, which dates back to 4000 BC (Fischer-Elfert, 2005). Briefly, 10% of all humans are experiencing an unprovoked seizure event during their lifetime according to the World Health Organization (WHO). Spontaneous seizures are the main characteristic of epilepsy, however seizures in general can be provoked by various influences that interrupt the neuronal crosstalk within the brain. Patients suffering from epilepsy are characterized by frequent unprovoked seizures. The International League against Epilepsy (ILAE) defined that patients suffer from epilepsy once they have two or more unprovoked seizures within a day, one unprovoked seizure and the probability of creating further seizures or the diagnosis of a known epilepsy syndrome (Fisher *et al.*, 2017). During a seizure, multiple behavioral aspects can happen from twitching or jerking movements to loss of consciousness. Tonic-clonic seizures display the mainstream view of epileptic seizure, in which the patient experiences characteristic spasms of the rump and extremities (clonic) accompanied by a sudden stiffness (tonic) (ILAE). These behavioral phenomena go along with specific hypersynchronous activity in the brain that can be picked up by an EEG. Epileptic seizures are grouped into either focal or generalized seizures. Focal seizures are found in one region of the brain with simple connotation causing for example twitching or sensational changes and complex connotation that might cause loss of consciousness depending on the severity of the seizure. Generalized seizures are subdivided in two categories. Primary generalized seizures have no clear onset zone and are seen all over the cortex from the start of the seizure. Secondary generalized seizures are also considered focal due to the start in focal brain regions with a later generalization. Such secondary generalized seizures are seen in an epilepsy syndrome called Childhood Absence Epilepsy. The combination of seizure types and behavioral symptoms, as well as the cause of the epileptic seizures specifies different epilepsy syndromes. Their origin can be divided into five different categories, namely: Metabolic, infectious, structural, genetic or of unknown cause (Pitkänen *et al.*, 2006).

1.3. Childhood Absence Epilepsy

Childhood Absence Epilepsy (CAE) is seen in children between the age of 4 and 12 years. Between 10 - 17% of all children diagnosed with epilepsy develop this specific form of epilepsy (Loiseau *et al.*, 1995). The distribution of this syndrome is not equally dispersed within humans. As it is also known in other forms of epilepsy, girls show a higher prevalence of CAE in comparison to boys (Waaler *et al.*, 2000; Christensen *et al.*, 2005). CAE counts as a generalized non-convulsive form of epilepsy. Most of the time it is considered as a mild or benign form of epilepsy with low risk of injuries (van Luijtelaar *et al.*, 2017). Children are observed to show an impairment of their consciousness during seizure with little movements in their eyes and facial muscle (Panayiotopoulos, 1999).

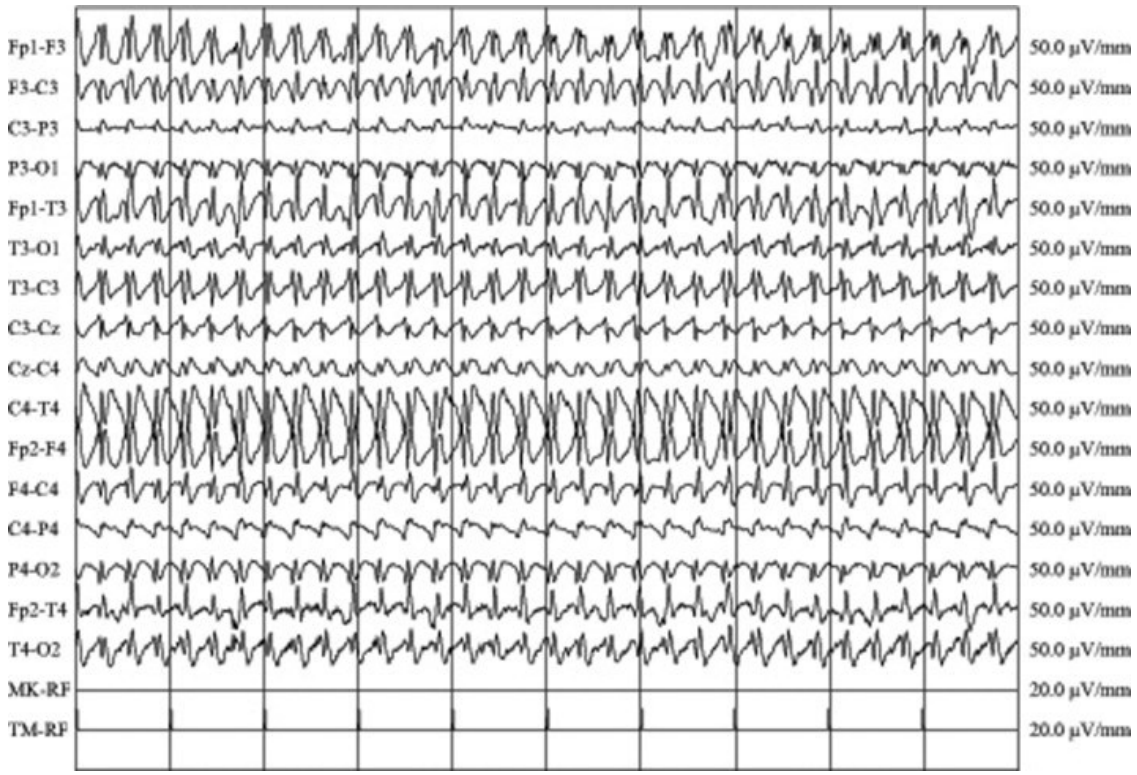


***Figure 3: EEG of a child with Childhood Absence Epilepsy**. EEG pattern contains typical spike and wave discharge (SWD) (taken from Cerminara* et al.*, 2012).*

In the EEG, absence seizures show a characteristic pattern within the whole brain, which is build up by a slow wave component and a spike component (Figure 3). This pattern is also known as spike and wave discharge (SWD) and has a predominant frequency between 2.5-4 Hz in humans (Figure 4) (Pitkänen *et al.*, 2006; Kessler & McGinnis, 2019). As the SWD pattern can be recorded all over the cortex, SWDs are commonly classified as generalized seizures. Moreover, as they arise rapidly from a seemingly normal background EEG, they have been regarded as sudden unpredictable events for a long

time. Detecting and sorting these specific patterns requires a specialist who can distinguish the known pathological rhythm of CAE (Pitkänen *et al.*, 2006).

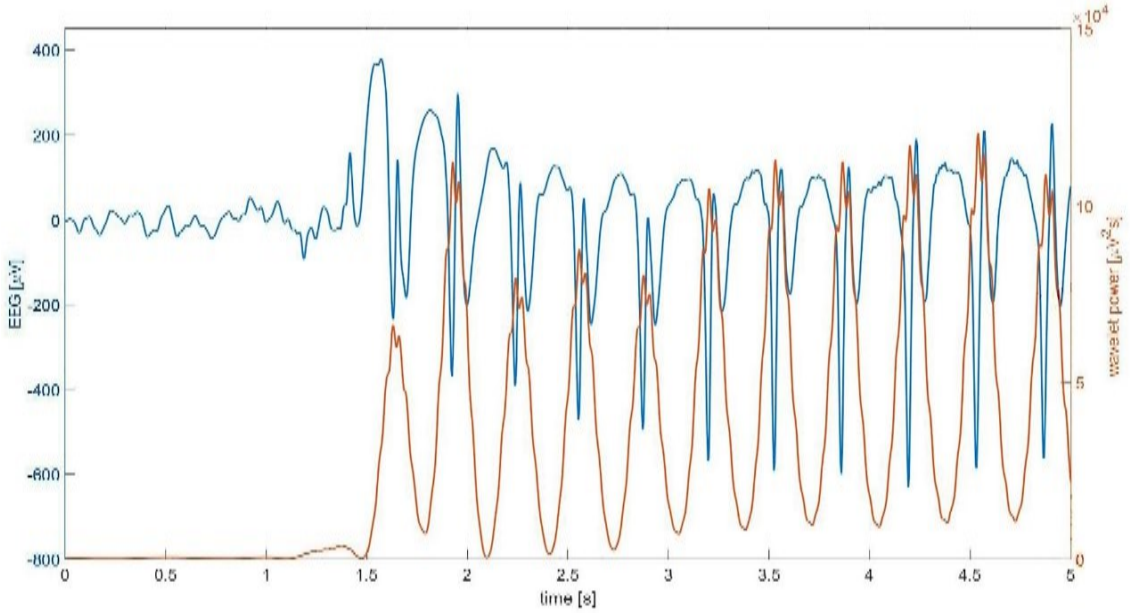


Figure 4: Close up view on SWD pattern showing repetitive spike and wave complexes seen in human EEG (Blue)*. Wavelet power depicted in Orange. (adapted from Glaba* et al.*, 2020).*

On average, seizure events take 10 seconds. However, episodes can last from 1 to 44 seconds and are known for their suddenness of arrival and passing (Sadleir *et al.*, 2006; Matricardi *et al.*, 2014). Patients can be observed to have brief staring spell during seizures, which can be misinterpreted as normal child behavior. During seizures, children can remain unresponsive to external stimuli like touch or questions addressed to them. Activities like chewing or playing are halted during seizures but continue after the event. The loss of consciousness might not be complete. After an episode, there are cases in which they do not remember what happened during it or even recognize the seizure event in general. However, sometimes they are able to recall the presented stimuli (Gloor, 1986; Rosenow *et al.*, 1998; Drinkenburg *et al.*, 2003). Although CAE seizures are mostly considered harmless, they do influence the child in their daily life and contribute to a deterioration of quality of life (Kessler & McGinnis, 2019). Time periods of reduced consciousness or transition phases between sleep and wakefulness make up the lion share of the overall seizure occurrence. Seizure events happen repeatedly during the day and are classified to occur up to several hundred times a day summing up to a considerable time interval. Most of the time, seizure appearance is linked to epochs of

low activity, also attributing the time spend in school (Kessler & McGinnis, 2019). Although CAE is a mild form of epilepsy, patients still show cognitive, behavioral and psychiatric comorbidities and are often affected by attention deficit hyperactivity disorders and anxiety (Caplan *et al.*, 2008; Shinnar *et al.*, 2017). At the beginning of adolescence, CAE patients have a remission of their symptoms with a chance of 56% to 95%. CAE patients with no remission make up 15%. They form other types of seizure, which have a more malign character like juvenile myoclonic epilepsy with its generalized tonic seizures, myoclonic seizures or persistent absence seizures (Wirrell *et al.*, 1997; Callenbach *et al.*, 2009; Puka *et al.*, 2020).

CAE patients have a high responsive rate of 60-95% to the antiepileptic medications available (Wirrell *et al.*, 1997; Grosso *et al.*, 2005). Although the gold standard prescribed medication was already introduced in the 1950's, valproic acid and ethosuximide are still the main first-line medications (Wheless *et al.*, 2007; Glauser *et al.*, 2010). There are several other medications available but valproic acid and ethosuximide have the most reasonable side effects paired with high effectiveness. Ethosuximide is considered the main choice for the initial starting drug due to its mild adverse effects in comparison to valproic acid. Where valproic acid targets the T-type sodium channel, ethosuximide is addressing the T-type calcium channel. Both medications work as antagonists on the respective channels. Altering the calcium and sodium channel results in changes of the polarization preventing the formation of burst firing and thereby the formation of SWDs (Coulter *et al.*, 1989; Gören & Onat, 2007). Interestingly, CAE symptoms of patients aggravate by using standard drugs like carbamazepine and phenytoine available for generalized convulsive and partial seizures (Panayiotopoulos, 1999).

Up to this point, CAE is considered an idiopathic disease with a polygenic background (Crunelli & Leresche, 2002; Asadi-Pooya & Homayoun, 2020). There have been several debates about involved regions but currently, there are two major theories on the origin of SWDs. While both theories coincide with the imperative importance of a functioning thalamo-cortical system, the first theory focuses on the thalamus as the most important brain structure, while the second considers the cortex as more important (see paragraph 1.5 for more detail).

1.4. The thalamo-cortical system

As mentioned above, both theories on the origin of SWD coincide with the imperative importance of a functioning thalamo-cortical system. The brain's function is heavily dependent on the internal interconnection it creates between brain regions. These interconnections are also known as neural circuits or loops, which carry information to manage bodily functions such as motor instructions, hormonal responses and many more (Purves *et al.*, 2001). Neuronal disorders like CAE are known to display pathological changes in the neural circuitry. These changes in the intricate circuits cause the formation of the SWD pattern. Connections between the cortical and thalamic cells are mostly reciprocal and excitatory. In between the cortex and thalamus lies the thalamic reticular nucleus (RTN) taking up a regulatory function with its mostly inhibitory connections produced by gamma-Aminobutyric acid (GABA)-producing neurons. Both inhibitory and excitatory signaling are involved in the feedforward and feedback connections in the cortico-thalamo-cortical network (Sherman & Guillery, 2005; McCafferty *et al.*, 2018). Thalamic nuclei involved in the somatosensory thalamo-cortical system can be classified

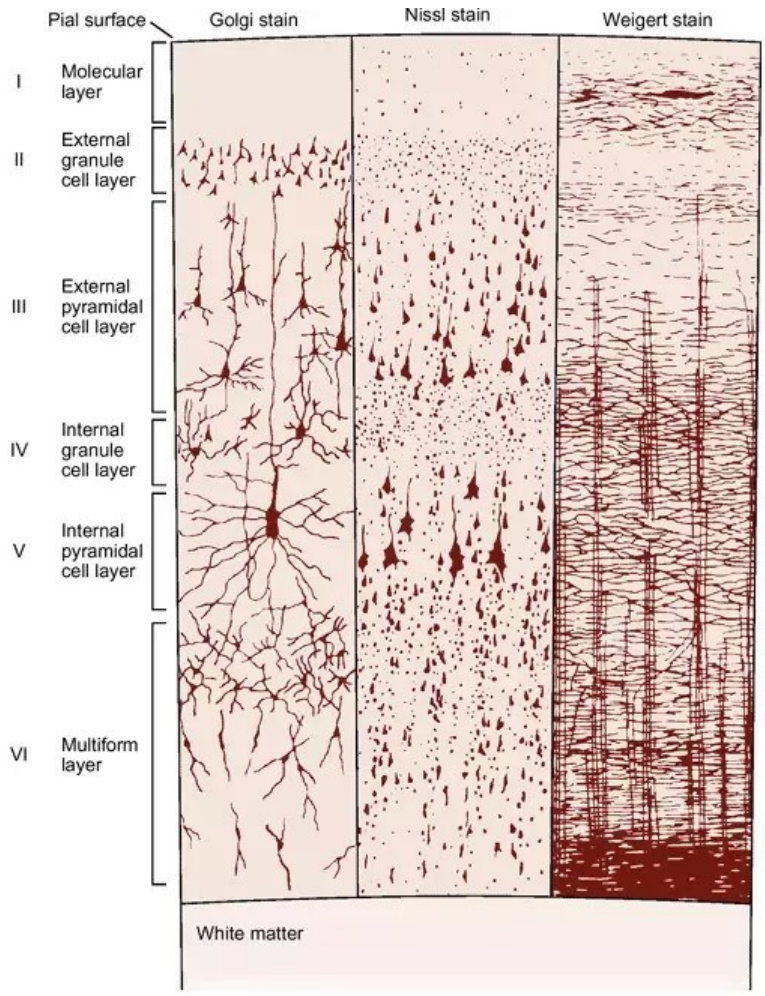


Figure 5. Cross section of the neocortex with its six layers. *Three different types of staining were used. Golgi staining to show the nervous tissue, Nissl staining to visualize the cell bodies and Weigert staining to show the myelin content (taken from Amaral, 2006).*

as either "first" order or "higher" order (Sherman & Guillery, 2002). First order nuclei receive driving inputs from subcortical areas, whereas higher order nuclei receive driving inputs from cortical areas.

The ventral-postero-medial nucleus (VPM) as an example is a first order nucleus and part of the cross talk between thalamic areas and cortical areas. It receives sensory information from the head and transmits them to the neurons of the somatosensory cortex in layer IV (Zhang & Deschênes, 1997; Oda *et al.*, 2004). The cerebral cortex has a unique morphology with its six known layers (I-VI) also known as the neocortex, each with a particular composition and a laminar build (Figure 5). These six layers are termed (I) molecular layer, (II) external granular layer, (III) external pyramidal layer, (IV) internal

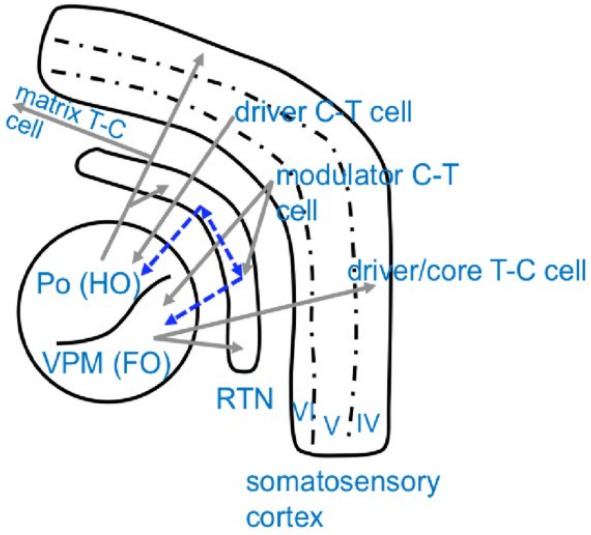


***Figure 6: Schematic display of the somatosensory thalamo-cortical loop**. GABAergic inhibitory neurons represented with blue arrows, glutamatergic excitatory neurons presented with grey arrows. HO = higher order; FO = first order (from Lüttjohann et al., 2015).*

granular layer, (V) internal pyramidal layer and (VI) the multiform layer. The thalamo-cortical system has multiple known connections to specific layers (Figure 6). Another thalamic nucleus involved in the somatosensory cortico-thalamo-cortical interaction is the posterior thalamic nucleus (Po), which acts as an example for a higher order nuclei. The PO receives driving inputs from the primary somatosensory cortex layer V and spreads the input reciprocally back onto the somatosensory cortex and other cortical areas (Sherman & Guillery, 2005). This reciprocal interaction between the thalamus and cortex is thought to be heavily involved in cortio-cortical communication (Sherman & Guillery, 2005). Neurons in layer VI of the somatosensory cortex send information to the VPM, PO and the RTN. The RTN is located between the cortex and thalamus like a barrier. It projects onto thalamic neurons via GABAergic interneurons and has an internal crosstalk via fast acting gap junctions as shown in Figure 6 (Pinault, 2004; Lüttjohann & van Luijtelaar, 2015). The RTN is divided into several compartments, which are responsible for different tasks (Guillery *et al.*, 1998). The caudal part of the RTN receives and projects inputs from and to the VPM. It is considered as the somatosensory part of the RTN (Pinault & Deschênes, 1998). The ATN is involved in the limbic loop receiving information from the rostral part of RTN (Gonzalo-Ruiz & Lieberman, 1995). Actions of the RTN via GABAergic neurons are involved in the change of thalamic firing. There are two main firing modes called “tonic” firing and “burst” firing. Tonic firing is considered the normal firing pattern during active wakefulness, whereas burst firing is only seen once

the neurons are hyperpolarized between -70 and -90 mV, which is common in phases of rest and sleep. It is known that a genetic rat model of CAE has more hyperexcitable cells in the somatosensory cortex, which makes them more susceptible to create burst firing. Burst firing prevents signal information to be passed to several brain compartments that would process the input. It is postulated that this hyperpolarization is the main culprit for the formation of seizures (Hughes & Crunelli, 2005). Hyperpolarization of thalamic cells by the RTN results in the activation of hyperpolarization-activated cyclic nucleotide-gated (HCN) channel, which produces a cation current called I_H. This causes the activation of low threshold calcium currents (I_T). Both the I_H and I_T currents depolarize thalamic cells leading to burst activation of Na^+/K^+-dependent fast action potentials. The depolarization by the action potentials inactivates I_T and I_H currents leading to the repolarization of the cell membrane. The action potential burst firing activates thalamic cells, which have collaterals to GABAergic RTN cells. The cycle starts anew with rhythmic hyperpolarization of thalamic cells by GABAergic signaling and thereby I_H and I_T dependent burst discharges (Steriade, 2003).

The cortex as information processing region extends collaterals additionally to the RTN, which in turn can modulate the thalamic cells to a certain degree. Therefore, it is theorized that the RTN is responsible for modulatory, limiting and synchronizer actions of the thalamo-cortical system (Steriade 1998, Hugenard and McCormick 2007). To understand seizure generation, it is important to unravel interaction with this intricate brain network.

1.5. Animal models of absence epilepsy

Up to this day, animal research is irreplaceable for studying types of neural disorder like CAE. However, not every model has the same face and predictive validity. For CAE, there are several ways to manifest absence seizures either chemically, electrically induced or genetically inbred (Pitkänen *et al.*, 2006). The first model of CAE was in feline animals. Absence seizures were induced by a chemical injection of penicillin into the carotid artery acting as a gamma-Aminobutyric acid ($GABA_A$) receptor antagonist (Avoli, 1995). The $GABA_A$ receptor is an ionotropic receptor and is also activated via ligand binding. In these

animals, absence epileptic seizures can be monitored in the EEG in the form of SWDs with a frequency of 3 Hz for a duration of 5-8 hours. Another chemically induced absence seizure is created by injecting gamma-Hydroxybutyric acid (GHB) into various test subjects like monkeys, cats and rats (Pitkänen *et al.*, 2006). The injection of GHB created a seizure with a mean frequency of 7-9 Hz varying from the human form of CAE. In contrast to the chemically induced absence epileptic seizures, rodent models are more common. The available different mouse models of CAE have a single point mutation mostly affecting calcium channel efficiency. The phenotypes of these mouse models varied strongly from each other and also from the human version (Pitkänen *et al.*, 2006). Therefore, the face validity is not as strong as preferred. At this point, two rat models, both derived from the Wistar rat strain, are common in the field of CAE research. The two models are termed Wistar Albino Glaxo Rijswijk (WAG/Rij) rats and Genetic Absence Epilepsy Rats from Strasbourg (GAERS) (Pitkänen *et al.*, 2006). Seizures are unprovoked spontaneous, generalized and react similarly to medication. The animal models face the same symptoms of loss of consciousness with occasional facial twitching and chewing (Depaulis & van Luijtelaar, 2006). Their genotype is not known; however, evidence points to the direction of an autosomal-dominant inheritance (Engel, 2001). Studies have identified specific chromosomal loci, which control features of the SWDs. Interestingly, chromosomes influencing these SWDs are specific for each rat model. In WAG/Rij rats, parts of chromosomes 5 and 9 are responsible for features of SWDs, whereas parts of chromosomes 4, 7 and 8 influence SWDs in GAERS (Gauguier 2004. Rudolf 2004). The polygenic character of Absence Epilepsy could also be shown through the modulation of cumulative seizure duration in GAERS by T-channel calcium channel mutations. Contradictory, this modulation was not found in WAG/Rij rats (Powell *et al.*, 2009). According to specialists, there are three distinctive features necessary for a validated animal model. The model needs to have the relevant features in the EEG pattern, the same pharmacological sensitivity, similar brain structure involved with the same behavior during seizure (Depaulis *et al.*, 2016). The rodent model reacts similarly to medication, as the valproic acid/ethosuximide prevents the seizure whereas carbamazepine and phenytoin aggravate it (Feddersen *et al.*, 2007; Nguyen-Michel *et al.*, 2009). Both models show secondary generalized seizures with a spike and wave morphology lasting on average ~10 seconds with frequencies between 7-11 Hz. During the first two seconds of

the seizure, higher frequency in comparison to the later parts of the seizure are shown (Durón *et al.*, 2005). Both rat strain have a polygenic background with unprovoked formation of fully formed SWDs from 3 month and 6 month for GAERS and WAG/Rij rats, respectively (Schridde & van Luijtelaar, 2004; Jarre *et al.*, 2017). These rodent models still display differences to the human CAE. In the rodent model SWDs occur relatively late in their development and are persistent even after reaching adolescence. While the WAG/Rij rats have a higher discharge frequency of SWDs and displays a higher power in the frequency band of 8-14 Hz in the preictal timeframe, the GAERS animal line shows a longer cumulative total duration, a higher number and a higher mean duration of SWDs. The GAERS animal line also has differences in the morphology of a single SWD. These single cycle SWD display more energy in the faster components (Akman *et al.*, 2010).

1.6. Spike and wave discharge onset

For a long time, it was debated where the starting point of the SWDs is. Advances in techniques and equipment allow a more elaborate view on theories done in the last century Currently, there are two major theories available concerning the origin of SWDs (Gloor, 1968; Meeren *et al.*, 2005). Both studies focus on the crosstalk between the thalamus and the cortex. However, these studies show a switching focus point. Gloor et al. proposed the cortico-reticular theory suggesting the thalamus to play a major role in the formation of absence seizures (Gloor, 1968). The thalamus is in parts responsible for the generation of sleep spindle oscillation. Gloor's theory states that the whole cortex is hyperexcitable picking up sleep spindle oscillations, which in turn transforms them into the pathological pattern of the SWDs. This thesis was based on the SWDs display in the feline model, which created SWDs once the $GABA_A$ antagonist penicillin was injected into the cortex. Interestingly, injection of penicillin only into the thalamus did not produce the SWD pattern (Gloor *et al.*, 1977). Upon transition towards sleep, the RTN creates a rhythmic oscillation frequency of 7-14 Hz corresponding to sleep spindles via inhibitory postsynaptic potentials (IPSP) being transmitted onto the thalamus via collaterals. Thalamic cells are consecutively hyperpolarized by the IPSP, send the oscillation back onto the cortex and via collaterals back onto the RTN. The RTN reactivates and restarts

the rhythmic firing (Steriade, 2003). Contrary to the suggestion that SWDs originate from sleep spindle, lesion studies of parts of the RTN showed that even in the absence of sleep spindles, SWDs were created (Meeren *et al.*, 2009). Human data showed that patients produce SWDs not exclusively during slow wave sleep but also during wakefulness (Pinault, 2003).

The current view on the origin of SWDs is based on research done by Meeren and colleagues (Meeren *et al.*, 2002). By using the local field potential of a cortical grid and thalamic nuclei, they showed a local driving area in the barrel field of the somatosensory cortex. The somatosensory cortex has the driving role in the first 500 ms whereas afterwards the cortex and thalamus alternated in the driving role. It is now postulated that the cortex comprises a hot spot, which is responsible for the SWD generation once certain unknown criteria are fulfilled (Meeren *et al.*, 2002, 2005; van Luijtelaar & Sitnikova, 2006). The onset zone and the highly repetitive SWD pattern make seizures of absence epilepsy (absence seizures) a good target for alternative ways to study it.

1.7. Towards treatment approach

The overarching goal when working in epileptology is the complete abolishment of seizures in general, which is not yet possible at this point. With the advances in technology it is now possible to record and store long periods in a multitude of animal models like GAERS, WAG/Rij rats, non-epileptic control (NEC) rats and many more (Pitkänen *et al.*, 2006). These datasets of EEG support the study of the epileptic brain and its circuitries, epileptogenesis, consequences of seizures, circadian rhythm (van Luijtelaar *et al.*, 2013; Smyk & van Luijtelaar, 2020) and the efficacy of new invention of medication or other techniques.

Although medications are available for CAE and reach high response rates, they can be accompanied with heavy adverse effects (Cunha, 2019). Recent advances in computational strength and techniques shine a new light on alternative treatment options. Deep brain stimulation (DBS) as one of the alternatives shows remarkable results in several neurological disorders like Parkinson's disease, dystonia and depression

(Sironia, 2011; Hariz, 2014; Cleary *et al.*, 2015; Vanhoecke & Hariz, 2017). This treatment option uses a neurostimulator to apply electrical impulses onto specific brain nuclei via implanted depth electrodes. Electrical impulses can modulate neuronal activity. DBS has already been used in several absence epilepsy animal models with varying success. High frequency stimulation showed suppression of neuronal activity once applied to either subcortical nuclei or cortical structures. Interestingly, open loop stimulation of a few short pulses can cause after-discharges to occur with the same common pattern as SWDs (Lüttjohann *et al.*, 2011, 2013). However, repetitive stimulation caused a passive conduct to the stimulation after a short period, pointing to a possible habituation mechanism (Kossoff, 2004; Osorio, 2005; Huang & van Luijtelaar, 2013). Based on the possible acclimatization of neurons to the DBS, a closed-loop system would be more effective in comparison to an open system. A closed-loop system would reduce the adaptation to the electrical stimulation, decrease the behavioral and physiological effects of the stimulation and might also increase the shelf life of the DBS application later on (van Luijtelaar *et al.*, 2017).

1.7.1. Detection of SWDs

Jandó and colleges took the first step towards the development of such a closed loop system by creating the first detection system for SWDs with the usage of an artificial neural network (ANN) (Jandó *et al.*, 1993). He trained a neural network with an offline dataset of several rat strains (F344, BN, F1, F2), using a High Voltage Spindle (HVS) approach as deciding factor between interictal, i.e. the non-epileptic EEG in-between seizures and ictal, i.e. the seizure. However, HVS show variations with either suppressed spikes or missing wave components. The ANN had a “small” structure with just one hidden layer, which is considered the simplest neural network. This sturdy used a sliding window of 10 ms of either raw EEG data or FFT electrocorticography (ECoG) data. Subsequently, the ANN preformed rather well with a sensitivity of 93-99% of the manually marked HVS with an error rate of falsely detected seizures of 18-40%. This detection was based on a trade-off between the number of positive detected seizures, false positive detected seizures and missed detections. This rather limited approach was impeded by the low computation power of that time. He postulated that with better-tuned

hyperparameter, like the number of hidden layers, neurons in each layer or other parameters, his ANN would produce better results (Jandó *et al.*, 1993). Another detection system was created by Westerhuis who used frontal-parietal ECoG measurements from the of WAG/Rij rats (Westerhuis *et al.*, 1996). This approach was based on the maximized steepness of 250 ms intervals. Once the steepness surpassed an automatically detected threshold value based on interictal periods of wakefulness for more than one second, a SWD was detected. His algorithm produced a sensitivity of 97% with minimal error rate (movement artifacts). Sitnikova and colleagues developed an offline sleep spindle detection based ECoG measurement that detected seizures by using an automatic detection based on complex Morlet wavelet-based algorithm (Sitnikova & van Luijtelaar, 2009). The differentiation between SWDs and sleep spindle was made possible by this implementation. Sleep spindle (7-14 Hz) of the rodent animal model WAG/Rij and GAERS produce similar frequencies to interspike frequencies of SWDs (8-12 Hz), which cause a high error rate. Another detection algorithm working on-line is based on the detection of SWDs by the frequency band of 30-50 Hz with a sensitivity and specificity reaching of 99% (Ovchinnikov *et al.*, 2010). Such high frequency components are characteristic for sharp EEG waveforms like a spike of the SWDs and are therefore well suited for the detection of seizure events.

Up to this point, there are several approaches to detect SWDs either working with raw EEG/ECoG, processed data like Morlet wavelet energy (Sitnikova & van Luijtelaar, 2009; Richard *et al.*, 2015) or harmonic analysis (Van Hese *et al.*, 2009) with varying success. The improvement of detection systems in general enables a more reproducible and stable way to analyze bigger datasets that are commonly produced during EEG measurements.

1.7.2. Prediction

The description of SWDs with its abrupt appearance was challenged by the detection of precursor discovered within thalamus and cortex (Panayiotopoulos, 1999; Sitnikova & van Luijtelaar, 2007; Ovchinnikov *et al.*, 2010). EEG and magnetoencephalography of CAE patients showed an increase in power of the low frequencies within several brain regions

(Inouye *et al.*, 1994; Gupta *et al.*, 2011). Research on WAG/Rij rats showed equal increases in power levels in both frontal ECoG and thalamic ECoG measurements (van Luijtelaar *et al.*, 2010). Transition of interictal to ictal intervals were analyzed with continuous wavelet transform (CWT) and Morlet wavelet to localize the time difference of onset and "precursor" (Sitnikova & van Luijtelaar, 2009; van Luijtelaar *et al.*, 2016). Results showed an increase in delta-theta (3-5 Hz) power, which almost simultaneously preceded the start of the SWDs in cortex and thalamus (van Luijtelaar et al., 2010). This change in frequency, now labeled as a precursor, was measureable up to two seconds prior to the transition phase. The precursor was present over a period of 500 ms in thalamic and cortical readings. Precursor activity was shown to be present in both cortical and thalamic areas throughout the brain. Most of the time, seizure events preceded a precursor in both thalamus and cortex (79%) but some events only showed a cortical precursor (11%) or a thalamic precursor (5%) with a small percentage of total absence of precursor activity (5%) (van Luijtelaar *et al.*, 2010, 2016). This occurrence of a defined precursor in both thalamic and cortical regions allowed a multitude of approaches to examine and alter the formation of the onset of SWDs. This approach allowed an investigation of precursor activity. However, seizures occurred during the transition of wakefulness to sleep, displaying sleep spindle oscillation (7-14 Hz) interfering with accurate prediction (van Luijtelaar *et al.*, 2016).

1.7.3. Prediction algorithm

The first on-line approach for the prediction of the seizure used 4 h periods of WAG/Rij rats EEG measurements with three electrodes of which two were placed in the cortex (layer IV, V) and one in the Po (Maksimenko *et al.*, 2017). Precursor activity was found in period preceding a seizure in the frequency band of 5-10 Hz. It was shown that this precursor activity produced local synchronization of neurons in the area of the electrode and global synchronization between cortex and thalamic nuclei. The prediction algorithm used the CWT and wavelet energies of three bands, which correspond to the precursor (5-10 Hz), and two frequency bands aiming to reduce some of the false detection. These two frequency bands resemble sleep spindle activity (7-20 Hz) and light slow wave sleep (3-5 Hz). A high occurrence of false predictions was noticed during light slow wave sleep,

where the brain is in a slightly hyperpolarized (Drinkenburg *et al.*, 1991). In comparison, passive wakefulness, deep slow wave sleep and active wakefulness did not produce high levels of false predictions. The algorithm defines that a positive detection is achieved once the precursor band energy exceeds a rat specific threshold and has more power than the two other frequency bands. This algorithm achieved a high sensitivity of SWD precursor detection, which is necessary for the application in medical devices (88-100%). However, there is still a multitude of error produced with this prediction model, which might be attributable to the chosen recoding sites (Cortex layer IV, V and Po) (Maksimenko *et al.*, 2017).

1.8. Artificial neural network and machine learning

With the advent of big data collections, a need arises for automated, reproducible processing. The idea of automated functioning and even thinking machines was already thought upon by authors during the 19th century (Butler, 1872). However, the current view on artificial intelligence and machine learning is rather an elaborated way of teaching an algorithm to decide upon a question based on statistical regression analysis (Davison, 2008). Although machine learning is not a new method, it received more attention in recent years based on the improvement of available computation systems.

With a more efficient way to analyze and react at the same time to certain signals, a method called brain-computer interface (BCI) comes into consideration. BCI systems act as a computation bridge and are already used in several tasks such as controlling robotic movements, artificial limbs or a cursor positioning on a computer (Bozinovski, 2017; Murphy *et al.*, 2017). With the usage of EEG data acquired directly from the patient, correlations to pathological and physiological changes could be traced back more easily to a specific EEG pattern.

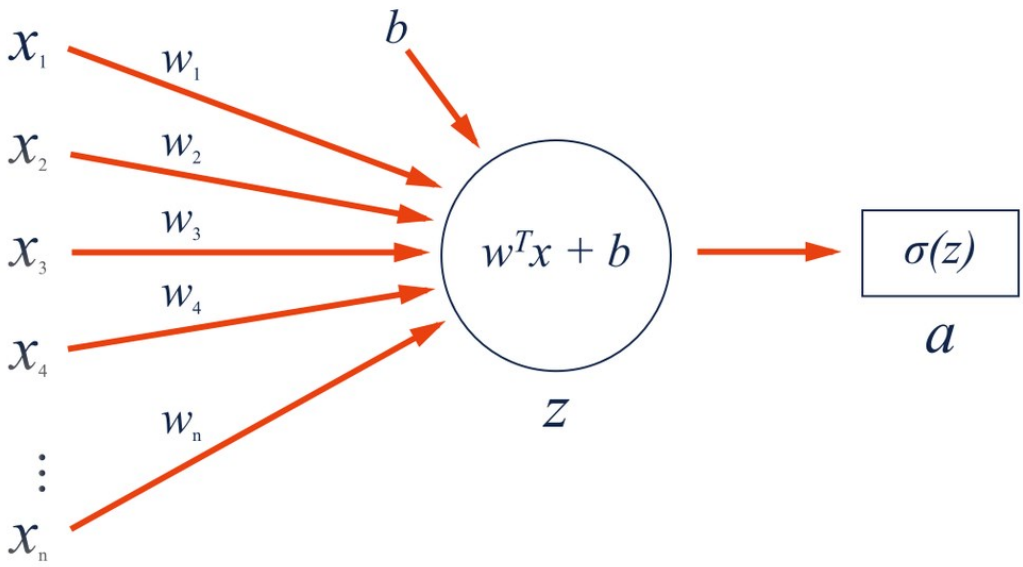


Figure 7: Schematic illustration of process within a single neuron*. (left) Input layer X with weighted connections w, (middle) processing (z) of input data corresponding to the inputs times weights with added bias, (right) output data (a) processed by an activation function.* (Castrounis, 2019).

EEG data are highly specific and minuscule changes need a certain set of techniques to be processed and analyzed. One of those techniques is machine learning. The history of machine learning by neural networks goes back to the 1940's, when Warren McCulloch and Walter Pitts laid the ground stone for further developments (McCulloch & Pitts, 1943). The simple principle from McCulloch was based on his so called "neurons", or computing units (Figure 7). During computing, neurons receive inputs of binary character (0, 1) and aggregate these within the neuron. Neuronal inputs have a certain range of modulation. They used the simple mechanical functions like "and", "or", "neither" and many more. With this set of binary input, decisions are made via a previous defined threshold. Once a value surpasses a certain threshold, the output from this neuron would then represent a decision corresponding to either a value of 0 or 1 (Mcculloch & Pitts, 1990). The first neural networks contained three layers; namely one input, one hidden and one output layer. This type of shallow neural network was commonly used until Hinton and colleagues showed the ability of deep neural networks to achieve a higher learning capacity (Hinton & Salakhutdinov, 2006; Polson *et al.*, 2015).

This simple setup of a neuronal network went through many changes over the years. However, the principle of McCulloch is still used in the machine learning algorithm available at this moment. Nowadays there is a multitude of variations available designed

for specific tasks. Specified neural networks can best work with the input previous defined for its type. The development of specified neural networks aimed to create a sorting mechanism for every kind of problem. Recently, deep neural networks with multiple hidden layers have achieved positive results in the recognition tasks with even better accuracies in comparison to humans on a benchmark dataset (Schmidhuber, 2015). Deep neural networks offer a greater capacity to extract features out of big data sets (Hinton & Salakhutdinov, 2006). The depth of a neural network needs to correspond to given features available in the dataset. Although large neural networks have a higher learning capacity, they are more resistant to the learning process (backpropagation) and thereby features are learned much slower.

The most common recognition tasks comprise of image, video, speech and text recognition. The advantages of machine learning algorithms are related to their ability to work with almost any type of data, the amount of data processed, its ease to train, the relatively short time each decision takes and the high accuracy.

1.9. Artificial intelligence, machine learning and deep learning

The current definition of artificial intelligence (AI) is a computer-based intelligence. In its early development, artificial intelligence resembled the learning process and reasoning equal to human beings. Machine learning as a subcategory of AI is defined as an algorithm that is able to "learn" by applying statistical methods without being explicitly programmed for a given task. An example of such a machine learning approach is the so called Random forest, which will be explained in more detail in the method section of this thesis. Several neural networks available and in use are considered deep learning algorithms. Neural networks are exclusive to deep learning that can adapt and adjust to process large data sets.

Deep learning algorithms are grouped in three main categories: supervised, unsupervised and reinforcement learning (Ayodele, 2010). Supervised learning has input and output data provided at the beginning. Classification is previously done to the dataset to measure the accuracy and loss of the training efficiency. The input data is split in a training

set and a test set in a commonly used ratio like the 70:30 split. This ratio is heavily dependent on the actual size of the original dataset. However, more advanced comparison methods are also available like cross validation. In supervised learning, the output can only represent the probability of a sample data to belong to a previously defined class.

Unsupervised learning has no classification done to the initial input data. The algorithm tries to create classes based on pattern recognition or find other relationships between inputs. This type of clustering tries to learn to distinguish similarities and differences of data points that are presented in the data and thereby group the data (Schmidhuber, 2015).

Reinforcement learning is a decision-based algorithm that reacts on given input. The algorithm has a goal set for a desired output. However, the way to achieve the output is unknown. The algorithm's first iteration has randomly chosen decisions that are scored and related to the final output. Decisions are either punished or rewarded after the iteration. In the next iteration, a new solution is evaluated on their effect on the final output. There are two types of reinforcement learning, with positive reinforcement highly rewarding good decisions or negative reinforcement penalizing bad decisions (Schmidhuber, 2015; Bajaj, 2020).

2. Aims of this study

Although there is evidence of precursor events in cases of CAE, up to this point there is no accurate way to detect them. Therefore, this thesis aims for a more accurate prediction of absence seizures by applying several techniques to improve the prediction of SWDs in the two Absence Epilepsy rat models WAG/Rij and GAERS. The algorithm previously published for the prediction produces high sensitivity of absence seizures prediction (Maksimenko *et al.*, 2017). However, a big problem with the algorithm is the high false positive rate, which resulted in a low specificity. Therefore, this thesis has three distinct aims:

I) **Elucidating the influence of different EEG recording site combinations within the thalamo-cortical system on the accuracy of the algorithm.** Additionally, the characteristics of false predictions are being compared to true prediction by applying spectral analysis.

II) **Optimization of SWD prediction by applying machine learning techniques.** This study will implement a sorting algorithm to further divide the true positive from the false positive prediction of the above-mentioned algorithm. For this, a standard machine learning algorithm (Random forest) is used.

III) **Using a deep learning neural network approach evaluating the prediction possibility from unbiased unprocessed raw EEG data.** For this, a deep learning neural network will be built of a combination of layers that are responsible for the extraction of temporal dependencies in EEG data and pattern recognition component distinguishing the features responsible for the formation of a SWD.

3. Materials and methods

The following chapter will concentrate on the techniques used to optimize SWD prediction within two genetic rat models of absence epilepsy, WAG/Rij and GAERS. At the beginning, the focus lay on the acquisition of two datasets in WAG/Rij rats and GAERS (see Paragraph 3.1). Further, the background and application of a wavelet-based algorithm (Maksimenko *et al.*, 2017) used in off- and online SWD prediction will be introduced (see Paragraph 3.2). This algorithm was used to compare the accuracy of SWD prediction (i.e. the sensitivity and specificity of SWD prediction), which can be obtained based on local field potential (LFP) recordings, acquired at several distinct recording sites within the thalamo-cortical system.

In a next step, several techniques of machine learning and deep learning will be introduced, which were applied to the output generated by the wavelet-based algorithm of Maksimenko et al. (2017), as well as to unprocessed raw LFP signals derived from GAERS and WAG/Rij rats (see Paragraph 3.3).

3.1. Data sets

3.1.1. Wistar Albino Glaxo Rijswijk (WAG/Rij)

A previous recorded dataset containing LFP recordings of 16 male WAG/Rij rats aged between 6 and 9 months, were analyzed (Lüttjohann & van Luijtelaar, 2012). In each rat, LFP signals were simultaneously measured in eight different brain structures within the thalamo-cortical system including the Po, VPM, caudal and rostral part of thalamic reticular nucleus (cRTN and rRTN), anterior thalamic nucleus (ATN) as well as layer IV, V, VI of the right somatosensory cortex (S1). LFP signals were gathered at a constant sample rate of 2048 Hz and filtered between 1 Hz high pass (HP) and 100 Hz (LP) low pass as well as by a 50 Hz notch filter, over a period of at least 4 hours in freely moving animals (Lüttjohann & van Luijtelaar, 2012). A WINDAQ-recording-system was used to digitize EEG signals (DATAQ-Instruments Inc., Akron, OH, USA). Rat movement was registered via a PIR detector (RK2000DPC LuNAR PR Ceiling Mount, Rokonet RISCO Group S.A., Drogenbos, BE).

A second dataset was acquired consisting of six male WAG/Rij rats aged between 6-9 month from the secondary motor cortex of layer Va and b and layer VI, representing an offset region located in relative distance to the onset zone of SWDs in S1.

3.1.2. Genetic absence epilepsy rat from Strasbourg (GAERS)

Electrophysiological *in vivo* recordings were performed in 6 male GAERS aged between 3 and 9 months over a period of 24 hours. They were born and raised at the animal facility of the Institute of Physiology I, Westphalian Wilhelms-University Münster. Prior to the surgery (electrode implantation see chapter 3.1.2.1.), the animals were housed in pairs with access to food and water *ad libitum*. Efforts were made to keep the discomfort of the animals as minimal as possible. After the surgery, they were placed separately to ensure integrity of the electrode head mount. All animals were housed with a 12:12 hours light:dark cycle (light on at 6 AM). All experimental procedures are in accordance with the guidelines and regulations of the council of the European Union (Directive 2010/63/EU) and approved by local authorities (LANUV Nordrhein-Westfalen; approval ID number: 84-02.04.2016.A297). The EEG data were collected unilateral from layer IV, V and VI of the right somatosensory cortex.

3.1.2.1. Surgery

The stereotactic surgery was performed under an intraperitoneal applied pentobarbital (Narcoren, 50 mg/kg; Boehringer Ingelheim Vetmedica GmbH, Ingeheim am Rhein, Germany) anesthesia using a commercial stereotactic frame (Stoelting Co., Wood Dale, IL USA) with bregma and lambda in a horizontal alignment. Local anesthetic lidocaine (Xylocain Gel 2%, AstraZeneca GmbH, Wedel, Germany) was applied to the ear canal, which was exposed to the stereotactic frame. Body temperature was controlled and conserved via a heating pad. Before trepanation, the skull was cleaned with H_2O_2 (4%) and saline (0.9%). Three electrodes were implanted within the right hemisphere in layer IV of the somatosensory cortex (anterior-posterior: -1.8, M/L: -3.6, H: -3.2 mm to bregma). Tip distance between electrodes was less than 1 mm to ensure local recording without interference due to direct contact. A reference and a ground electrode were

implanted epidural on top of the cerebellum. All coordinates were determined according to the stereotactic atlas of Paxinos and Watson (1998). Electrodes (Stainless steel electrode isolated with poliamide, Ø 0.2 mm) were fixed to the skull with first glass Ionomer dental cement and dental acryl cement (Pulpdent Corporation, Watertown MA, USA). The rats were preoperative injected with Rimadyl® (carprofen, 5 mg/kg body weight; zoetis, Louvain-la-Neuve, Belgium) 30 minutes prior to the surgical procedure. Another two treatments with Rimadyl® were applied postoperative 24 and 48 hours after surgery. EEG recordings started two weeks after surgical recovery.

3.1.2.2. Recording of local field potentials

Animals were placed in a 43x28x42 cm Plexiglas recording box and were connected to the recording leads for LFP recordings. Using a physiological amplifier (TD 90087, Radboud University Nijmegen, Electronic Research Group, The Netherlands), LFP signals were amplified. Movements of the subject were controlled by a Passive Infrared Registration system (PIR, PK2000DPC LuNAR PR Ceiling Mount, Rokonet). Data were filtered in the same way as LFP signals of the WAG/Rij rat dataset, except a lower sample frequency of 500 Hz, allowing faster data processing.

3.2. Data processing

3.2.1. Wavelet-based prediction algorithm of Maksimenko et al. (2017)

Both datasets were analyzed by an offline-version of the SWD prediction algorithm developed by Maksimenko et al. (2017), which predicts SWDs based on wavelet analyses, performed in three LFP signals, acquired simultaneously in three different recording sites. In addition, a modified offline-version of the SWD prediction algorithm predicting SWDs based on wavelet analyses, performed in two LFP signals, acquired simultaneously in two different recording sites were equally applied to the datasets.

The prediction based on two EEG traces screened 28 sets, which differed in the combination of recording sites within the thalamo-cortical system (see appendix 1b),

whereas for the prediction based on three EEG traces, 56 sets of different recording combinations (see appendix 1a) were tested on their SWD prediction performance.

At each recording location, the algorithm assessed the mean wavelength energy within a time window of 500 ms shifting along the complete EEG trace sample by sample (Figure 8). In each channel (i) and at each time step (t), the wavelet energy (W) within the frequency range of 5-10 Hz corresponding to the precursor ($W_{i\Delta S1}(t)$) is calculated using wavelet transformation with a modified Morlet mother function (Maksimenko *et al.*, 2017). This energy obtained in each channel is then multiplied to determine the occurrence of thalamo-cortical synchronization at each moment in time ($W_{\Delta S1}(t) = W_{1\Delta S1}(t)$ X $W_{2\Delta S1}(t)$ X $W_{3\Delta S1}(t)$).

At the same time, wavelet energy is calculated and multiplied in each channel for a frequency range of 3-5 Hz in accordance to the light slow wave sleep ($W_{\Delta S2}(t) = W_{1\Delta S2}(t)$ X $W_{2\Delta S2}(t)$ X $W_{3\Delta S2}(t)$) and within a frequency range of 7-20 Hz representing sleep spindles ($W_{\Delta S3}(t) = W_{1\Delta S3}(t)$ X $W_{2\Delta S3}(t)$ X $W_{3\Delta S3}(t)$).

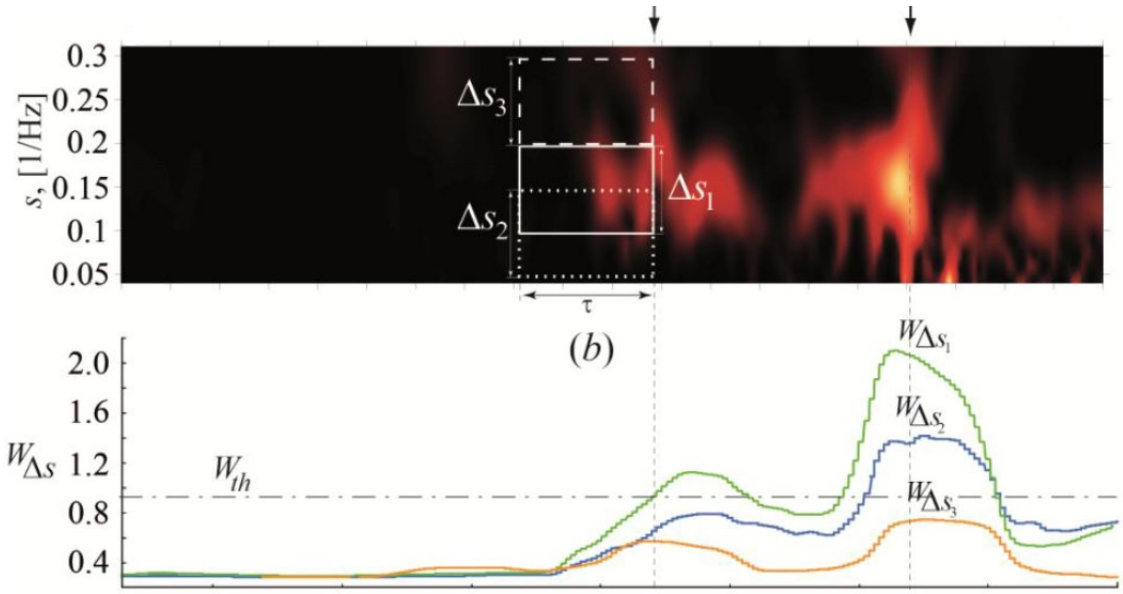


Figure 8: Wavelet analysis criterion for precursor detection. *(Top) Wavelet energy spectrum prior to a seizure. Mid-section arrow represents time point of precursor detection. Right-side arrow represents start of seizure. Rectangular window corresponds to frequency bands of defined: precursor ($W_{\Delta S1}$),* slow wave *sleep ($W_{\Delta S3}$) and sleep spindle ($W_{\Delta S2}$). (Bottom) Product of the three frequency bands with rat specific threshold value W_{th} (Maksimenko et al., 2017).*

To decide whether a SWD precursor is present in the EEG the algorithm, three criteria apply:

1. Energy of $W_{\Delta S1}(t)$ needs to exceed an individualized specific threshold.
2. Energy of $W_{\Delta S1}(t)$ must exceed energy of $W_{\Delta S2}(t)$.
3. Energy of $W_{\Delta S1}(t)$ must exceed energy of $W_{\Delta S3}(t)$.

Since SWD prediction quality depends on the above mentioned individualized thresholds, SWD prediction performance of each channel combination was determined for a total of 14 fixed threshold values ranging from 0.1 to 0.75 in case of three channel combinations, and a total of 16 fixed threshold values ranging from 0.005 to 0.04 in case of two channel combinations. Threshold range varied based on acquired product from either two or three channel.

3.2.2. Sensitivity, specificity and number of false positive

For each analyzed channel combination, the algorithm created a list of time points, at which a potential SWD precursor was detected. With the aid of this list, the number of true positive predictions, the number of false positive predictions as well as the number of missed (unpredicted) SWDs were determined.

In addition, sensitivity (sensitivity = number of correctly predicted SWDs / (number of correctly predicted SWDs + number of unpredicted SWDs) × 100%) as well as number false positives per hour were calculated for each channel combination and for each of the tested individualized specific threshold values.

3.2.3. Histology

After recordings, the animals were euthanized by an overdose of intraperitoneal injected pentobarbital (Narcoren, 100 mg/kg; Merial GmbH, Münster, Germany). A direct current (9 V, 25 µA, 2 s duration) was pathed though each electrode to create an electrolytic micro-lesion at the location of the electrode tip. The brain was removed carefully and fixed in 4% paraformaldehyde (PFA). After 24 hours, the brain was moved to a 30% sucrose solution for three days and then cut into coronal slices of 60 µm using a

microtome. Fixated brain-slices were stained with cresyl violet and inspected under a light-microscope (dnt, DigiMicro Profi, Germany).

3.2.4. Statistical analysis of SWD prediction performance based on the algorithm of Maksimenko et al. (2017)

In order to compare SWD prediction performance, the sensitivity and the number of false positives were visualized in histograms for each rat plotting channel combinations on the x-axis against fixed threshold values on the y-axis.

Linear regression analysis (Pearson correlation) was used to determine the degree of interdependence between the sensitivity of prediction as well as the number of false positives per hour.

An analysis of variance (ANOVA) was performed with sensitivity as dependent variable, channel combination as between subject factor 1, version of algorithm (2channels, 2ch; 3channels, 3ch) as between subject factor 2, threshold as covariate 1 and number false positives per hour as covariate 2.

Likewise, an ANOVA with number of false positives per hour as dependent variable, channel combination as between subject factor 1, version of algorithm (2ch, 3ch) as between subject factor 2, threshold as covariate 1 and sensitivity as covariate 2 was performed thereafter.

To avoid the multiple comparison problem for post-hoc analyses, all channel combinations were grouped into different categories including category 1: two intracortical recording sites (CC), category 2: one cortical and one thalamic recording site (CT), category 3: two intrathalamic recording sites (TT), category 4: three intracortical recording sites (CCC), category 5: two cortical recording sites and one thalamic (CCT), category 6: one cortical recording site and two thalamic recording sites (CTT) and category 7: three intrathalamic recording sites (TTT).

Post hoc analyses included:

1) ANOVA with sensitivity as dependent variable, category of channel combinations (CC, TC, TT, CCC, CTT, CCT, TTT, MCCC) as between subject factor 1, version of algorithm (2ch, 3ch) as between subject factor 2, threshold as covariate 1 and number of false positives per hour as covariate 2.

2) ANOVA with number of false positives per hour as dependent variable, category of channel combinations (CC, TC, TT, CCC, CTT, CCT, TTT, MCCC) as between subject factor 1, version of algorithm (2ch, 3ch) as between subject factor 2, threshold as covariate 1 and sensitivity as covariate 2.

All statistical analyses were performed using IBM SPSS version 25. Data are expressed as the arithmetic mean values ± standard error of the mean (S.E.M.). Differences were considered statistically significant when $p \leq 0.05$ (*), $p \leq 0.01$ (**) and $p \leq 0.001$ (***).

3.2.5. Comparison of wavelet spectra displaying true positive or false positive detections.

All true positive detections and a total number of 50 randomly selected false positive detections per rat, both identified by the wavelet-based algorithm of Maksimenko et al. (2017), were extracted from three cortical recordings within the somatosensory cortex of WAG/Rij rats and GAERS. Wavelet power within different frequencies were determined and statistically compared using repeated measures ANOVA with wavelet power as dependent variable, type of detection (true positive, false positive) as within subjects factor 1 and frequency bands $W_{\Delta S1,}$ $W_{\Delta S2 \text{ and}}$ $W_{\Delta S3}$ as between subjects factor 2 was used for statistical comparison.

3.3. Machine learning and deep learning techniques

The following section will focus on the technical background of machine learning and deep learning techniques applied in this thesis. Concurrently, the implementation of the deep learning techniques to the LFP data and the prediction of SWDs will be described. The general buildup of neural networks and its different conventional forms used in this thesis will be introduced including the techniques and concepts of convolutional neural networks, recurrent neural network and long short-term memory (LSTM) networks (Hochreiter & Schmidhuber, 1997).

3.3.1. Random forest

Simple decision trees are used in a broad context, for example in the algorithm of Maksimenko et al. (2017) that is used in this thesis. Here, a decision tree determines with a few parameters that need to be fulfilled the detection of the precursor. Once the performance of one single decision tree is not optimal, the machine learning paradigm Random forest might improve the accuracy. A random forest is an analogy of a forest that is made up by multiple trees. A single decision tree within a random forest algorithm on its own could possibly provide only mediocre results. However, as a group of trees, a random forest provides accurate decisions based on majority voting (Figure 9). In this research, a random forest with 1000 trees is utilized for further differentiation of features used in the wavelet based precursor detection, described in paragraph 3.2.1. The wavelet power of three EEG traces in three frequency bands is used for the classification, which results in a decision based on 9 alternating wavelet values. Two different dataset were measured with the random forest. The first dataset consisted of six GAERS in which the EEG was only measured within the somatosensory cortex. The second dataset comprised of six of the ten WAG/Rij rats that had matching electrode locations (appendix 2). Time points and wavelet features for true and false prediction were extracted at a threshold value resembling >60% sensitivity of given combination and rat resulting in small changes of the amount of predicted SWDs. These small changes are attributed to the fixed threshold value. Five combinations of channels were tested from EEG traces of WAG/Rij rats dataset: 1. Recordings in layer V and VI of the somatosensory cortex further referred

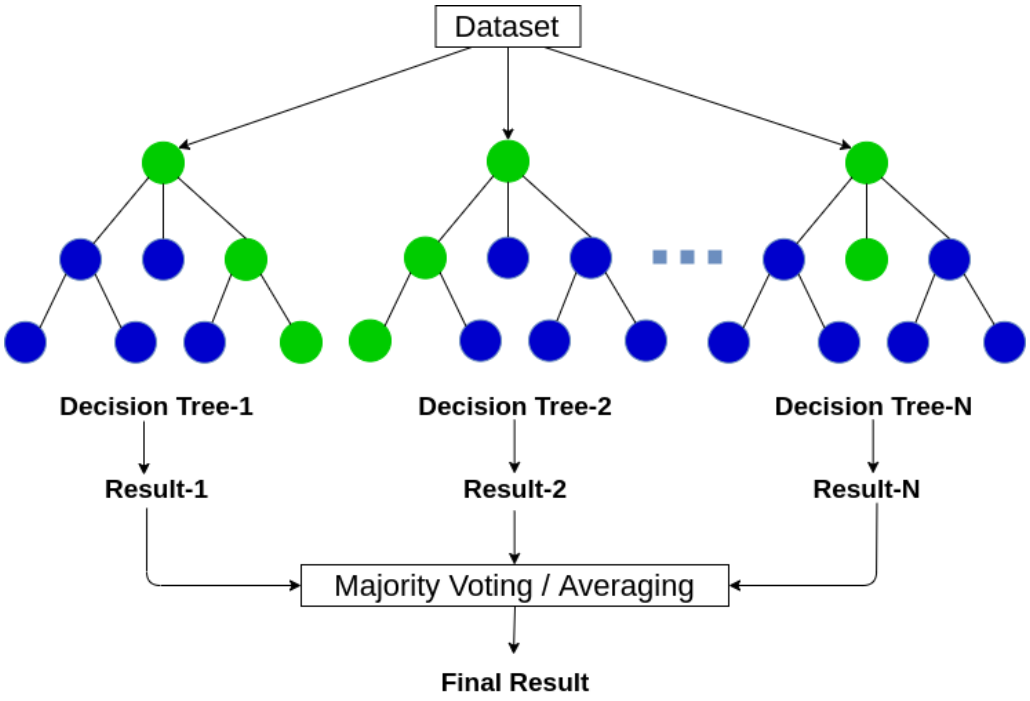


Figure 9: Random forest example. *Decision based on majority vote of multiple single decision tree results (taken from Sharma, 2020).*

to as "CC" (n=177); 2. Recording in layer IV, V and VI referred to as "CCC" (n=161); 3. Recordings obtained in layer IV, VI and VPM as "CCT" (n=161); 4. Recordings acquired in Layer VI, VPM and cRTN as "CTT" (n=145) and lastly 5. Recordings acquired in VPM, cRTN and Po as "TTT" (n=162). In addition to the WAG/Rij rat dataset, one combination of intra-cortical EEG measurements for the GAERS dataset from layer IV, V and VI (GCCC) of the somatosensory cortex was tested. Lastly, to get an impression on the effect of sample size, intra-cortical GAERS data containing either 1844, 161 and 145 true and false detections were compared to each other (GCCC 1844, GCCC 161 and GCCC 145).

In this thesis, the technique called "random-undersampling" was applied to reduce the training time of the model with the least amount of data loss and counter the data imbalance. When applying random-undersampling, all events of the underrepresented class are selected but only equal amounts of the overrepresented class. This results in a dataset consisting of 50:50 preictal time intervals and interictal events. To visualize the range of the differentiation in each combination, each random forest (n=100) was fed with all available data samples of the true positive class, and different sets of randomly selected false positive detections. The balanced accuracy was calculated for each run and statistically compared between groups (CCC, CCT, CTT, TTT and CC) using an ANOVA. In

addition to that, the accuracy of one tree revealing the average (median) accuracy was statistically tested against chance level using permutation testing. To this end, true and false detections were randomly shuffled to one of the two classification classes and for each randomization the maximal accuracy was determined and displayed in a histogram. In case the accuracy computed from the real dataset was positioned above the 95th quantile of the histogram, the real dataset deviated (with a certainty of 95%) from randomized datasets, which did not have a difference between true positives and false positive detections.

3.3.2. Raw EEG data analysis

EEG contains a multitude of information corresponding to brain activity. Here, the information from the precursor of a SWD can already be found prior to seizure onset (Ovchinnikov *et al.*, 2010; Lüttjohann & van Luijtelaar, 2012). To further improve the detection of SWD precursor, a deep learning approach was applied specifically designed for the analysis of rapidly fluctuating sequential data-like EEG. More precisely, a modified LeCun network, an artificial neuronal network that was previously used in image recognition (Figure 10) as well as the detection of epileptic seizures (LeCun, 1998; Mirowski *et al.*, 2008), was applied for the sorting of interictal and preictal EEG fragments.

This deep learning neural network, adapted from LeCun's LeNet-5 (Figure 17), contains multiple convolutional layers, which are responsible for extraction of the internal structure within EEG signals. Two MaxPooling (retaining only maximum value in given window) layers reduced the amount of parameters from the data, which counters an overfitting of the model and increases the processing speed. Afterwards, features learned and extracted by the convolutional layer are presented to a layer of long short-term memory (LSTM). LSTM as recurrent neural networks retain information to a higher degree than other deep learning algorithms (e.g. like a support vector machine) and thereby are better suited to investigate the temporal dependencies within the EEG signal. The output of the LSTM layer is then fed into several fully connected (dense) layers that are used to further analyze the data. Lastly, a softmax activation layer sorts the output into the two classes: "preictal" and "interictal". Dropout layers that nullify changing

neuronal output over each training are implemented between each layer to counter overfitting. Before describing this highly specialized network in more detail (section 3.3.4.2), a general background on the concepts of ANN and deep learning including the technical background of the previous mentioned subparts (e.g. layers) of the applied LeCun network are introduced next.

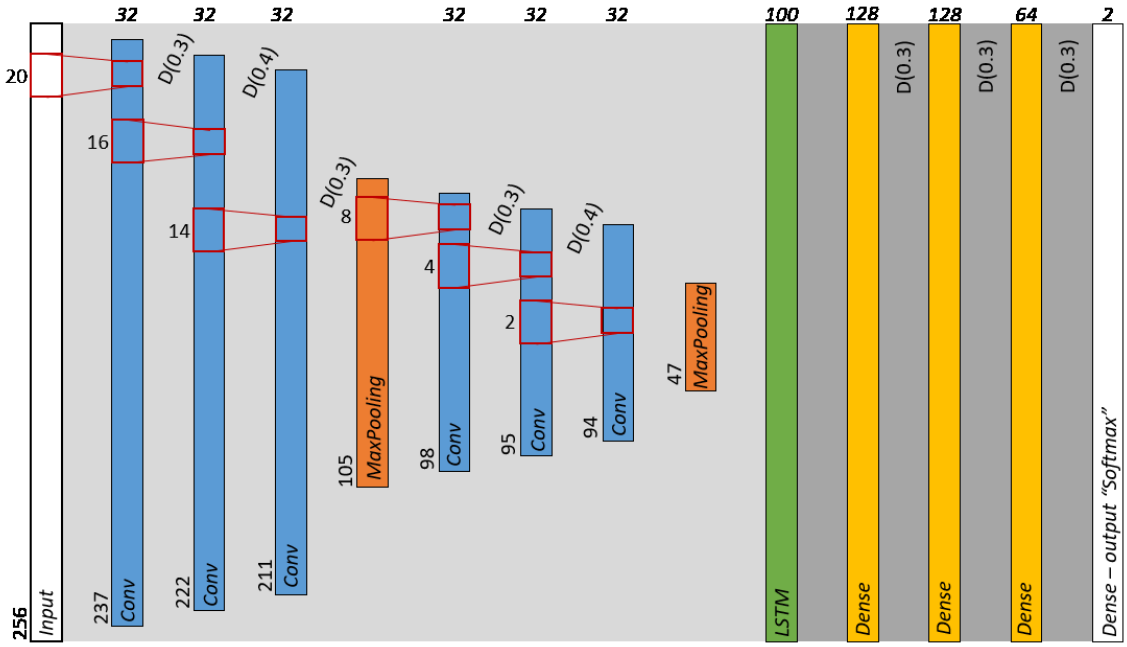


Figure 10: Modified LeCun network with implemented long short-term memory (LSTM). *Far left showing the input layer (white) with 256x3 data points, corresponding to 512 ms and the three EEG traces. Numbers over the layer name display the change in data points remaining. Red boxes show the shifting kernel size with each one stride over the data. Blue Layers correspond to Convolution 1 dimensional layer with each 32 neurons per layer. MaxPooling layer is depicted in orange, which reduces the data by half (shifting kernel of 2). Long short-term layer (100 neurons) shown in green for recurrent processing. Three dense layer (128, 128 and 64 neurons) displayed in yellow. Far right shows the output layer with two neurons corresponding to the two available classes. The output layer applies the "softmax" activation function. Convolution and Dense layer each use the activation function "ReLU". Between each layer, a Dropout layer (D) is implemented that counter overfitting by setting the output of randomly selected neurons (30-40%) to 0.*

3.3.3. Artificial neural network

Artificial neural networks or short neural networks are processing systems inspired by the biological brain (Hopfield, 1988). ANN are built of interconnected nodes, generally called neurons, as stated earlier. The neurons are situated in layers where the input $\mathbf{x}_{(t)}$ is modulated by a matrix out of weights w_x and biases b and then passed through some

activation functions Ø. Generally, ANN are built up by stacking several layers after each other (Figure 11).

Each neuron within the layer has connections to one or multiple neurons in the following layer. The connections between neurons are weighted and should mimic synapses of the biological system. The most important part of an ANN is the ability to learn. The learning process of ANN is comparable to the brain's learning in a way that it can be traced back to the interconnections between neurons (Haykin, 2009). ANN are built of one input layer, at least one hidden layer and one output layer. Each layer receives inputs from the previous layer. The nodes of the input layer have the dimensionality of the given data and are the first nodes that process the data. Between the input and the output layer is the hidden layer, which is the key computing unit responsible for the processing and final decision. The output layer of an ANN is used for the classification tasks and is typically made up of only a few nodes, in which each node represents a different decision. The calculated value in the output neuron corresponds to the probability of the input *x* belonging to the class *y* (Nielsen, 2019).

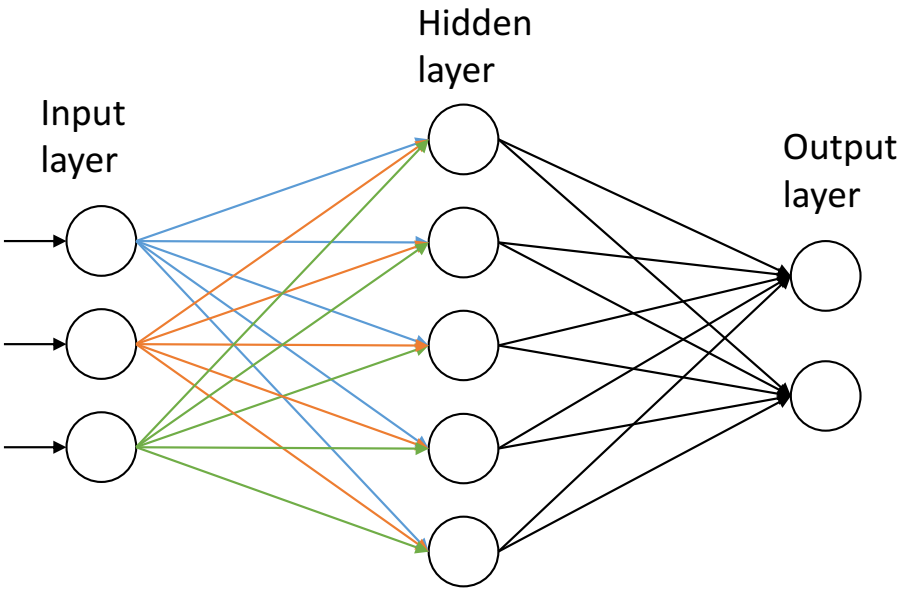


Figure 11: Schematic structure of a shallow artificial neural network. *(left to right) The input layer with three neurons is connected to all neurons of the hidden layer. The output layer receives inputs from the last hidden layer. (adapted from Castrounis, 2019)*

At the beginning of each training session, the weights and the bias are set to initial values. The weights are set randomly in the range between values of 0 and 1 to break the symmetry. The bias on the other hand are initialized as a value close to zero to increase the dimensionality. Weights and biases are learnable and adjustable parameters corresponding to the learning of a neural network and are responsible for the signal furthered through the layers.

An important feature of supervised machine learning, as applied in this thesis, is the cost function also known as loss function or error function. It represents the average difference between the predicted output and the desired output. The overall goal of an ANN is the reduction of cost function to zero for the training and the validation set, which would indicate that every feature of the training set is learned. The cost function represents miscalibration of the weights to separate the given classes. This allows the ANN to apply optimization methods for weight adjustments. Tracing back the influence of certain weights on the prediction is done via backpropagation.

Backpropagation, also applied in this thesis, is an often-used optimization method for ANN. It is used to compute the contribution of each neuron/weight to the predicted output. Backpropagation starts by calculating the error of the last hidden layer to the output layer. By transitioning through the model layer by layer and changing the weights correspondingly to its average relation to the output, the ANN tries to optimize itself for the given classification task. Backpropagation algorithms use a gradient descent to optimize the weight configuration based on the cost function (Figure 12). Each iteration tries to reduce the loss to converge to a minimum.

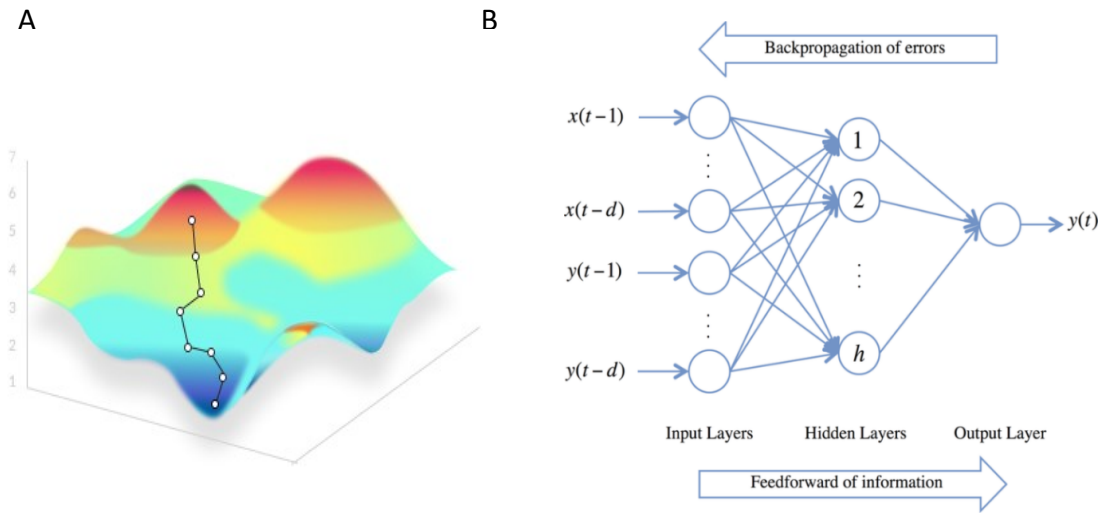


Figure 12: Backpropagation in neural networks. *(A) Gradient descent of a function is illustrated by a stepwise (white dots) nearing data minim (dark blue). Data values are displayed from high to low by a color code from red to blue. (B) Backpropagation of errors from the output layer back onto the weights of previous layer (Missinglink.ai, 2020, Chan Phooi M'ng & Mehralizadeh, 2016).*

3.3.3.1. Activation functions

Activation functions are a necessary tool to modulate neuron output. Several activation functions are used to modulate and increase the dimensionality of the data helping to

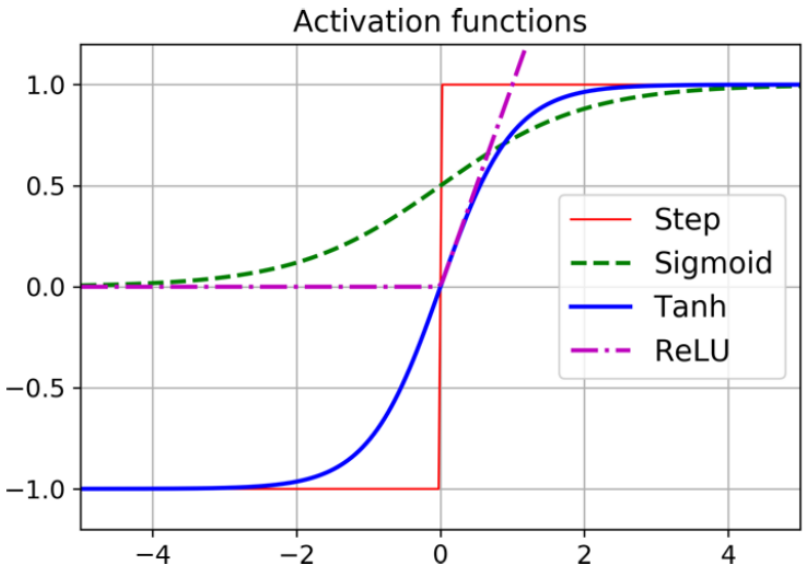


Figure 13: Most common activation functions. *The three activation functions rectified linear unit (ReLU), Tangens hyperbolicus (Tanh) and Sigmoid are shown in comparison to a simple step. (Missinglink.ai, 2020).*

learn complex patterns. Activation functions also limit the output to a certain range. The most commonly used activation functions also used in this thesis are Sigmoid, Tangens hyperbolicus (Tanh) and rectified linear unit (ReLU) (Figure 13).

Each activation function alters the linear data and thereby increases the dimensionality using the following algorithms.

- Sigmoid function

$$Sigmoid(z) = 1/(1 + \exp(-z))$$

- Tanh function

$$Tanh(z) = 2\sigma(2z) - 1$$

- ReLU function

$$ReLU(z) = \max(0, z)$$

The ReLU function is a default function for most feedforward neural networks. ReLU changes linear output transformation to nonlinear transformation. This activation function remains close to linear so that optimizations devised for linear improvements and gradient-based modulations remain possible (Goodfellow *et al.*, 2011). In this thesis, the ReLU activation function is used in the convolutional and dense layer of the modified LeCun network (see Figure 10 and LeCun et al., 1998) as it reduces the activation of neurons if the input value is negative.

The output layer of the modified LeCun network on the other hand uses the so called softmax activation function, which applies the metric categorical cross-entropy. Categorical cross-entropy is specifically used, if only one class can be present simultaneously, e.g. EEG can only represent either interictal or ictal not both at the same time. The softmax activation function transforms the output into probabilities between 0 and 1. Each value corresponds to a previously defined class with the sum adding up to 1.

3.3.4. Recurrent neural network

The modified LeCun network is a so called recurrent neuronal network (RNN). Sequential data such as natural language processing, weather prediction, speech recognition and in this case EEG data processing require this type of ANN, because RNN are specialized to work with varying lengths of sequential data. Its internal state correlates to a shared weight across several time steps, which resembles a form of feedback connection within layers. By contrast, convolutional neural networks (CNN) with their convolution operation have the same capacity to share information over time, however they remain shallow on their ability to learn from long sequences (Sainath *et al.*, 2015). ANN can work with one-to-one data structures where the algorithm works with a fixed size input and gives out a classification. In this thesis, the approach many-to-one is applied. Many-to-one receives a sequence of inputs and produces a classification out of it.

There are three types of mappings in sequential data:

- sequence to sequence
    - *Time series is forecasted of stock values or weather data*
    - *Sequences of data like time series data predicts multiple time points after the end of the training data*
- sequence to vector
    - *Sequence data is analyzed and converted into a specific state e.g. Seizure/Baseline, vigilant, sleeping*
- vector to sequence
    - *Images are converted and described into a specified image description*

In this project, sequential EEG data are analyzed with a sequence to vector mapping, where the output is a prediction of a chance in which class the next time interval is situated. RNN consist of similar core computing units within the layers as seen in normal ANN. This core unit has an internal state that will be updated every time a new input is presented to the cell. The internal state will be fed back to the model the next time it reads an input, thereby creating a memory function to a certain extend.

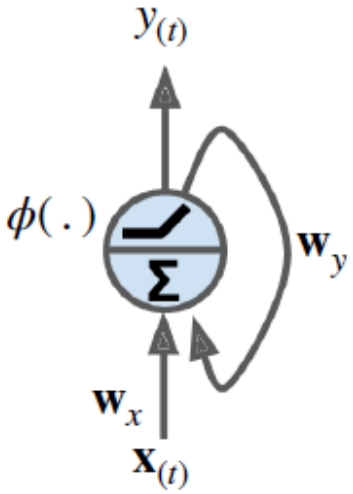


***Figure 14: Core node of Recurrent Neural Networks**. X resembles the input at a given time (t) with the current weights set for this input (W_x). As a core mechanism, the inputs from all connections are summed up (∑) and processed by an activation function (φ) and the internal weights (W_y). The core cell has two outputs representing the input to the next layer (γ (t)) and the alternated internal weight (Missinglink.ai, 2020).*

The common procedure of RNN units uses the ANN algorithm with the addition of reapplying another internal weight to the data that is retained during training to a certain degree, thus creating the memory function (Figure 14) (Missinglink.ai, 2020).

3.3.4.1. Problems encountered in RNN: Vanishing and exploding gradient

RNN can encounter a vanishing or exploding gradient with gradient-based learning once the derivatives of activation functions attain close to zero and large values, respectively. Gradient descent becomes less effective in big neural networks with many layers. Additionally, the noise on learning increases alongside with the length of temporal dependencies (Bengio *et al.*, 1994; Goodfellow *et al.*, 2011). Both problems inhibit the network's ability to change weights by backpropagation due to long-term dependencies. Once the gradient is close to zero, weights are unable to adjust and change in a significant

way to achieve learning. However, if the gradient is big, changes in weight increase too much so that the learning ability is still impeded.

3.3.4.2. Convolutional neural network

As displayed in Figure 10, the modified LeCun network contains a number of different types of layers including convolutional layers, a LSTM layer and fully connected (or dense) layers. Temporal convolutional neural networks (CNN) as a subcategory of Deep learning algorithms are designed to work best with a grid-like topology of the data. CNN can assign importance to aspects found in the data. They are used commonly in image recognition due to its 2-D grid like structure (Figure 15) (Geitgey, 2016). Similar to image recognition, EEG data can be viewed as a 2-D grid like structure as well. CNN can break down EEG pattern to learn, which temporal frequency changes are present in the data and adapt to it (Wang *et al.*, 2017).

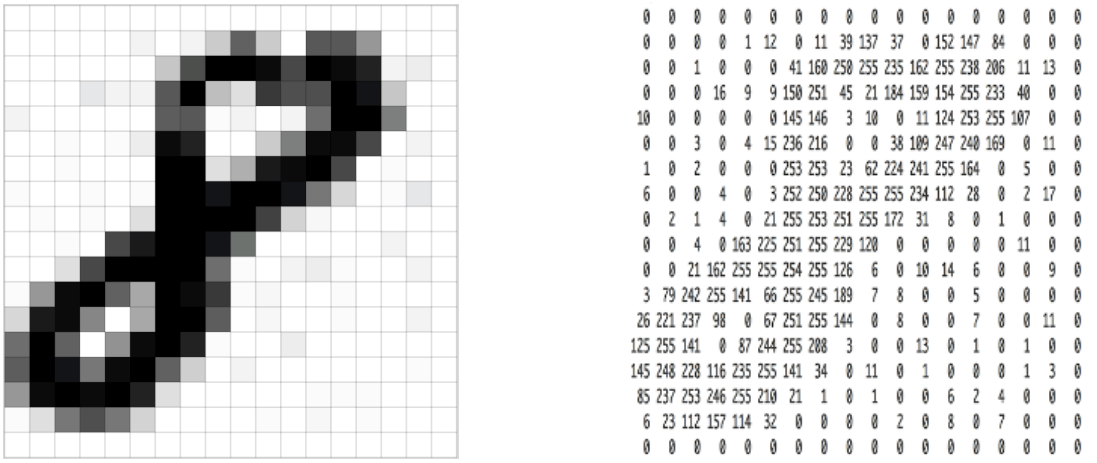


Figure 15: Transfer of a picture to numerical display. *(Left) Pixel matrix of a greyscale image depicting the handwritten digit eight. (Right) Pixel saturation indicated by a range from 0 to 255. Handwritten digits provided MNIST data set (Geitgey, 2016).*

CNN use the linear mathematical operation called convolution, which is used instead of matrix multiplications on the layers (Lecun *et al.*, 1998). The result of convolution operation is termed feature map. In comparison to old methods where hand-crafted features/filters were applied to images, CCN have the ability to define learned features themselves. Each layer has the ability to extract several features depending on the size of the applied window and stride of the shifting window. This results in locally learned features, which can be globally traced back. In this way, a feature like a precursor can be traced back irrelevant of the position in the EEG signal. The irrelevance of signal

positioning like an edge detection is achieved by a sliding window. This window convolves over the data with a set of filters like width and height in image recognition. The filter runs over the entire input data so that in images all color channels are covered. This results in a two dimensional activation filter at each spatial location. Interestingly, CNN were designed after the organization structure of the visual cortex, where neurons in restricted areas of the visual field respond to a given stimulus (Fu *et al.*, 2016). Alterations and optimizations are done by the size of the kernel (i.e. the analysis window) going over the grid like data. Correlations of features can be enhanced by a small sliding window

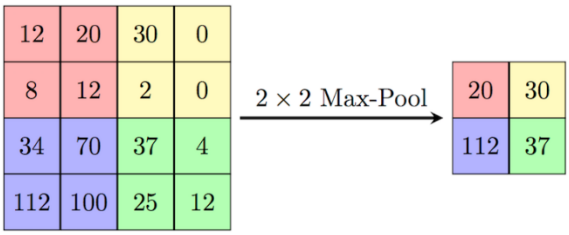


Figure 16: Max-pooling. *Processing a dataset of 16 data point with a 2x2 kernel that keeps only the maximum values of each kernel. (Geitgey, 2016).*

(receptive field on the temporal EEG data) increasing the spatial shift. Further, a technique called polling is applied to reduce the information of the raw data. Pooling uses a shifting window that process the data in some ways e.g. retaining only the highest value, thereby increasing the processing speed and reducing the noise (Figure 16, Figure 17) (Goodfellow *et al.*, 2011; Schmidhuber, 2015; Geitgey, 2016).

CNN are not only built of convolutional layers. Convolutional layers are paired with several mathematical optimizations such as ReLU activations, pooling layers (Max,

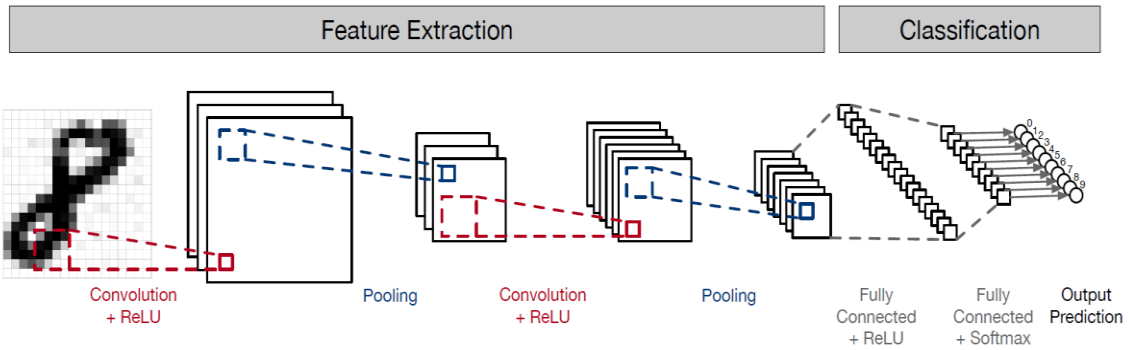


Figure 17: LeNet-5 CNN architecture*. The handwritten digit (left) is processed first by feature extraction moving different sized kernels over the image and later processed by a pooling operation. Features are learned during classification, which then gives a prediction on the digit. (Geitgey, 2016).*

Average, Global), fully connected layer (dense layer), dropout layer, batch normalization layer and finally the output layer. With a wisely chosen combination of these compartments (Figure 10, Figure 17), CNN are a more efficient tool to process image data and sequence data as CNN display better results for classification accuracies that surpass human based precisions.

In this thesis, multiple convolutional layers are responsible for extraction of EEG frequency features. EEG features are extracted based on different time intervals with a decreasing kernel size from the stating layer to the last convolution layer. This change in kernel size enables the focus of extracting specific temporal frequency dependencies at high-level layers and the extraction of small temporal dependencies at lower levels. Each layer applies the ReLU activation function with a dropout layer (30-40%) in between each convolutional layer to reduce overfitting. Additionally, two MaxPooling steps are implemented to reduce the data size while retaining most of the information and increasing the processing speed (see Figure 10).

3.3.4.3. Long short-term memory

The convolutional layers produce a feature map of the EEG data, however temporal dependencies between these features need to be processed with a memory retaining RNN. Normal RNN (see paragraph 3.3.4) retain information only to a certain degree and experience gradient based problem. A solution to this problem was the introduction of LSTM layers, which are also part of the modified LeCun network used in this thesis.

Long sequence data such as EEG data require a certain RNN model depth to extract features accurately. In addition, RNN are not efficient in extracting long-term temporal dependencies. Models with a high number of hidden layers can suffer a problem by backpropagation. To address this problem, Hochreiter & Schmidhuber proposed the LSTM network in 1997. In this network, simple neurons are replaced with intricate computing cells that capture long-term dependencies while not diminishing short-term effects.

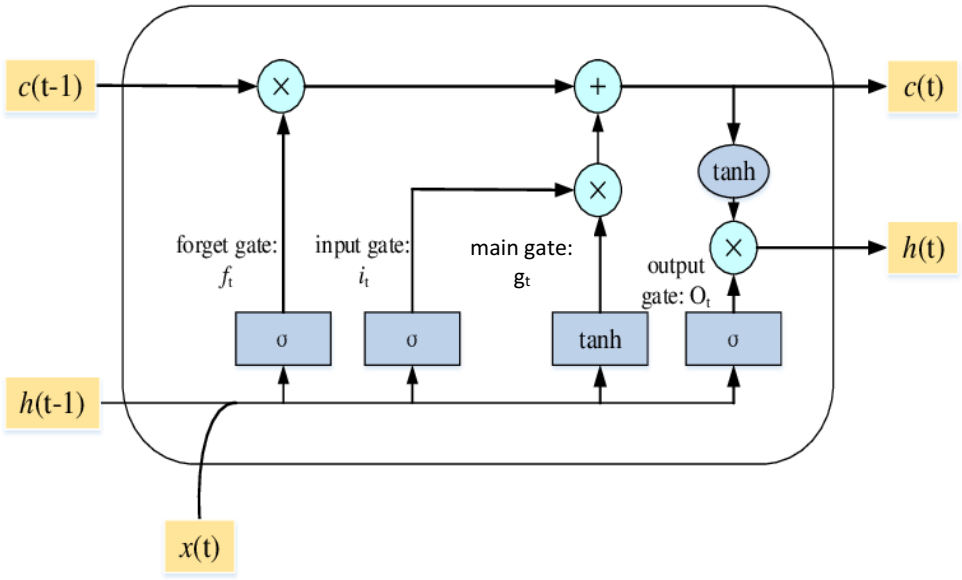


Figure 18: Schematic display of a LSTM cell. *(c(t-1)) represents the input from the previous LSTM memory function. (h(t-1)) represents the previous LSTM output, (c(t)) current LSTM memory, (h(t)) current LSTM output. The two activation functions (Tanh and sigmoid) are displayed in grey. Point wise operations are displayed in blue. (modified from Yuan et al., 2019).*

LSTM networks can retain information to a certain degree over longer training sets and have the ability to forget certain information thereby countering the vanishing or exploding gradient. They have a certain range of outputs that present data to the model and to further layers. These functions are called hidden state/short-term output $h_{(t)}$, long-term memory state $c_{(t)}$ and output for the current time step $y_{(t)}$. To achieve these specific outputs, LSTM are built of four main components: an input gate, a forget gate, an output gate and a main gate (Figure 18). Within a single LSTM cell two activation functions, Tanh and ReLU are used for specific alterations of the data.

- g_t main gate: keeps track of inputs
- f_t forget gate: controls parts of c(t) output
- i_t input gate: controls, which parts should be added from g(t) to c(t)
- O_t output gate: controls, which parts of c(t) should be included in y(t) and h(t)

3.3.5. Training of ANN: Validation of a dataset

Supervised machine learning algorithms need a similar unseen dataset to test their real accuracy. For this, the whole dataset is divided in a training and a test set (Figure 19A) (Scikit-learn, 2020).

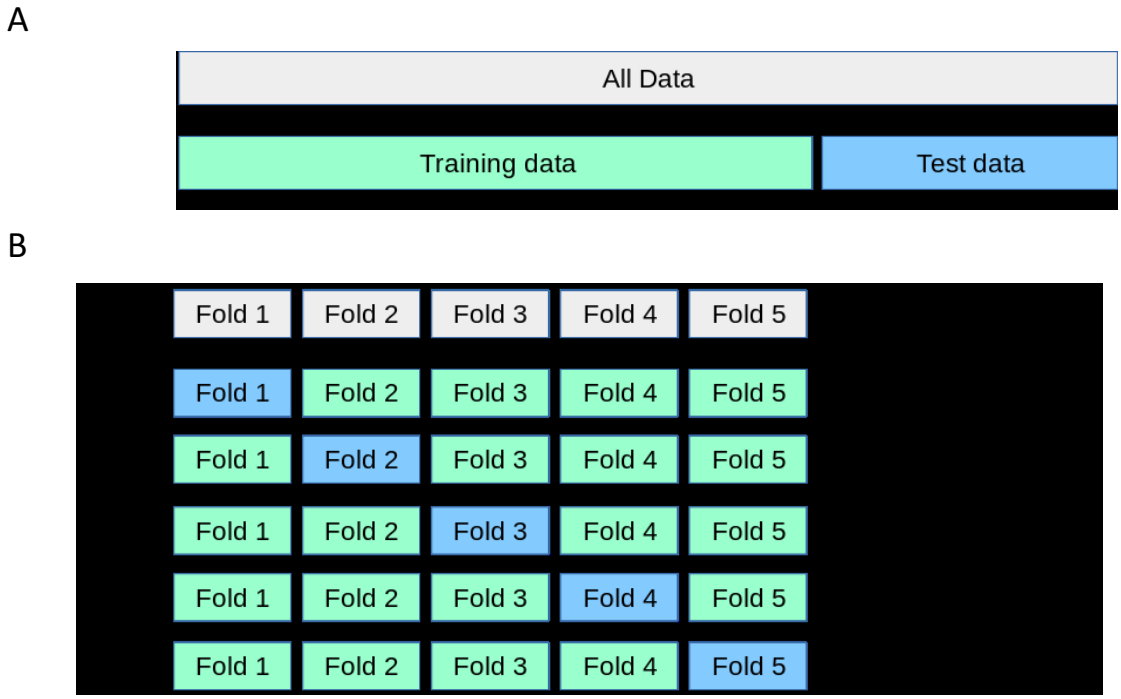


Figure 19: Data processing for validation. *(A) The whole dataset is divided into a training set and a test set. (B) K-fold cross validation. Data is split into 5 different sets with an always shifting test data* (Scikit-learn, 2020).

Data sets may not always represent the real world scenario due to missing information or unbalanced representation of a class. In our case for example, there are much more interictal periods available compared to precursor events. Cross-validation is a technique to ensure equal distribution. To achieve an accurate representation of all classes, several validation techniques are used. One of these techniques is called K-Folds cross-validation, where the training data is further randomly split in training data and test data (Figure 19B).

In this thesis, a 10-fold split is used splitting the 144 hours of LFP data derived from six rats into 10 equal parts. Each split shifts the data preserved for testing to ensure that each data set was used as validation set. At the end the parameters of each split is averaged and tested on the remaining unseen test data set.

3.3.6. Preprocessing of the dataset fed to the modified LeCun network (Deep learning dataset)

For the deep learning part of this thesis, it is imperative to preselect the data due to an unbalanced representation of classes. The first analysis considered WAG/Rij rat data. The same animals were selected for the analysis with the first deep learning algorithm focusing on on-line features as well as for the second algorithm working with unprocessed EEG data. The two prominent classes in this thesis are baseline and precursor, the precursor being highly underrepresented. In the second analysis, somatosensory EEG data was used from six GAERS, which were measured for 24 hours with a sampling frequency of 500 Hz resulting in 43.200.000 time points per animal (24 h x 60 min x 60 s x 500 Hz = 43.200.000). To address this problem, random undersampling was applied to reduce the interictal event to equal values of preictal event. The dataset was manually labeled in the WinDaq Playback software prior to the deep learning analysis to define the start and end of SWD seizures.

The raw data was split into bins of 512 ms with an analysis window of 256 ms directly afterwards (Figure 20). The bins were shifted over the whole dataset by the size of the analysis window, which created an overlap by 50% of the raw data. The analysis windows was labeled to display the state of the animal with: “0” corresponding to baseline events, “1” transition time interval from baseline to SWDs and “2” corresponding to SWD intervals. For the purpose of the detection of a precursor prior to a SWD, the Deep learning model was trained to detect only baseline and transition intervals. To account for rat specific EEG measurements, raw EEG data was normalized. Shifting window of the raw EEG traces corresponded to 512 ms with a shift and analysis window (directly connected) of 256 ms. Data size was chosen based on exponents of two to allow faster processing.

This intricate preprocessing and training of the deep neural network was additionally illustrated by video presentation (Appendix 3)

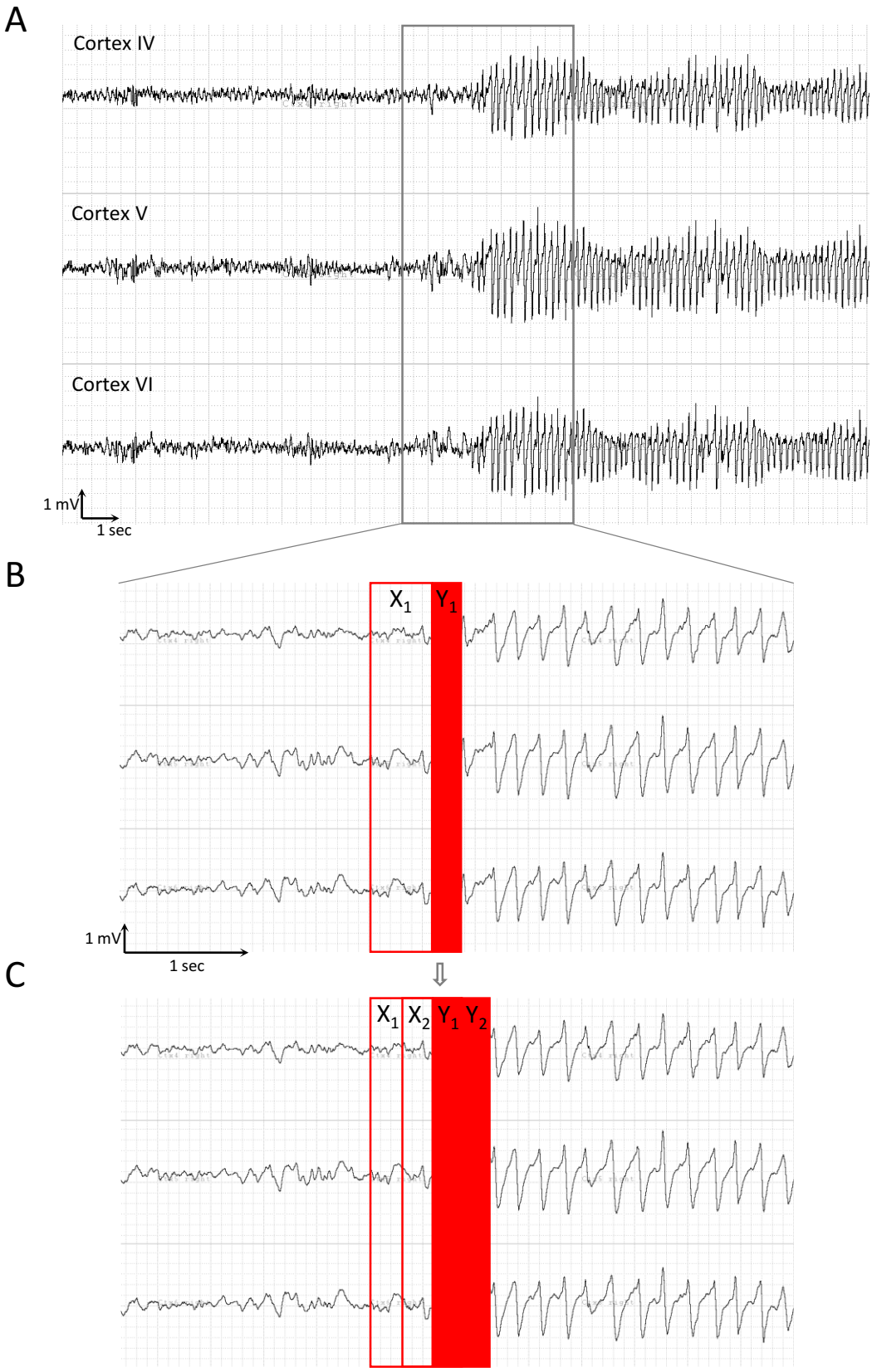


Figure 20: EEG deep learning preprocessing. *(A) Raw EEG from GAERS animal recorded within cortex IV, V, VI. (B) Extracted raw data at position 1 (X) with bin length of 512 ms and an analysis window (Y) of 256 ms. (C) Moving window through the entire dataset at position 2 with an overlap of raw EEG data equal to the analysis window. Value of the analysis window ("0" for interictal and "1" for preictal) corresponding the state of the animal within set time.*

3.3.7. Hyperparameter configuration

The deep learning model was trained with the optimizer "Adam" designed to work with stochastic optimization (Kingma & Ba, 2015). Adam has pre-set hyperparameter settings with a learning rate of 0.001. This optimizer is chosen to reduce overfitting of the model while keeping the pattern learning at a reasonable level. The batch size was set to 64 corresponding to the presented data within one training run. The model was trained with each data set split for 75 training epochs (Figure 19).

The deep learning algorithm was built (programmed) and run (performed) with Keras version 2.3, Tensorflow version 2.2 and Python 2.6.

3.3.8. Performance evaluation deep learning algorithm

The evaluation of the 10-fold cross validation is done by averaging the balanced accuracy of the validation data set from each trial. Balanced accuracy is calculated by adding the sensitivity (True positives/the total number of seizure sample) and the specificity (True negative/the total number of interictal sample) and then dividing it by two. The loss function and validation loss (val-loss) representing the learning of the features with 0 equals no information loss is checked for the steady decrease in both val-loss and loss. Decreasing val-loss represents accurate detected features in novel data sets. The data is then permutated as the data in the Random forest to test SWD prediction of the modified LeCun network against chance level prediction. A confusion matrix is further extracted to investigate the actual number of predicted interictal and preictal events (Figure 21).

	Predicted as interictal fragment	**Predicted as pre-ictal fragment**
Actual interictal fragment	Correctly classified as interictal	Incorrectly classified as interictal
Actual pre-ictal fragment	Incorrectly classified as pre-ictal	Correctly classified as pre-ictal

Figure 21: Confusion matrix. *Sum of all windows resembles data set size. Predicted positive and negative are correlated to the actual positive and negative present within the data set.*

Precision (correctly classified as preictal/(correctly classified as preictal + incorrectly classified as preictal) and predictive negative value (correctly classified as interictal/(correctly classified as interictal + incorrectly classified as interictal) is calculated with the help of the confusion matrix.

4. Results

4.1. Cortical and thalamic relationships in precursor detection

In a previous study, SWD prediction in WAG/Rij rats reached high sensitivities, however, produced a high degree of false positive predictions (Maksimenko *et al.*, 2017). As the generation of a SWD is thought to rely on an intricate interaction between cortex and thalamus. It was postulated that the brain areas used for the prediction of SWDs in that study might not yet have been chosen optimal. To elucidate, if there is an optimal channel combination that is suited to improve the prediction of SWDs, this work combined EEG recordings acquired from eight different brain regions of the thalamo-cortical system of WAG/Rij rats and compared those combinations on their SWD prediction performance. This comparison resulted in alternating sensitivities of SWD prediction (Figure 22) [$F(84, 10869) = 6.093$, $p<0.001$] and different rates of false positive (Figure 23) [$F(84, 10869) = 22.411$, $p<0.001$].

Taking a closer look at the specific sensitivities that were achieved by the different channel combinations, it was revealed that both the number of EEG traces used for prediction (i.e. two (22.8 ± 0.5%) vs. three channels (41.0 ± 0.4%)) [$F(1, 10946) = 644.348$, $p<0.001$] as well as the type of combination (CC (33.0 ± 0.9%), CT (24.2 ± 0.6%), TT (13.7 ± 0.7%), CCC (61.7 ± 1.5%), CCT (48.9 ± 0.5%), CTT (36.7 ± 0.4%), TTT (32.2 ± 0.8%), MCCC (33.3 ± 2.5%)) mattered for the degree of sensitivity that could be reached for SWD prediction (Figure 22A) [$F(7, 10946) = 250.6$, $p<0.001$]. Intracortical measurements within S1 brain region showed significant differences between the prediction based on two channel CC (33.04 ± 0.9%) and its three channel counterpart CCC (61.7 ± 1.5%) ($p<0.001$) (Figure 22B,C). In general, two channel prediction showed lower sensitivities in comparison to three channel prediction ($p<0.001$). However, intracortical prediction based on two channel CC did not vary significantly from three channel of intrathalamic recording sites TTT (32.2 ± 0.8%) and three motorcortex recording sites MCCC (33.3 ± 2.5%; $p=0.918$). Thalamo-cortical CT prediction showed significantly lower sensitivities to intrathalamic measurements (24.2 ± 0.6%; $p<0.001$). Lowest sensitivities were achieved by intrathalamic measurements TT (13.7 ± 0.8%) which were significantly lower to all other groups (all, $p<0.001$) (Figure 22).

As for the predictions based on two recording traces, differences in SWD prediction sensitivities was seen between intrathalamic, intracortical and thalamo-cortical (Figure 22). Prediction based on three channels showed similar behavior as the prediction based on two channels. Highest sensitivity was achieved by intracortical measurements within layer IV, V and VI of the somatosensory cortex (61.7 ± 1.5%). CCC showed significant differences to combinations using thalamic regions for the prediction (Figure 22C) (all, $p<0.001$). Significant differences were additionally seen between sensitivities reached with either one thalamic channel used (CCT: 48.9 ± 0.5%; $p<0.001$) or two (CTT: 36.7 ± 0.4%; $p<0.001$). No significant difference was measured between CTT and SWD onset zone far region MCCC ($p=0.179$). However, sensitivities reached with intracortical motorcortex combinations (MCCC) were significantly lower in comparison to intracortical somatosensory combinations (CCC) ($p<0.001$). CCT (48.9 ± 0.5%) showed significantly higher sensitivities in comparison to intrathalamic (TTT) combinations (32.2 ± 0.8%, $p<0.001$). The motorcortex (MCCC) combination did not differ significantly from intrathalamic measurements ($p=0.68$).

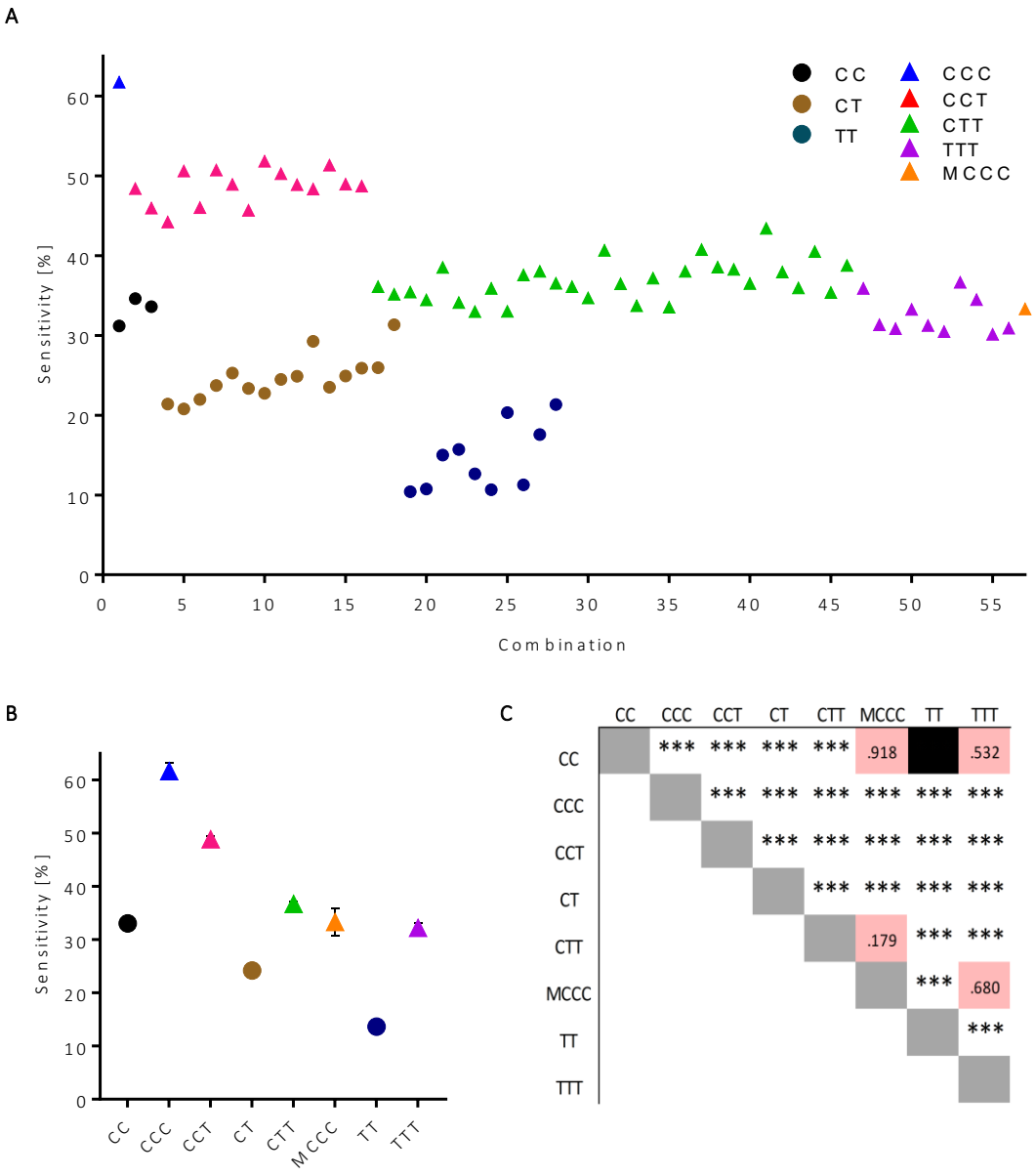


Figure 22: Average sensitivities of precursor detection achieved with different channel combinations. *(A) Sensitivities [%] of SWD prediction (displayed on y-axis) achieved for different combinations of brain regions displayed on the x-axis (see Appendix 1a, 1b or bookmark for numerical indexing of the combinations). Triangles (▲) correspond to channel combinations with three electrodes while round dots (●) refer to two electrodes used. Covariate factor threshold was calculated for 0.2719 and for 116.9715 nFP/h. (B) Average sensitivities [%] achieved for group-types with 1: two intracortical recording sites (CC), category 2: one cortical and one thalamic recording site (CT), category 3: two intrathalamic recording sites (TT), category 4: three intracortical recording sites (CCC), category 5: two cortical recording site and one thalamic (CCT), category 6: one cortical recording site and two thalamic recording sites (CTT) and category 7: three intrathalamic recording sites (TTT). (C) Results of statistical comparison between groups (*** indicates significance at p<0.001).*

Taking a closer look at the number of false positive per hour (nFP/h) the number of EEG traces again reveals to be of significant importance [$F(1, 10869) = 187.683$, $p<0.001$] (Figure 23A). Predictions based on two EEG traces showed significantly higher nFP/h (162.6 ± 3.3) than predictions based on three EEG traces (88.5 ± 2.3; $p<0.001$). The type of combination (intracortical, thalamo-cortical or intrathalamic) displayed significant difference in achieved nFP/h as well [$F(1, 10869) = 6.093$; $p<0.001$]). In contrast to the results seen for the achieved sensitivity above, intracortical CC combinations showed highest nFP/h (221.1 ± 6.2, $p<0.001$) (Figure 23B,C). Intrathalamic TT (164.9 ± 5.1) displayed the second highest false positive rate in comparison to CC (221.1 ± 6.2, $p<0.001$) and CT (144.7 ± 3.8, $p<0.001$). Interestingly, thalamo-cortical CT showed significantly reduced nFP/h to intrathalamic measurements (164.9 ± 5.1, $p<0.001$) and thereby produced the least errors of the two EEG trace version (Figure 23B,C).

Considering predictions based on three EEG trace, MCCC (129.8 ± 17.9) was the only group that produced similar nFP/h as the two channel version (CT) (144.7 ± 3.8, $p=0.396$). Additionally, MCCC showed no significant difference to prediction based on three intrathalamic channel (TTT) ($p=0.270$). Intracortical CCC (85.2 ± 10.6) nFP/h did not differ significantly to groups with thalamo-cortical groups including either one thalamic EEG trace (CCT 70.6 ± 3.5, $p=0.181$) or two thalamic EEG traces (94.7 ± 2.9, $p=0.181$). However, CCC produced significantly less nFP/h than TTT (110.2 ± 5.4, $p<0.001$). CCT displayed a significant reduction of false positives in comparison to CTT (94.7 ± 3.0, $p<0.001$).

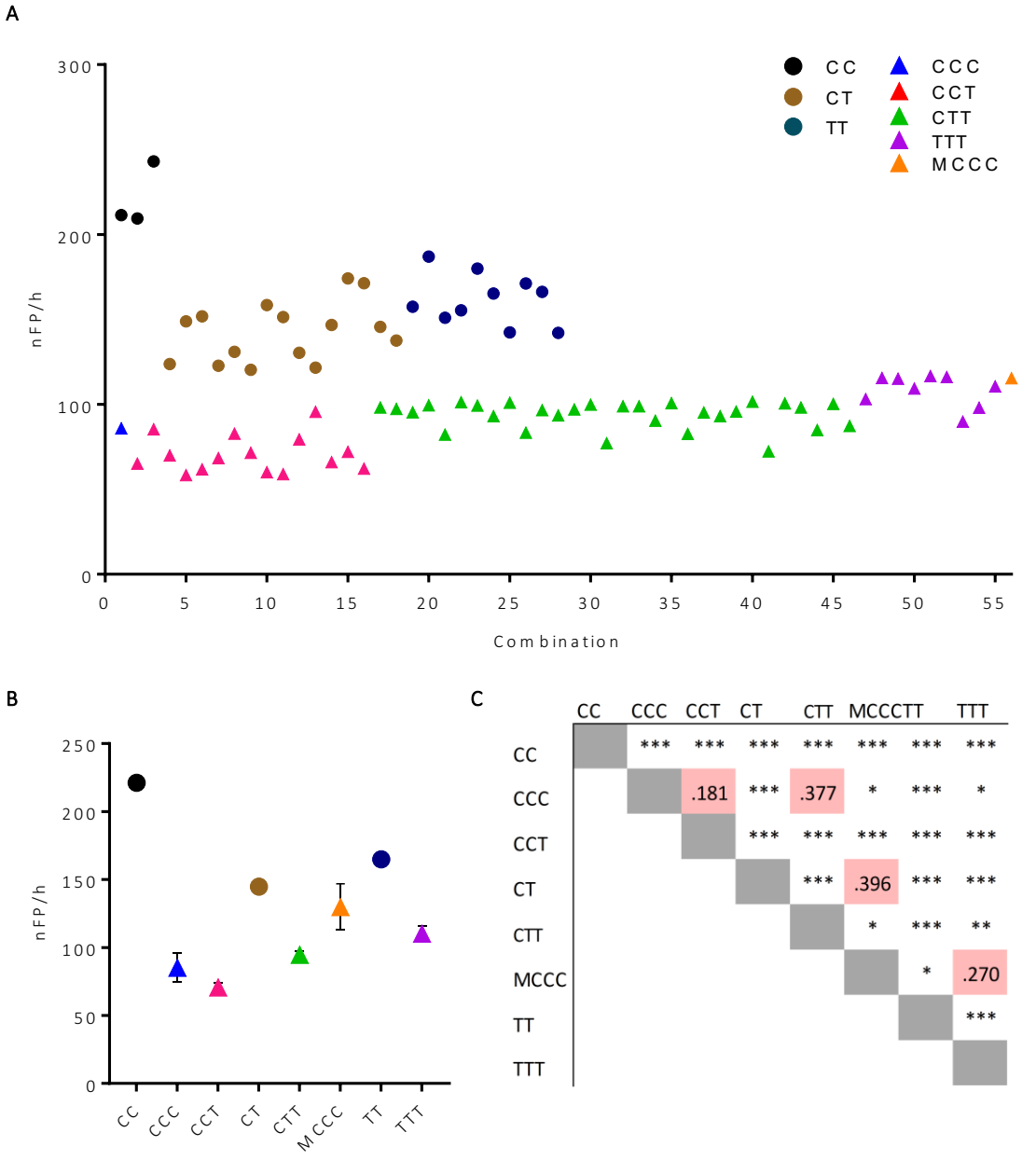


Figure 23: Average number of false positive detections per hour. (A) *Number of false positive displayed for each channel combination for three (triangle (▲)) and two (round dot (●)) brain areas used in the algorithm. Covariate factor for all combinations are calculated for threshold values of 0.2719 and 34.00% for sensitivity. (B) Groups of brain regions differ from each other on their average number of false positives per hour. (C) Significance between groups is depicted by the p-values (p<0.05 (*); p<0.01 (**); p<0.001 (***)). Red underlay represents non-significant differences. Combination composition indicated with groups 1: two intracortical recording sites (CC), category 2: one cortical and one thalamic recording site (CT), category 3: two intrathalamic recording sites (TT), category 4: three intracortical recording sites (CCC), category 5: two cortical recording site and one thalamic (CCT), category 6: one cortical recording site and two thalamic recording sites (CTT) and category 7: three intrathalamic recording sites (TTT). (for further information see Appendix 1a and 1b).*

Furthermore, a significant correlation between the achieved sensitivity and the nFP/h rate was revealed (Figure 24) (r=0.716; p<0.001). Again, a cloud formation of channel-combinations could be observed (Figure 24). Differences could be seen between predictions based on two recording traces displaying lower sensitivities and higher nFP/h in comparison to three channel. The addition of a third EEG trace for SWD prediction produced lower nFP/h and higher sensitivities. Intracortical CC group displayed with its high nFP/h and relatively high sensitivity for predictions based on two channels a clear differentiation to the decreasing slope displayed from intrathalamic TT to thalamo-cortical group CCT (Figure 24). Higher number of EEG traces (two channel and three channel) and influence of higher number of somatosensory cortex EEG traces within the prediction showed an interesting pattern that decrease the nFP/h and increases the sensitivity. Intracortical measurements for either two channel or three channel did not aligned with the decrease of nFP/h and increase of sensitivity displayed by the other groups. However, intracortical measurements displayed highest sensitivity of each version. Interestingly, two channel version thalamo-cortical prediction (CT) showed high sensitivity of two combinations (13 (Ctx V, rRTN): 29.26 ± 2.43%; 18 (Ctx VI, rRTN): 31.35

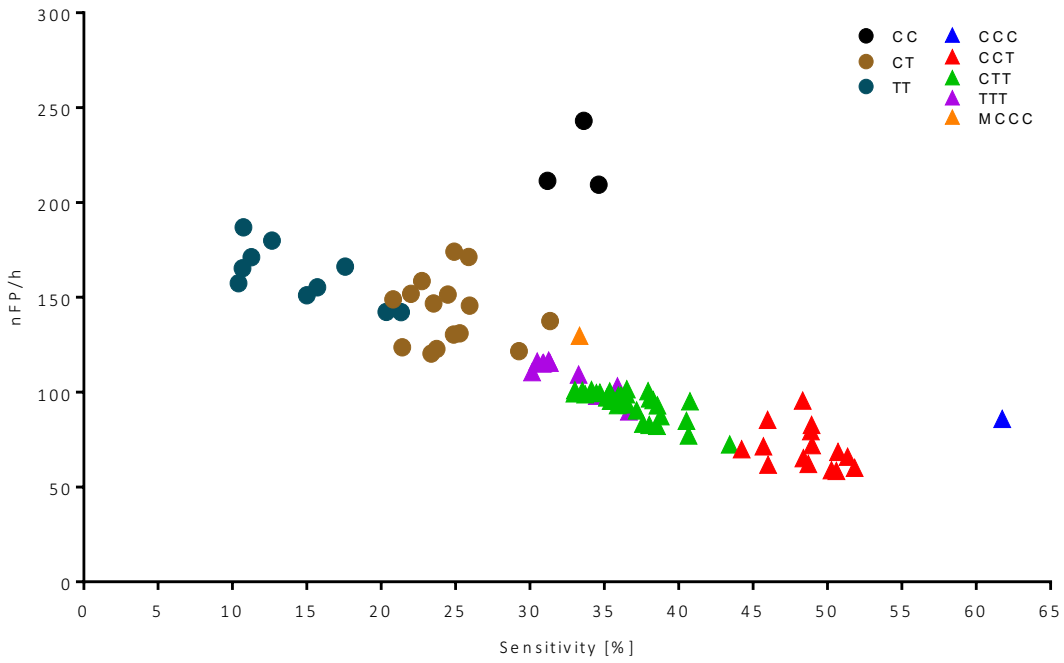


Figure 24: Correlation between nFP/h and sensitivity on precursor detection. *Combinations are depicted as either triangles (▲; three channel combinations) or round dots (●; two channel combinations). Different classes of channel combinations include groups 1: two intracortical recording sites (CC), group 2: one cortical and one thalamic recording site (CT), group 3: two intrathalamic recording sites (TT), group 4: three intracortical recording sites (CCC), group 5: two cortical recording site and one thalamic (CCT), group 6: one cortical recording site and two thalamic recording sites (CTT) and group 7: three intrathalamic recording sites (TTT) (for further information see Appendix 1a and 1b).*

± 2.43%) that did not vary significantly from intracortical measurements CC (1 (Ctx IV, Ctx V): 30.9 ± 2.6%; 2(Ctx IV, Ctx VI): 31.2 ± 2.6%; 3(Ctx V, Ctx VI):33.6 ± 1.5%) (all, p>0.05).

4.2. SWD prediction: Differences between rat strains

The overarching goal of seizure prediction is the detection of a precursor irrelevant of strain and species. As some differences between the two most widely used genetic rat models are already reported for some features of the SWDs itself (see introduction) it is aimed to compare the prediction-performance of the Maksimenko et al. algorithm based on EEG traces acquired in GAERS and WAG/Rij rats, respectively. More specifically, comparison of SWD prediction performance was compared between predictions based on three intracortical recordings acquired in either the somatosensory cortex (layer IV, V and VI) of GAERS or WAG/Rij rats (see paragraph 3.2.1), as highest sensitivities were reached for this channel combination in WAG/Rij. No significant differences were seen between GAERS and WAG/Rij rats for the sensitivity of precursor detection (GAERS: 61.0 ± 3.1%; WAG/Rij: 58.3 ± 1.9%; p=0.477) (Figure 25A). However, highly significant differences between GAERS and WAG/Rij rats were observed regarding the nFP/H (GAERS: 66.2 ± 21.5; WAG/Rij: 217.4 ± 13.2; p<0.001) (Figure 25B). Overall, the SWD prediction performance was significantly better in GAERS. Thus, indicating alternating

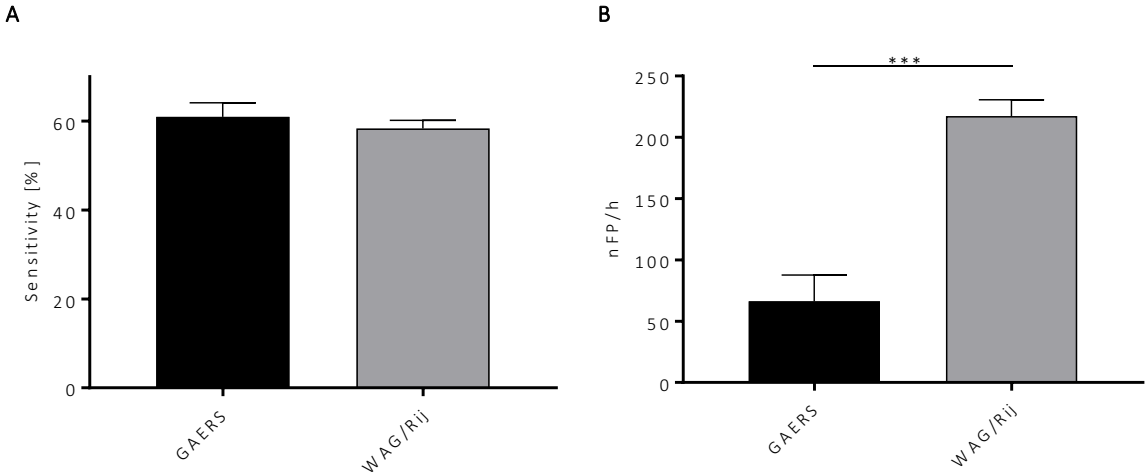


Figure 25: Differences in SWD prediction performance of the Maksimenko et al. algorithm in WAG/Rij rats and GAERS. *(A) Average sensitivity [%] of SWD prediction based on three intracortical recordings of the somatosensory cortex of either WAG/Rij rats or GAERS with covariate factor value 0.4127 for threshold and 175.911 for nFP/h. (B) Average NFP/h based on the same recordings in WAG/Rij rats and GAERS seen in A, with covariate factor value 0.4127 for threshold and 59.07% for sensitivity. *** indicates significance at p<0.001.*

efficiency of frequency band ranges for predicting absence seizure in GAERS compared to WAG/Rij rats.

4.3. Spectral comparison of true positive and false positive detections in GAERS and WAG/Rij rats

As seen above, SWD prediction resulted in a higher number of false positive detections in WAG/Rij rats compared to GAERS. This indicates different reliance on the calculated spectral features applied by the algorithm for precursor detection. To further investigate this issue and to uncover possible other spectral features, which might improve SWD prediction, spectral features of true and false detections were compared for GAERS and WAG/Rij rats.

Figure 26A-C depicts exemplary spectrogram images of true and false positive SWD predictions. Time point -0.5 to 0 features the analysis window in which either the true positive precursor or the false positive was detected. Time point 0 was used as a standardized reference point for averaging across multiple true positives and false positives respectively (Figure 26D-E).

The SWD start is depicted as time point 0.4 seconds on the x-axes (Figure 26A-C2). Additionally, high synchronicity is indicated at the starting point between channels by high values of the wavelet product ranging from 0.1 to 0.2 in the 5-10 Hz band shown on the y-axes. Comparing the true (Figure 26B) and false (Figure 26C) detection, differences in the degree of synchronicity can be seen. In addition, SWD patterns are clearly seen starting from the red line (2) in Figure 26B. SWDs show initially high synchronicity that recedes shortly after.

Next Figure 26D,E depicts average differences between true and false predictions. Average spectra of the wavelet product of true detections are subtracted from false detection to highlight the difference between both events. Here animal line specific differences are observed between WAG/Rij rats and GAERS (Figure 26D,E). In GAERS, higher power can be seen in comparison to WAG/Rij rats (Figure 26D,E) indicated by the maximum value of 0.02 (GAERS) to 0.004 (WAG/Rij). WAG/Rij rats and GAERS, display

negative values within the precursor detection window imply higher power within the false detections in both animal lines (Figure 26D,E-1). WAG/Rij rat EEG power showed a higher variance at the starting point, indicated by the less pronounced pattern between time points 0 to 1 second. However, at time point 0 higher power can be observed in comparison to the next second within WAG/Rij rats. GAERS product deduction of three

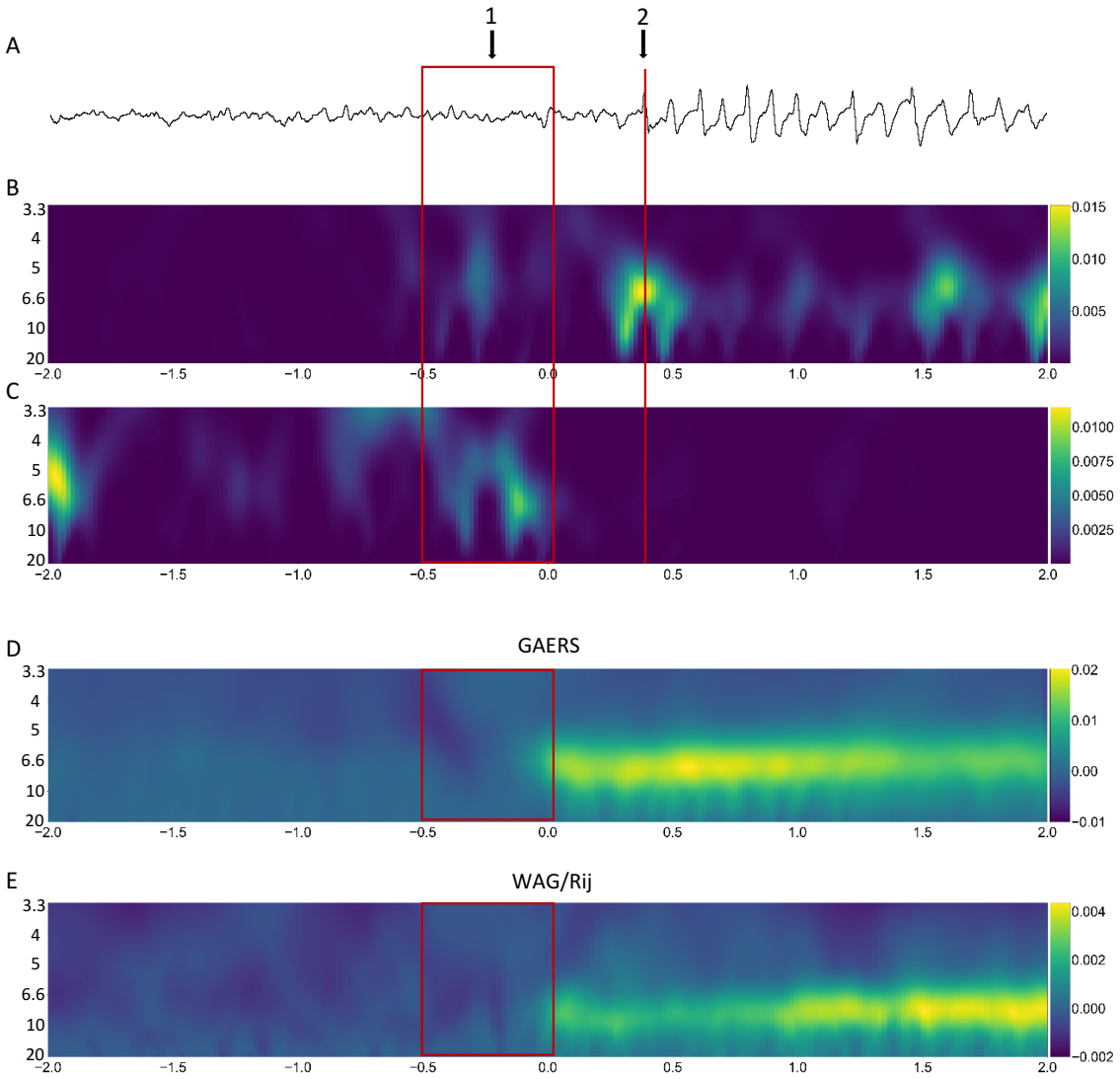


Figure 26: Spectrograms (product derived from three intracortical recordings) of true and false positive detections. *(A) Example of 4 s- trace of a somatosensory cortex FP measurement in GAERS. "1" denotes the analysis window including the true or false positive detection used for calculation of wavelet power. "2" denotes the start of the SWD. (B) Product of Wavelet power from three intracortical EEG traces belonging to a true detection of a precursor event. Analyzed Frequencies range from 3.3 Hz to 20 Hz and are displayed on the y-axes. Time point 0.0 on the x-axes represents calculated point of precursor event detection. Power differences are displayed on the right side with color coding from blue to yellow. Bright colors correspond to synchronized EEG signals of the three recorded traces. (C) Product of Wavelet power from three intracortical EEG traces belonging to a true detection (D-E) Average difference between true and false detections. Grand Average products of 300 true detection are subtracted from grand average product of 300 false positive detection events from GAERS (n=6). For WAG/Rij rats (n=10), grand average product of 262 true detections were subtracted from grand average product of 500 false detections.*

cortical EEG true events subtracted from false events, showed decreasing power beginning form time point 0.5 to 1.5 (Figure 27A).

The product of wavelet power seen for the three frequency bands $W_{\Delta S1}$ (5-10 Hz), $W_{\Delta S2}$ (3-5 Hz) und $W_{\Delta S3}$ (6.6-20 Hz) was statistically compared between true and false detections for GAERS and WAG/Rij rats (Figure 27). $W_{\Delta S1}$ of WAG/Rij rats exhibited significantly lower energy within true predictions (0.073 ± 0.021) in comparison to false predictions (0.123 ± 0.026; $p<0.01$) (Figure 27A). For $W_{\Delta S2}$ displayed differences between true (0.089 ± 0.023) and false (0.135 ± 0.023; $p<0.01$) as well. The frequency band resembling sleep spindles $W_{\Delta S3}$ on the other hand, did not significantly differ between true (0.066 ± 0.019) and false (0.088 ± 0.019; $p=0.118$) predictions.

GAERS data revealed a similar trend as WAG/Rij rats, in that $W_{\Delta S1}$ and $W_{\Delta S2}$ showed differences between true and false but $W_{\Delta S3}$ did not. $W_{\Delta S1}$ resembling the precursor activity within GAERS showed significant differences between true (0.128 ± 0.04) and false (0.320 ± 0.103; $p<0.01$) predictions (Figure 27-B). Light slow wave sleep ($W_{\Delta S2}$) had similar effects showing significant differences between true (0.280 ± 0.155) and false (0.431 ± 0.151; $p<0.05$) predictions. Both frequency bands showed lower energy during

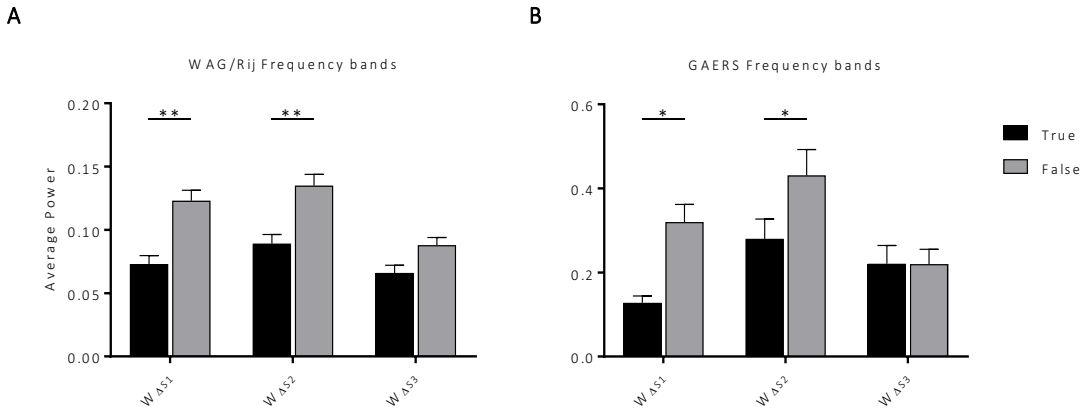


Figure 27: Comparison of product wavelet power of true and false positive detections for the three frequency bands used by the Maksimenko et al algorithm*. (A) Average product of wavelet power for the three frequency bands $W_{\Delta S1}$ = 5-10 Hz, $W_{\Delta S2}$ = 3-5 Hz, and $W_{\Delta S3}$ = 6.6-20 Hz calculated for true and false positive detections based on three intracortical EEG recordings in the somatosensory cortex of WAG/Rij rats. True detections are depicted in black while false positive detections are shown in grey. (B) Average product of wavelet power for the three frequency bands $W_{\Delta S1}$-$W_{\Delta S3}$ calculated for true and false positive detections based on three intracortical EEG recordings in the somatosensory cortex of GAERS. * denotes significance at p<0.05 level, ** denotes significance at p<0.01 level.*

epochs of true detection. $W_{\Delta S3}$ on the other hand did not express significant differences between true (0.221 ± 0.106) and false (0.220 ± 0.086; p=962) predictions.

Interestingly, differences in average power between true and false detections of the $W_{\Delta S1}$ band from GAERS were more pronounced in comparison to WAG/Rij rats ($p<0.05$). The same interaction could also be seen in the average power of $W_{\Delta S2}$ with GAERS having a more pronounces difference between true and false positive as compared to WAG/Rij rats ($p<0.05$).

4.4. Differentiation of true and false predictions using the random forest machine learning approach.

Frequency band based detection of precursor events prior to SWDs indicated differences between frequency bands of true and false detections. To further analyze and distinguish true and false detections from each other, the machine learning technique random forest was used. This technique tries to identify useful features for the differentiation of true and false detections within their spectral components. Nine spectral parameters ($W_{\Delta S1\text{-}3}$) of each true and false detection including intracortical, intrathalamic and thalamo-cortical measurements were fed to a random forest containing a total number of 1000 decision trees (Table 1).

Combination	*Area 1*	*Area 2*	*Area 3*	*N*
CC	Ctx IV	Ctx VI		177
CCC	Ctx IV	Ctx V	Ctx VI	161
CCC 90%	Ctx IV	Ctx V	Ctx VI	232
CCT	Ctx IV	Ctx V	VPM	161
CTT	Ctx IV	cRTN	VPM	145
TTT	Po	cRTN	VPM	162
GCCC	Ctx IV	Ctx IV	Ctx IV	1844/161/145
GCCC 90%	Ctx IV	Ctx IV	Ctx IV	2440

Table 1: Different groups of training sets used for the random forest machine learning approach. *Combination resembles the key reference for data used in random forest algorithm with each representing the three EEG trace regions used for prediction. Intracortical CC measurement featured two brain regions. N depicts the number of positive detections used in the data set with three variations for intracortical GAERS data. GCCC is the only combination done in GAERS whereas the rest is done in WAG/Rij rats.*

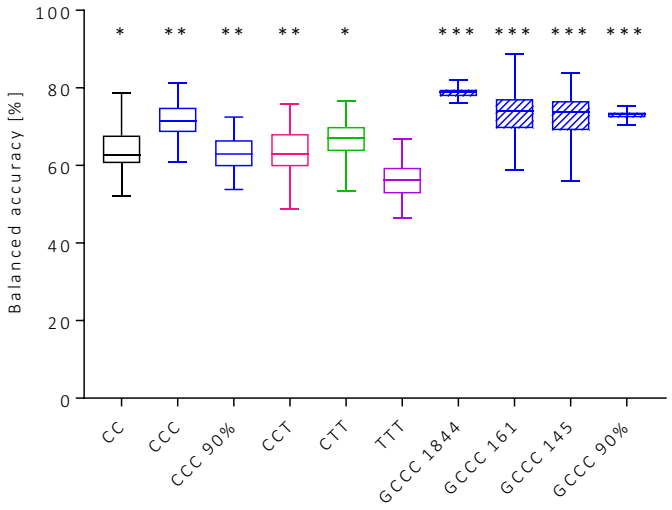
For training, each tree was fed with the parameters of 70% of the positive detections as well as an equal amount of randomly selected false positive detections (see methods section for details). For performance evaluation features of the remaining 30% were fed to the trained trees. Final decision whether a true or false prediction was presented to the forest is based on a majority voting of the trees. As in paragraph 4.1, the ability of the random forest to differentiate between true and false predictions was compared between different channel-pairs (Table 1). To further investigate the influence of random undersampling on the classification of events, 100 individual datasets of false detections where chosen. Deviations in balanced accuracy of combinations are attributed to dataset changes (Figure 28).

Applying the random forest algorithm to different brain region combinations, produced balanced accuracy ranging from 56.2 to 78.8% (Figure 28-A/B). Except for intrathalamic channel pair TTT ($p=0.167$), all groups produced significantly higher than chance classification, as proven by the permutation test ($p<0.05$ revealed for combinations CC and CTT, $p<0.01$ revealed for CCC, CCC 90% and CCT and $p<0.001$ for GCCC (1844, 161, 145) and GCCC 90%) (Figure 28A). Comparison of average balance accuracies between different combinations of analyzed brain region showed significant differences between the groups (Figure 28C) [$F(8, 891)=223.451$, $p<0.001$].

For WAG/Rij rats, sorting of true and false detection in the intracortical CCC dataset produced the highest accuracy (71.47%) which significantly varied from the balanced accuracies reached by all other groups (Figure 28C) ($p<0.001$). Using 90% instead of 60% sensitivity- based input data of the intracortical CCC measurements did not improve the separation of true and false prediction instead CCC with 60% showed higher balanced accuracies ($p<0.001$). CCC 90% on the other hand did not significantly vary from the intracortical CC (62.5%, $p=0.398$) combination and thalamo-cortical CCT (66.7%, $p=0.517$) combination. Sorting based on two intracortical EEG traces (CC) did not differ in balanced accuracy significantly from balanced accuracy achieved in the sorting based on CCT recordings ($p=0.844$). Sorting based on the thalamo-cortical combination CCT significantly differed from sorting based on two thalamic and only one cortical recording CTT ($p<0.001$). Sorting based on intrathalamic measurements without cortical influence produced the lowest balanced accuracy (56.2%, $p<0.001$).

Comparison of sorting-balanced accuracies form intracortical measurements of GAERS and WAG/Rij rats showed significantly higher differentiation ability of the trained random forest in GAERS animals. GCCC with n=1844 reached the highest balanced accuracy (78.8%, $p<0.001$). Significant deterioration of trainings efficiency (balanced accuracy) was observed once the trainings data set was reduced to equal values applied in the WAG/Rij rat intracortical measurement. However, no difference was seen in GAERS intracortical GCCC measurement with n=161 (73.58%) to the lowest WAG/Rij rats trainings set size (73%, n=145), indicating the irrelevance of data set size differences between 145 and 161. GCCC 90% shows no significant differences of its achieved classification success of 73.1% balanced accuracy between both reduced data sample 145 ($p=0.873$) and 161 ($p=0.409$) but still displayed significant lower balanced accuracies in comparison to GCCC measurements adjusted to threshold levels of 60%. This difference indicated a slightly lower classification success but overall classification is more successful due to 30% more preictal segments used in total.

A



B

Combination	Balanced accuracy
CC	62.5%
CCC	71.5%
CCC 90%	63.3%
CCT	66.7%
CTT	63.2%
TTT	56.2%
GCCC 1841	78.8%
GCCC 161	73.6%
GCCC 145	73.0%
GCCC 90%	73.1%

C

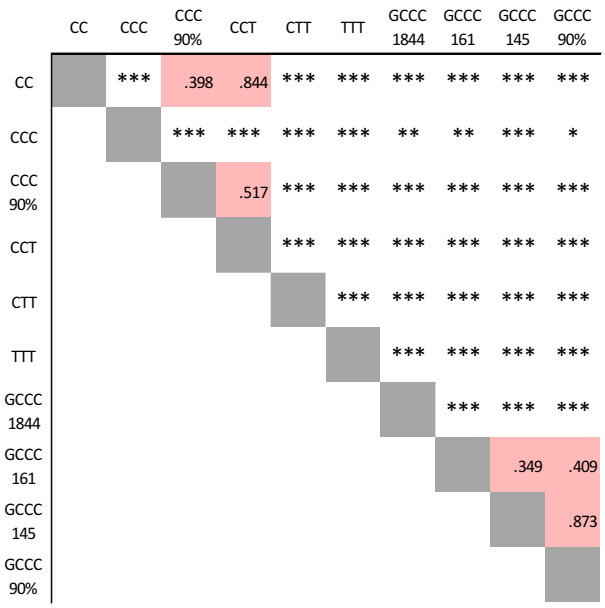


Figure 28: Average balance accuracies achieved for SWD prediction by the random forest algorithm. *(A) Boxplot analysis of balanced accuracy from random forest analysis of different combinations in WAG/Rij rats and GAERS. Boxplot analysis consists of 100 alternating seeds of chosen false positives (random undersampling) with all true positive predictions. Deviations in balanced accuracy of combinations are attributed to dataset changes. Tested data corresponded to intrathalamic, intracortical and thalamo-cortical prediction combinations (CC (n=177): Ctx V, Ctx VI; CCC (n=161): Ctx IV, Ctx V, Ctx VI; CCT (n=161): VPM, Ctx IV, Ctx VI; CTT (n=145): VPM, cRTN, Ctx VI; TTT (n=162): Po, VPM, cRTN). GCCC indicated intracortical measurements within GAERS animal with following training data size (n= 1844, 161, 145) to compare to WAG/Rij rats intracortical measurements and adjusting for lowest trainings number. All combination thresholds were adjusted to reach ~60% sensitivity. Comparison to the adjusted ~60% sensitivity is done with the intracortical WAG/Rij rats and GAERS measurement (~90% sensitivity). Note the non-significant training in combination "TTT". (B) Mean balanced accuracy from all 100 seeds of given combinations. (C) Statistical difference between groups. Note the non-significant difference between "GCCC 161" and "GCCC 145" ((p<0.05 (*); p<0.01 (**); p<0.001 (***)).*

Given this non-significant difference the GCCC 90% dataset can be regarded as most promising for SWD prediction. Validation of the algorithm performance on 30% (i.e. approximately 7.2 hours) of unseen data correctly detected 520 out of 754 SWD with only 161 false positives remaining, which equals a sensitivity of 68.97% with 77.32% specificity (Table 2). The high specificity shows that the random forest can more easily classify interictal events (prior false positive of the Maksimenko-algorithm). The validation set used here consists of 30% of the total data set (n=4880) made up of equal amounts of interictal and preictal events (random undersampling).

	Predicted as interictal fragment	*Predicted as preictal fragment*
Actual interictal fragment	549	161
Actual preictal fragment	234	520

Table 2: Classification matrix shown of GAERS intracortical somatosensory SWD prediction classification. *Data represents true and false positive received by the Maksimenko-algorithm, which was adjusted to achieve 90% sensitivity. A total number of 1464 true and false positive were used as a validation set corresponding to 30% of the trained data.*

Training of the random forest on five animals and validating the trained model on a complete unseen 24h data set showed lower averaged balanced accuracies (56.2 ± 1.93%) in comparison to previous used validation technique (70:30 split). However, it must be noted that high values SEM were seen between the rats for both specificity (57.4 ± 10.19%) and sensitivity (55.23 ± 10.18%) of prediction, indicating rat specific differences between obtained accuracies.

Corresponding to the hypothesis that the sleep spindle does not influence the classification of the random forest due to the non-significant effects seen prior in the surface plot, only the decision parameter for the light slow wave sleep and the precursor band are used for the classification. Classification was done on GAERS datasets adjusted

to reach significance levels of 90% to see differences to the prior determined balanced accuracies. Although there was no significant difference seen between true and false predictions within the sleep spindle band, classification was slightly reduced to 69.8%.

4.5. A deep learning algorithm for SWD prediction based on raw, unprocessed EEG traces

To investigate if next to the above used spectral features, additional features, which might be present in the EEG, are able to further improve the prediction of SWDs, a modified LeCun network was constructed (Figure 9). Considering this form of SWDs without prior feature specification and modulation, a highly specific deep learning algorithm was applied analyzing the raw EEG of three intracortical recordings of GAERS with 512 ms time windows to predict if a SWD will occur within the following 256 ms (Figure 20). Random undersampling was applied to the data to counter unbalanced training of features. Training of the deep neural network with a data set containing 2695 interictal sample to match the 2695 preictal sample. After 75 training epochs, classification of interictal and preictal events on unseen test data was performed and resulted in an averaged balanced accuracy of 75.1%. This balanced accuracy was shown to be significantly better ($p<0.001$) than chance-based prediction by permutation statistics (see methods for detail). The training of features reached higher negative predictive value (82.4%) in comparison to the lower precision (67.75%) (Table 3). Figure 29A depicts the average-balanced accuracy at each training epoch. Averaged balanced accuracy over 10 validation sets of SWD prediction stagnated between the 40^{th} (76.06%) and 50^{th} (75.68%) trainings epoch, indicating that features of SWD precursor within this dataset are fully learned. This notion was also supported by the average loss and validation loss (val-loss) values depicted in Figure 29B for each run. Training loss of the trainings set showed improved learning to the last training epoch. However, validating error of the learned features with the unseen data (val-loss) stagnated beginning from the 40^{th} epoch (Figure 29B). After the 40^{th} epoch, val-loss values kept increasing, indicating slight overfitting of learned features.

	Predicted as interictal fragment	*Predicted as preictal fragment*
Actual interictal fragment	2221	474
Actual preictal fragment	869	1826

Table 3: Confusion matrix of SWD prediction based on raw EEG data using a modified LeCun network. *A total number of 2695 preictal and 2695 interictal events had to be classified.*

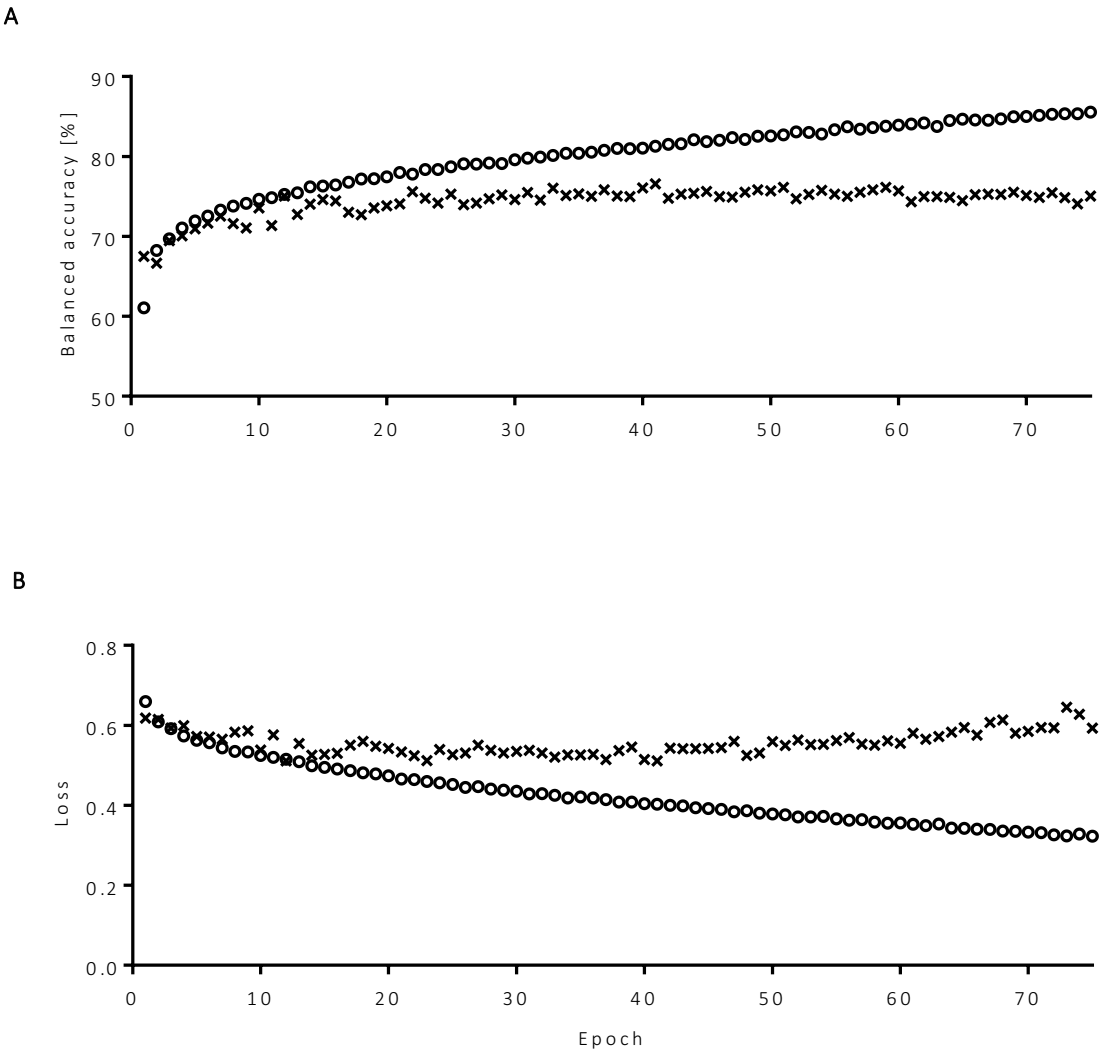


Figure 29: Training efficiency of the modified LeCun network in the prediction of SWD by raw EEG analysis. *512 ms raw EEG is analyzed on precursor activity signaling EEG occurrence within the next 256 ms. (A) Average-balanced accuracy of a 10-fold data split is depicted for each training epoch. Circles (○) indicate balanced accuracy during training, while crosses (×) indicate balanced accuracy of unseen test/validation set. (B) Average loss of training difference between real and predicted data is displayed for each epoch for the training set (round dots) and validation set (crosses).*

To validate the algorithm on a full 24h dataset, the algorithm is trained on 5 animals with one unseen animal functioning as a validation set. Validation on a full 24h dataset showed a balanced accuracy of 68.2 ± 0.04% (

Table **4**). Although a high number of false negatives where correctly detected (8579), they appear comparable small in relation to the accurately detected false negatives (144829) resulting in a precision of 94.4 ± 0.8%. Less than halve of the preictal EEG fragments were accurately classified resulting in 42.3 ± 4.0% sensitivity.

	Predicted as interictal fragment	*Predicted as preictal fragment*
Actual interictal fragment	144829	8579
Actual preictal fragment	259	190

Table 4: Confusion matrix of SWD prediction based on raw 24h EEG data using a modified LeCun network. *Confusion matrix shows classification average of 6 tested 24h measurements with a total of 162963 tested EEG intervals for each rat.*

5. Discussion

The aim of this thesis was to optimize the prediction of absence seizures using a combination of different techniques. For this, the influence of different EEG recoding sites within the thalamo-cortical system on the predictability of absence seizures in WAG/Rij rats was elucidated using a previously developed wavelet-based SWD prediction algorithm of Maksimenko et al. (2017). Additionally, the performance of the algorithm was investigated on its applicability to another absence epileptic rat model GAERS. Furthermore, surface plots of true and false prediction of WAG/Rij rats and GAERS were compared and investigated regarding their suitability of being used by a random forest machine learning algorithm in order to further classify true and false positive predictions to improve overall prediction of SWDs. Finally, a deep learning algorithm was constructed and trained to investigate the possibility of SWD prediction based on unbiased and unprocessed raw EEG data of preictal and interictal epochs.

Corresponding to the three postulated aims, the results reported in this thesis demonstrated significant improvement for each method used to optimize SWD prediction: A significant increase in predictability was shown by using three intracortical EEG traces in the somatosensory cortex in respect to the first aim regarding the question about the best fitting brain region combination. Comparison of wavelet spectra of true and false positive predictions revealed significant differences in two of the three frequency band energies within true and false predictions. Using the extracted features of the Maksimenko-algorithm with either threshold levels corresponding to 60 or 90% sensitivity, allowed further classification of true and false predictions, reaching a balanced accuracy of 78.8% and 73.1% respectively by applying a machine learning algorithm called random forest. Lastly, the applied deep learning algorithm was able to predict absence seizure with an average accuracy of 75.5% based on raw EEG traces without previous self-defined feature engineering.

The following chapter will concentrate on the optimization of pre-existing and new founded methods predicting absence seizures and their correlation to the already published literature. Finally, the achieved findings of the efficiency of these techniques will be evaluated.

5.1. Region specific SWD prediction

This study expanded the view on the involvement of brain regions within absence seizure prediction in the way that an improvement of prediction with EEG recordings in the vicinity to the onset zone of SWDs can be achieved. The research aimed to improve the prediction of SWDs by the algorithm developed by Maksimenko et al. (2017) that previously showed high sensitivity for precursor prediction but a high number of false positive as well. It was postulated that high number of false positive predictions might be correlated to the single brain region combination used for the prediction. Therefore, the effectivity of the algorithm for the prediction of seizures with combinations out of eight different brain regions within the rat model WAG/Rij rats was investigated in this thesis.

The algorithm of Maksimenko et al. determines the synchronicity of brain regions by calculating the product out of different EEG traces. Therefore, it is not surprising that the use of either two or three channels within the prediction showed a significant difference with three channels providing higher predictability of seizures. Using the product for an indication of synchronicity provides an intricate play between enough information present to detect certain patterns and random synchronization of brain regions. The product of two EEG traces might provide less differentiable information between preictal and interictal events. Increasing the number of used EEG traces to three offers a better indication of synchronization.

This work pointed out the importance of represented brain regions within the prediction, as increasing the number of somatosensory EEG traces resulted in the highest predictability. In comparison to previous research where two cortical and one thalamic region (Cortex IV, V and Po corresponding to combination no. 2) were used for the prediction, prediction based on intracortical regions in S1 tested in this study provided significantly better results (see Figure 22 and Figure 23 combination no. 1 (intracortical) and no. 2 (thalamo-cortical)) (cf. Maksimenko *et al.*, 2017). The coupling of brain regions preictally might indicate a possible explanation for the higher influence of cortical regions on the predictability (Sysoeva *et al.*, 2016). Sysoeva investigated the dynamic coupling seen prior and during SWDs showing a gradual increase of coupling between intracortical regions with additional increase in coupling between cortical and thalamic regions shortly

thereafter. Somatosensory cortical regions IV, V and VI are seen to drive the thalamus preictally except for bidirectional crosstalk seen between Po and somatosensory cortical region IV and V (Lüttjohann & van Luijtelaar, 2012). Surprisingly, in this study the thalamo-cortical combination with two cortical regions and the thalamic region Po did not produce significantly different predictabilities compared to other combinations with one thalamic and two cortical regions. As there are time delays between cortex and Po of approximately 10-12 ms (Lüttjohann and van Luijtelaar 2012; Lüttjohann and van Luijtelaar 2013) better results might be achieved once this signal delay is attributed for. The algorithm applied in this study did not adjust the delay between different brain regions. Implementing a function of adjusting the physical delay between brain regions might provide further insight into the precursor occurrence and increase SWD predictability. The high predictability of SWDs seen once using intracortical somatosensory cortex EEG traces can nicely be related to the assumed onset zone of SWDs, in which the precursor determining parameter are more prominent (Meeren et al. 2002; Polack et al. 2007; van Luijtelaar et al. 2011). This is further supported by the fact that SWD prediction within motorcortex (MCCC) reached a less accurate prediction.

Interestingly, intracortical prediction based on only two intracortical recording sites in S1 was not distinctively better than two thalamo-cortical predictions with one thalamic and one cortical region. These two thalamo-cortical combinations showed surprising results. More specifically, both combinations used two channels for the prediction with both using the rRTN and alternating cortex V and VI for the prediction. Previous research showed initial interactions of cortical regions with the rRTN, which might influence the predictability (Sysoeva et al. 2016). Additionally, the rRTN contains a high importance for seizure formation as pharmacological inhibition or lesions of it decrease SWD activity (Aker *et al.*, 2006; Meeren *et al.*, 2009). A recent study debating over the influence of the centromedian thalamus (CM) on the seizure formation showed alterations of SWD patterns once local pharmacological inhibition is applied to the CM (Terlau *et al.*, 2020). This novel view on the role of the CM to seizures in general might be another target for seizure prediction or treatment via stimulation, due to its high relevance to SWD formation.

The predictability of SWDs were analyzed in two rat models GAERS and WAG/Rij rats to investigate the algorithms interchangeability. Contrary to previous assumptions, WAG/Rij rats and GAERS showed a significant difference on the predictability of SWDs. GAERS produce less false positive, which might be related the higher power seen preictally (Akman *et al.*, 2010). GAERS and WAG/Rij rats display slightly different prominent frequencies during seizure in any stage (start, mid, end), with GAERS showing slightly lower frequency of 8 Hz and WAG/Rij rats showing 9 Hz at the start of the seizure (Akman *et al.*, 2010). Additionally, waveform differences of GAERS and WAG/Rij rats might also be extended to the precursor. Contrary to the finding in this study, Akman et al., (2010) showed that WAG/Rij rats display higher power between 8 and 14 Hz preictal, which would influence the sleep spindle parameter used in the applied algorithm. Moreover, the study showed a generally higher difference between baseline events and pre-SWD events, which would lead to contrary believes that precursors can be more accurately detected in WAG/Rij rats than in GAERS (Akman *et al.*, 2010). This difference might be correlated to the EEG recording area within the frontal cortex. Furthermore, a highly specific pattern was seen within S2 and S1 regions preictally with the temporal starting point occurring previously in S2 regions, further supporting the importance of onset zone specific precursor prediction (Pinault et al. 2001; Zheng et al. 2012). This could be an indication for the efficiency of chosen frequency bands for the prediction, which might be more fitting for GAERS SWD prediction. This shows the first indication of distinct differences of precursor morphology.

Concluding, intracortical SWD predictions based on three distinct EEG traces showed significantly higher predictability in comparison to two intracortical EEG traces. The additional information presented by the third channel might be responsible for the detection of more SWDs. The interplay between cortex IV, V and VI might be the information needed within this prediction. Resulting grouping accuracies seen between all combinations can also be traced back to the cortical focus theory in that the network aspect is further described by the sensitivities achieved (Meeren *et al.*, 2005).

5.2. Frequency analysis

Since the earlier mentioned precursor detection based on the algorithm of Maksimenko et al. (2017) had three distinctive criteria all concerning the energy within frequency bands, investigation of the frequency band differences between true and false positive prediction were done. Comparing the two animal models GAERS and WAG/Rij rats, true and false predictions displayed a lower overall product in WAG/Rij rats. This supports the idea that precursor wavelet energy is higher even in preictal events as it was shown in initial starting 500 ms of SWDs (Akman *et al.*, 2010). Interestingly, precursor analysis of true and false predictions within GAERS displayed the slightly different pattern in that false predictions produce higher average power (Figure 25). As mentioned above, differences can be observed at the start of the seizure in GAERS and WAG/Rij rats. While GAERS show an increased power and synchronicity right at the beginning with steadily decreasing power after the initial second, WAG/Rij rats show low power at the start while after the first second the power and synchronicity increase. This difference in power could be attributed to species specific SWD patterns that are slightly different throughout a SWD (Akman *et al.*, 2010).

The research in this study showed considerable differences in the first and second parameter corresponding to slow wave sleep and precursor frequency respectively between true and false predictions. Both parameters showed lower average power in true predictions. It cannot be ruled out that physiological brain activity produces similar synchronization interactions of brain regions due to the fact that frequencies between 4-12 Hz are seen within normal wake related oscillations (Pinault et al. 2001). However, such combined delta and theta interaction are seen rarely (Sitnikova *et al.*, 2009). A higher chance of SWD formation is seen during vigilance state corresponding to drowsiness and transitional states where EEG frequencies are slowing down (Drinkenburg et al. 1991; van Luijtelaar et al. 2011). Episodes of drowsiness as well as preictal temporal EEG segments are proven to include higher delta activity with additionally seen theta activity (van Luijtelaar *et al.*, 2010). The precursor seems to consist of a lower power in comparison to the rarely occurring delta and theta synchronously appearing events. Considering the third parameter, frequency band corresponding to sleep spindle did not display any differences between true and false predictions. There is no indication that

sleep spindle activity, which display similar frequencies (7-14 Hz) as the precursor (5-10 Hz), it not important for the differentiation between normal sleep activity and pathological SWD formation (Maksimenko *et al.*, 2017). SWDs and sleep spindles are in some ways related; however, SWDs produce higher total power in beta frequencies (12.5-30 Hz) (Gloor, 1968; Kostopoulos, 2000; Sitnikova & van Luijtelaar, 2009). Therefore, it is not surprising that energies of true and false prediction did not differ within the third parameter due to the abundantly present sleep spindle during drowsiness and the higher error rate (Sorokin *et al.*, 2016).

Differences between true and false predictions energy band power were significantly more pronounced in GAERS compared to WAG/Rij rats. True to false prediction values displayed higher difference in GAERS for both the precursor and the light slow wave sleep parameter in comparison to WAG/Rij rats. These results were controversial due to previous studies showing similar differences between baseline and pre-SWD events in both GAERS and WAG/Rij rats (Akman *et al.*, 2010). The length of the applied analysis window and grouping of the frequencies could explain this disagreement. However, it cannot be ruled out that other factors like the duration of the precursor event and prominent frequencies are involved in the slope difference as there are other differences seen between SWDs, which are in some ways relatable to the somatosensory recoding site (Akman *et al.*, 2010). The more pronounced differences between true and false detections in this study might also explain the better predictability of SWDs in GAERS compared to WAG/Rij rats.

5.3. Prediction optimization using random forest

As stated in previous experiments, initial differences were seen in frequency bands of true and false predictions. To analyze this pattern, it was hypothesized that a machine learning algorithm termed random forest is able to divide true and false predictions to improve overall prediction of SWDs. Using a novel approach to apply a random forest to prediction decision parameter might lead to a significant reduction of the encountered problem of a high number of false positive (Maksimenko *et al.*, 2017).

The machine learning approach showed that additional information is still present in the chosen parameter. The parameter allowed significant differentiation of precursor and false positive predictions. As it was mentioned before, seizure formation comprises of an intricate interplay of a network of neurons. Therefore, the given results are not surprising in that efficient predictions require more logical decisions as provided in the algorithm used (Blumenfeld, 2005; Maksimenko *et al.*, 2017). This novel approach to add a random forest to the prediction output is hard to put in context to previous research. Although other groups have shown efficient seizure prediction when working with EEG data, the dataset was mostly comprised of different types of seizures (Wang *et al.*, 2019). Using the CHB-MIT database providing intractable seizures of pediatric subjects, they achieved high accuracies (84.8%) by analyzing time intervals of 15 minutes with wavelet packet features and later on decoding the output with a random forest algorithm. These results show that a random forest can be used to further decode preprocessed EEG data and thereby function as an early seizure indication system. Manzouri et al. (2018) discussed the improved efficiency of random forest algorithms in comparison to another machine learning algorithm termed support vector machine (SVM). SVM try to find a hyperplane within the data that separates patterns linearly. Here, the random forest reacted quicker to detect the focal seizure in comparison to the SVM, which would coincide with the postulated goal to predict seizure online without considerable delay.

Multiple recoding site combinations were tested for their still retaining information considering the seizure predictions. The results showed the same trend as received from the Maksimenko-algorithm in that processing intracortical recordings with the random forest reached higher balanced accuracies in comparison to combination groups having thalamic EEG traces. These results coincide with the previous experiments by the earlier shown synchronization of somatosensory cortical regions than in thalamic regions (van Luijtelaar *et al.*, 2010). Interestingly, differences between GAERS and WAG/Rij rats could still be seen within intracortical measurement. It can be excluded that this difference is solely attributed to the higher training samples in that the influence of the small dataset size present in WAG/Rij rats was tested by reducing the training data from intracortical GAERS measurements to match the WAG/Rij rat intracortical sample size. This reduction to either a sample size of 145 or 161 showed still improved balanced accuracies. The

frequency bands between 3-5 Hz and 5-10 Hz are the main driving forces for the further differentiation. However, there is still an interaction of all three parameters due to a slight decrease of 3-5% balanced accuracy once the third parameter corresponding to sleep spindle (7-20 Hz) is excluded for the classification. Furthermore, the classification result of the random forest further supports the notion that the selected frequency bands are better suited for precursor detection in GAERS.

Interestingly, parameters extracted from WAG/Rij rats and GAERS were chosen to resemble equal sensitivities (e.g. Maksimenko-algorithm thresholding to 60% or 90%), however classification success of true and false positive predictions was improved in GAERS. To test the information available, in higher sensitivities, two different threshold levels applied in the Maksimenko-algorithm were chosen to either reach a prediction sensitivity of ~60% or ~90% for intracortical measurements in both WAG/Rij rats and GAERS. Both intracortical predictions with threshold levels adjusted to reach ~60% sensitivity showed higher predictabilities in comparison to intracortical measurements in 90% sensitivity. This difference demonstrates the more complex features, present in the additionally detected SWDs within the 90% version. The reduction of the rat specific threshold reduced the required energy within the frequency bands. Reducing the threshold is always accompanied by an increase in number of false positive due to the limiting factor of irrelevant EEG pattern that cause false positive declaration. Although additionally obtained true and false positive parameters reduce the balanced accuracy (-5.7% in GAERS) of the trained model, but still higher sensitives are obtained in the end. However the additionally detected precursors by the lower threshold values introduce more versatile problems of classification that need more information as the extracted parameters that were retrieved from to the Maksimenko-algorithm.

Although training of features was heavily influenced by the training sample size, quite convincing results were also reached with a medium sized dataset of 145 events. The required sample size to learn certain features is closely correlated to the information retained in the data. Generally, increasing sample sizes result in higher balanced accuracies and show stronger stability of predictions (Kim, 2009).

Finally, random forest algorithms are an efficient tool to differentiate true and false predictions further. Random forests show fast processing power and in many cases outperform other machine learning algorithms like SVM or neural networks (Liu *et al.*, 2013; Manzouri *et al.*, 2018). With the rather small sample size and its tabular preprocessed data format, the random forest is the optimal tool to show classification success. Another positive feature of random forest is seen in its internal hyperparameter that do not need to be heavily optimized (Nitze *et al.*, 2012). Highly unbalanced data structures such as in this study with low numbers of true possible predictions and high false positive predictions might benefit from bigger data sets as seen in the comparison of samples sizes in the intracortical GAERS data set. Although random undersampling is an established method to work with unbalanced data sets, valuable data received from experiments might not be presented to the algorithm (He & Ma, 2013). Easily discernable true and false predictions within the data set might result in a wrong assumption on the achieved balanced accuracies. False positive detections that are similar to true detections might not be discernable in some cases as it was seen by the reduction of balanced accuracy in intracortical WAG/Rij rats and GAERS prediction parameter with ~90% sensitivities.

However, by alternating the chosen false positive 100 times, as performed in this thesis, a good impression on the prediction range is provided. An alternative approach to random undersampling is an oversampling of the underrepresented class. While this approach allows to account for a bigger variety of false positives to be considered by the algorithm, this approach bears the risk of unbalanced learning and overfitting of predictions features, which is avoided by the random undersampling approach chosen in this thesis.

Applying both the Maksimenko adjusted to reach 90% sensitivity and random forest classification on-line would result in a reduction of precursor sensitivity by ~30%. However, only ~25% of the false positive would remain as well. High differences were seen between two types of validation. The 70:30 split showed higher accuracies in comparison to a validation of an unseen rat, indicating rat specific parameter like EEG measuring electrode location in relation to the onset zone. As a consequence, a trained model used in an on-line application might benefit from an initial training phase that

recalibrates the random forest model for the specific animal. Alternatively, longer lasting recordings of a single animal might allow training of the random forest algorithm in a rat specific manner.

5.4. Unbiased raw EEG prediction with deep learning

Previous approaches aimed for a precursor prediction with an a-priori preprocessing to extract features, which implies a parameter limitation and loss of information to a certain degree. Therefore, this part of the dissertation focuses on a deep learning algorithm that defines features for the prediction of absence seizure in intracortical EEG traces within GAERS. The deep learning approach was only applied in GAERS, as this method relies on a large amount of training data, which were only available in the long-lasting 24 hour measurements performed with GAERS (Alwosheel *et al.*, 2018).

In general, deep neural network models are an efficient tool to work with a multitude of data and thereby are promising candidates for EEG data. Previous research showed promising EEG information decoding using deep neural networks especially known for its high variability of hyperparameter and task specified layered build (Bashivan *et al.*, 2016; Li *et al.*, 2020). Here, a novel approach is presented to predict seizures applying a combination out of feature extracting layers, a LSTM layer responsible for decoding inherent temporal dependencies and three fully connected layers that are responsible for the differentiation of the classifiers preictal and interictal.

In the first step, multiple convolutional layers designed to work best with image or sequential data process the raw EEG data (LeCun, 1998; Bashivan *et al.*, 2016). This step is also known as feature extraction. Temporal and special features of EEG are picked up and identify different features corresponding to SWD formation (Goodfellow *et al.*, 2011). Here, the reducing kernel (shifting window) of the convolutional layer extracted EEG information from 40 ms to 4 ms. In this study's case, two max pooling steps after the consecutive convolutional layer reduced the data to decrease the training time. Following the convolutional layer, a LSTM with 100 units analyzed the inherent temporal dependencies within the features extracted from the convolutional layer. Although,

hyperparameter optimization like randomized search or grid search was not applied in this deep learning model, due to high processing power requirements and low chance of significant improvement of balanced accuracies over a few percent, significant predictions of absence seizure were still possible.

The deep learning model showed low levels of overfitting by the slight increase seen in validation loss values after the 50th training epoch. However, due to the addition of dropout layers, overfitting was reduced (Srivastava *et al.*, 2014; Cogswell *et al.*, 2016). This slight overfitting did not show significant alteration in levels of balanced accuracies between the 50th and 75th training epoch. The confusion matrix of the 75th run showed a higher detectability of interictal epochs in comparison to preictal epochs, which could also be translated to a 67.75% sensitivity and 82.4% specificity. The low level of sensitivity might be attributed to the randomized undersampling method used. Episodes of interictal EEG have a higher chance to appear with similar EEG pattern due to the unproportionally high count of interictal in general. This notion was already seen in van Luijtelaar et al., 2016 where false positive detection occurred with the highest frequency during light sleep. Additionally, active wakefulness and deep slow wave sleep showed seldomly SWDs. Randomly selected interictal as in this studies case might contain equal amounts of EEG fragments of wakefulness, light and deep sleep and therefore does not account for the imbalance seen of SWD frequency during all vigilance states. Further, deep learning in general need big data sets to train for specific pattern as seen in this case (Alwosheel *et al.*, 2018). In addition to that, the unknown length of the precursor and precursor pattern is hard to discern from normal interictal events with small sample sizes. In this case, interictal EEG data mostly comprises of background EEG during sleep and wakefulness and movement artifacts. Sleep and wakefulness seen in EEG show a high discernibility due to their steady character over longer periods. The exact time of precursor occurrence is not defined (van Luijtelaar *et al.*, 2010). However, previous research showed timely coupling changes between brain region and differences of delta and theta prominent frequencies between cortical and thalamic brain regions (Ovchinnikov *et al.*, 2010; van Luijtelaar *et al.*, 2010; Sysoeva *et al.*, 2016). These precursor indications were seen up to two seconds prior to the actual start of the seizure. The detection window size plays a major role in prediction delay in a way that the algorithm

needs to be fed with previously recorded data. Therefore, using an analysis window of two seconds would delay the processing of the data. Hence, the approach of this study was to investigate the direct time interval 512 ms prior to the seizure with a slight shift in actual seizure start of its predictability. A time window of 512 ms was chosen to allow measuring frequencies within previously used parameters. The results indicate that a time interval of 512 ms is sufficient to predict seizure to a certain degree.

Testing the learned features corresponding to the precursor on an unseen 24 H dataset showed on average a low sensitivity. The same problems seen within the application of the Maksimenko-algorithm combined with the random forest on an unseen dataset of 24 H was again encountered with the deep learning approach. This result adds to the prior stated assumption that the precursor prediction needs a specific recalibration phase where the weights within the model are fine tuned to a specific rat. These differences between animals was already seen before and addressed by applying a rat specific threshold (Maksimenko et al., 2017).

As mentioned before, a strong limiting factor of deep neural networks is its requirement of a high amount of samples to detect and extract precursor defining features (Alwosheel *et al.*, 2018). Deep neural networks processing of raw data require significantly higher amounts of samples in comparison to the earlier applied random forest that utilize preprocessed data. The precursor detection applied here aims to be applicable in patients, therefore a tradeoff of data size needs to be made.

In summary, it can be concluded that absence seizures can be predicted without prior feature engineering like wavelet transform or preprocessing by applying the Maksimenko-algorithm (Maksimenko *et al.*, 2017). However, further analysis is needed to reduce the analysis window to a minimum without a reduction in precursor predictability. Although this deep neural network shows efficient sorting of preictal and interictal events, the required heavy processing power will influence the speed of the prediction significantly. On-line approaches using this algorithm would most probably take more time in comparison to the previously applied algorithm, thereby delaying putative stimulation. These on-line predictions would show increased SWD predictability once the applied deep learning model is adjusted for the specific animal and more

training data is supplied. Comparing both data processing techniques random forest and the deep neural network, it is shown that similar accuracies were achieved. However, with the heavy preprocessing of the random forest by the Maksimenko-algorithm and its applicability in an on-line setting, random forest appears more appropriate to further investigate within new patients or experimental animals.

5.5. Translation to CAE

The ultimate goal on SWD prediction is a putative transfer of the above described improved SWD prediction algorithm(s) towards an application in human CAE EEG data. The biggest huddle in the comparison from rat to human EEG is might be the extraction of the optimal frequency bands for wavelet based precursor detection. While for the rats wavelet power in the frequency bands 3-5, 5-10 and 7-20 Hz were determined a study of Gupta et al., investigating the preictal period in EEG and MEG recordings of children with CAE, early preictal sources in the 2-4 Hz frequency were detected the difference in frequencies found and synchronizing brain regions preictally (Gupta *et al.*, 2011). Although human EEG did provide information about a putative precursor, it is not in the same frequency range as used in this study. A modulation of frequencies must be done prior to the translation. CAE in general consist of slower frequencies (3-5 Hz) in comparison to the most prominent frequencies in rat SWDs (7-11 Hz). Therefore, this change might also be extended to the precursor. After modulation of the parameter, an identification of the precursor might be possible with the Maksimenko-algorithm. The most promising improvement to the seizure prediction might be achieved by a combination of the Maksimenko-algorithm and a further classification of true and false positives by a random forest. Although the applied deep learning algorithm provides a classification of interictal and preictal EEG events, random forest provides a faster processing of information and is to a higher degree more stable to low sample sizes. The deep learning algorithm applied here needs multiple layers to process precursor irrelevant information and to learn deciding features for preictal and interictal events. These features might be rather small and are easily overlooked by big data structures as seen in raw EEG.

6. Conclusion and outlook

In this thesis, optimization of seizure prediction was shown to increase significantly by limiting EEG recording to intracortical brain regions in the proximity of the seizure onset. It was seen that precursor detection could be further improved by utilizing a classification step after the Maksimenko-algorithm preprocessed the EEG data, by applying a random forest approach. Lastly, by applying a deep learning approach it was shown that preprocessing of EEG data is not required to classify interictal and preictal EEG events. The reduction in unnecessary false positive might help in the development of a putative treatment approach. Roughly 15% of patients are unresponsive to available treatments, while available medication produces side effects like nausea, vomiting, sleepiness, insomnia and hyperactivity (Callenbach *et al.*, 2009; Glauser *et al.*, 2010; Puka *et al.*, 2020). An interesting alternative to standard medication and surgeries might be the application of brain computer interfaces (BCI). BCI provide multiple possibilities to work with patients experiencing seizures like the application of neurofeedback or deep brain stimulation (DBS) (Osterhagen *et al.*, 2010; Zangiabadi *et al.*, 2019). DBS was already successfully used in multiple settings, with Parkinson's disease as one of the most popular one (Halpern *et al.*, 2007). The idea to induce a stimulation onto specific brain regions thereby altering the firing pattern of neurons present is a promising approach to address neurological disorders. Previous research already showed initial success of high frequency stimulation within thalamic nuclei in both GAERS and WAG/Rij rats (Vercueil *et al.*, 1998; Lüttjohann & van Luijtelaar, 2013). This high frequency stimulation resulted in local inhibitory effects of affected neurons and thereby inhibiting pathological firing (Vercueil *et al.*, 1998). However, research showed that continuous stimulation by an open-loop approach caused the neurons to remain unresponsive due to putative habituation effects (Lüttjohann & van Luijtelaar, 2013). With a combination of techniques presented in this study, machines such as the RNS system (responsive cortical stimulation) could process EEG recordings and treatment of epilepsy (Sun & Morrell, 2014). With the optimization demonstrated in this study, lower rates of false positives were achieved as in comparison to already in use stimulation devices, which focused on partial onset epilepsy (Sun & Morrell, 2014). Using the improvements presented in this study would allow the development of new treatment options including an automated

SWD prediction triggering an on-demand kind of bio-feedback/stimulation in a closed loop BCI system with a significant reduction of wrong predictions

7. Summary

Childhood absence epilepsy affects about 10 - 17% of all children diagnosed with epilepsy and is not well treatable in some cases. The discovery of a precursor event shortly prior to seizures resulted in a putative treatment option via closed-loop deep brain stimulation. The previously developed online prediction system produces high sensitivity rates but also high number of false positive predictions. This thesis aimed to optimize seizure prediction by determining the impact of the number of EEG traces and the location of EEG recording sites in thalamus and in cortex on the prediction performance of the above mentioned algorithm in WAG/Rij rats and GAERS. Results indicate that three EEG traces in the somatosensory cortex close to the onset zone of SWD produce the highest predictability. Significant differences were observed between the two rodent models GAERS and WAG/Rij. Furthermore, comparison of wavelet spectra between true and false positive detections revealed significant differences, which were used to train a machine learning algorithm termed random forest. The random forest was able to accurately (73.1%) classify the true and false predictions. Comparing results from the random forest with an unbiased deep learning algorithm approach that was trained with raw EEG data also showed above chance classification success of interictal and preictal EEG epochs. However, the combination of the Maksimenko-algorithm and a further classification by the random forest showed a more promising predictability of SWDs. Both machine learning approaches bridged the problem of low sample size with the application of random undersampling. However, it cannot be excluded that an increase in training samples of somatosensory measured EEG and thereby an increase in trained interictal and preictal events would show higher classification success for unseen data. Consequently, the combination of a frequency modified Maksimenko-algorithm and random forest classification would allow a promising first view on human CAE seizure prediction.

8. Bibliography

Abhang PA, Gawali BW & Mehrotra SC (2016). Chapter 2 - technological basics of EEG recording and operation of apparatus. In *Introduction to EEG- and Speech-Based Emotion Recognition*, ed. Abhang PA, Gawali BW & Mehrotra SCBT-I to ES-BER, pp. 19–50. Academic Press. Available at: http://www.sciencedirect.com/science/article/pii/B9780128044902000026.

Aker RG, Yananli HR, Gurbanova AA, Ozkaynakçi AE, Ateş N, van Luijtelaar G & Onat FY (2006). Amygdala kindling in the WAG/Rij rat model of absence epilepsy. *Epilepsia* **47,** 33–40.

Akman O, Demiralp T, Ates N & Onat FY (2010). Electroencephalographic differences between WAG/Rij and GAERS rat models of absence epilepsy. *Epilepsy Res* **89,** 185–193.

Aljalal M, Ibrahim S, Djemal R & Ko W (2020). Comprehensive review on brain-controlled mobile robots and robotic arms based on electroencephalography signals. *Intell Serv Robot*; DOI: 10.1007/s11370-020-00328-5.

Alwosheel A, van Cranenburgh S & Chorus CG (2018). Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *J Choice Model* **28,** 167–182.

Asadi-Pooya AA & Homayoun M (2020). Idiopathic (genetic) generalized epilepsies with absences: clinical and electrographic characteristics and seizure outcome. *Neurol Sci*; DOI: 10.1007/s10072-020-04490-7.

Avoli M (1995). Feline generalized penicillin epilepsy. *Ital J Neurol Sci* **16,** 79–82.

Ayodele TO (2010). Types of machine learning algorithms. In *New Advances in Machine Learning*, ed. Zhang Y, p. 366. InTech. Available at: www.intechopen.com.

Bajaj P (2020). What is reinforcement learning. *geeksforgeeks.org*. Available at: https://www.geeksforgeeks.org/what-is-reinforcement-learning/.

Bashivan P, Rish I & Heisig S (2016). Mental State Recognition via Wearable EEG.

Bengio Y, Simard P & Frasconi P (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Trans neural networks* **5,** 157–166.

Berger H (1929). Über das elektroenkephalogramm des menschen. *Arch Psychiatr Nervenkr* **87,** 527–570.

Berger H (1939). Das Elektroenkephalogramm des Menschen. In *Nova Acta Leopoldina*, 6th edn., pp. 173–309.

Blumenfeld H (2005). Cellular and Network Mechanisms of Spike-Wave Seizures. *Epilepsia* **46,** 21–33.

Bozinovski S (2017). Signal processing robotics using signals generated by a human head: Frompioneering works to EEG-based emulation of digital circuits. *Adv Intell Syst Comput* **540,** 449–462.

Butler S (1872). *Erewhon*.

Callenbach P, Bouma P, Brouwera F, Geerts A, Arts W, Stroink H, Peeters E, van Donselaar C, Doudewijn Peters A & Brouwer O (2009). Long-term outcome of childhood absence epilepsy: Dutch study of epilepsy in childhood. *Epilepsy Res* **83,** 249–256.

Caplan R, Siddarth P, Stahl L, Lanphier E, Vona P, Gurbani S, Koh S, Sankar R & Shields WD (2008). Childhood absence epilepsy: Behavioral, cognitive, and linguistic comorbidities. *Epilepsia* **49,** 1838–1846.

Cash SS, Halgren E, Dehghani N, Rossetti AO, Thesen T, Wang C, Devinsky O, Kuzniecky R, Doyle W, Madsen JR, Bromfield E, Eross L, Halász P, Karmos G, Csercsa R, Wittner L & Ulbert I (2009). The human K-complex represents an isolated cortical down-state. *Science* **324,** 1084–1087.

Castrounis A (2019). *AI for people and business*.

Cerminara C, Coniglio A, El-Malhany N, Casarelli L & Curatolo P (2012). Two epileptic syndromes, one brain: Childhood absence epilepsy and benign childhood epilepsy with centrotemporal spikes. *Seizure* **21,** 70–74.

Chan Phooi M'ng J & Mehralizadeh M (2016). Forecasting east asian indices futures via a novel hybrid of wavelet-PCA denoising and artificial neural network models. *PLoS One* **11,** e0156338.

Chen G (2014). Automatic EEG seizure detection using dual-tree complex wavelet-Fourier features. *Expert Syst Appl* **41,** 2391–2394.

Christensen J, Kjeldsen MJ, Andersen H & Friis ML (2005). Gender differences in epilepsy. *Epilepsia* **46,** 956–960.

Cleary DR, Ozpinar A, Raslan AM & Ko AL (2015). Deep brain stimulation for psychiatric disorders: where we are now. *Neurosurg Focus* **38,** 1–24.

Coenen AML (1995). Neuronal activities underlying the electroencephalogram and evoked potentials of sleeping and waking: Implications for information processing. *Neurosci Biobehav Rev* **19,** 447–463.

Cogswell M, Ahmed F, Girshick R, Zitnick L & Batra D (2016). Reducing overfitting in deep networks by decorrelating representations. *4th Int Conf Learn Represent ICLR 2016 - Conf Track Proc*1–12.

Coulter D, JR H & Prince D (1989). Specific petit mal anticonvulsants reduce calcium currents in thalamic neurons. *Neurosci Lett* **98,** 74–78.

Crunelli V & Leresche N (2002). Childhood absence epilepsy: Genes, channels, neurons and networks. *Nat Rev Neurosci* **3,** 371–382.

Cunha J (2019). Zarontin (Ethosuximide). *rxlist.com*. Available at: https://www.rxlist.com/zarontin-side-effects-drug-center.htm#overview.

Davison A (2008). Cambridge series in statistical and probabilistic mathematics. *Cambridge Univ Press*.

Depaulis A, David O & Charpier S (2016). The genetic absence epilepsy rat from Strasbourg as a model to decipher the neuronal and network mechanisms of generalized idiopathic epilepsies. *J Neurosci Methods* **260,** 159–174.

Depaulis A & van Luijtelaar G (2006). Genetic models of absence epilepsy in the rat. In *Models of Seizure and Epilepsy*, pp. 233–248.

Drinkenburg W, Coenen A, Vossen J & Van Luijtelaar E (1991). Spike-wave discharges and sleep-wake states in rats with absence epilepsy. *Epilepsy Res* **9,** 218–224.

Drinkenburg W, Schuurmans M, Coenen A, Vossen J & Van Luijtelaar E (2003). Ictal stimulus processing during spike-wave discharges in genetic epileptic rats. *Behav Brain Res* **143,** 141–146.

Durón RM, Medina MT, Martínez-Juárez IE, Bailey JN, Perez-Gosiengfiao KT, Ramos-Ramírez R, López-Ruiz M, Alonso ME, Ortega RHCC, Pascual-Castroviejo I, Machado-Salas J, Mija L & Delgado-Escueta A V. (2005). Seizures of idiopathic generalized epilepsies. *Epilepsia* **46,** 34–47.

Ende M, Louis AK, Maass P & Mayer-Kress G (1998). EEG signal analysis by continuous wavelet transform techniques. In *Nonlinear Analysis of Physiological Data*, pp. 213–219. Springer Berlin Heidelberg, Berlin, Heidelberg.

Engel J (2001). A proposed diagnostic scheme for people with epileptic seizures and with epilepsy: Report of the ILAE task force on classification and terminology. *Epilepsia* **42,** 796–803.

Feddersen B, Vercueil L, Noachtar S, David O, Depaulis A & Deransart C (2007). Controlling seizures is not controlling epilepsy: A parametric study of deep brain stimulation for epilepsy. *Neurobiol Dis* **27,** 292–300.

Fischer-Elfert H (2005). *Abseits von Ma'at Fallstudien zu Außenseitern im alten Ägypten*. Ergon.

Fisher RS, Cross JH, French JA, Higurashi N, Hirsch E, Jansen FE, Lagae L, Moshé SL, Peltola J, Roulet Perez E, Scheffer IE & Zuberi SM (2017). Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. *Epilepsia* **58,** 522–530.

Fu H, Niu Z, Zhang C, Ma J & Chen J (2016). Visual cortex inspired CNN model for feature construction in text analysis. *Front Comput Neurosci* **10,** 64. Available at: https://www.frontiersin.org/article/10.3389/fncom.2016.00064.

Geitgey A (2016). Convolutional neural networks. *medium.com*. Available at: https://medium.com/@ageitgey/machine-learning-is-fun-part-3-deep-learning-and-convolutional-neural-networks-f40359318721.

Glaba P, Latka M, Krause MJ, Kuryło M, Jernajczyk W, Walas W & West BJ (2020). Changes in Interictal Pretreatment and Posttreatment EEG in Childhood Absence Epilepsy . *Front Neurosci* **14,** 196. Available at: https://www.frontiersin.org/article/10.3389/fnins.2020.00196.

Glauser TA, Cnaan A, Shinnar S, Hirtz DG, Dlugos D, Masur D, Clark PO, Capparelli E V. & Adamson PC (2010). Ethosuximide, valproic acid, and lamotrigine in childhood absence epilepsy. *N Engl J Med* **362,** 790–799.

Gloor P (1968). Generalized cortico-reticular epilepsies some considerations on the pathophysiology of generalized bilaterally synchronous spike and wave discharge. *Epilepsia* **9,** 249–263.

Gloor P (1986). Consciousness as a Neurological Concept in Epileptology: A Critical Review. *Epilepsia* **27,** 14–26.

Gloor P, Quesney LF & Zumstein H (1977). Pathophysiology of generalized penicillin epilepsy in the cat: the role of cortical and subcortical structures. II. Topical application of penicillin to the cerebral cortex and to subcortical structures. *Electroencephalogr Clin Neurophysiol* **43,** 79–94.

Gonzalo-Ruiz A & Lieberman AR (1995). GABAergic projections from the thalamic reticular nucleus to the anteroventral and anterodorsal thalamic nuclei of the rat. *J Chem Neuroanat* **9,** 165–174.

Goodfellow I, Bengio Y & Courville A (2011). Deep learning. *ICML2013 Tutor*1–800.

Gören MZ & Onat F (2007). Ethosuximide: From bench to bedside. *CNS Drug Rev* **13,** 224–239.

Grosso S, Galimberti D, Vezzosi P, Farnetani M, Di Bartolo RM, Bazzotti S, Morgese G & Balestri P (2005). Childhood absence epilepsy: Evolution and prognostic factors. *Epilepsia* **46,** 1796–1801.

Guillery RW, Feig SL & Lozsádi DA (1998). Paying attention to the thalamic reticular nucleus. *Trends Neurosci* **21,** 28–32.

Gupta D, Ossenblok P & van Luijtelaar G (2011). Space-time network connectivity and cortical activations preceding spike wave discharges in human absence epilepsy: A MEG study. *Med Biol Eng Comput* **49,** 555–565.

Halpern C, Hurtig H, Jaggi J, Grossman M, Won M & Baltuch G (2007). Deep brain stimulation in neurologic disorders. *Park Relat Disord* **13,** 1–16.

Hariz M (2014). Parkinsonism and related disorders deep brain stimulation: new techniques. *Park Realt Disord* **20,** S192–S196.

Haykin S (2009). *Neural networks and learning Machines*.

He H & Ma Y (2013). *Imbalanced learning: Foundations, algorithms, and applications*. Wiley-IEEE.

He Y, Eguren D, Azorín JM, Grossman RG, Luu TP & Contreras-Vidal JL (2018). Brain-machine interfaces for controlling lower-limb powered robotic systems. *J Neural Eng* **15,** 21004.

Van Hese P, Martens J-P, Waterschoot L, Boon P & Lemahieu I (2009). Automatic detection of spike and wave discharges in the EEG of genetic absence epilepsy rats from Strasbourg. *IEEE Trans Biomed Eng* **56,** 706–717.

Hinton G & Salakhutdinov R (2006). Reducing the dimensionality of data with neural networks. *Science (80-)* **313,** 504–507.

Hochreiter S & Schmidhuber J (1997). Long Short-Term Memory. *Neural Comput* **9,** 1735–1780.

Hopfield JJ (1988). Artificial neural networks. *IEEE Circuits Devices Mag* **4,** 3–10.

Huang L & van Luijtelaar G (2013). The effects of responsive and scheduled subicular high frequency stimulation in the intra- hippocampal kainic acid seizure model. *Epilepsy Res* **106,** 326–337.

Hughes SW & Crunelli V (2005). Thalamic mechanisms of EEG alpha rhythms and their pathological implications. *Neurosci a Rev J bringing Neurobiol Neurol psychiatry* **11,** 357–372.

Inouye T, Matsumoto Y, Shinosaki K, Iyama A & Toi S (1994). Increases in the power spectral slope of background electroencephalogram just prior to asymmetric spike and wave complexes in epileptic patients. *Neurosci Lett* **173,** 197–200.

Jandó G, Siegel RM, Horváth Z & Buzsáki G (1993). Pattern recognition of the electroencephalogram by artificial neural networks. *Electroencephalogr Clin Neurophysiol* **86,** 100–109.

Jarre G, Altwegg-Boussac T, Williams MS, Studer F, Chipaux M, David O, Charpier S, Depaulis A, Mahon S & Guillemain I (2017). Building up absence seizures in the somatosensory cortex: From network to cellular epileptogenic processes. *Cereb Cortex* **27,** 4607–4623.

Johnson EL (2019). Seizures and epilepsy. *Med Clin North Am* **103,** 309–324.

Kessler SK & McGinnis E (2019). A practical guide to treatment of childhood absence epilepsy. *Pediatr Drugs* **21,** 15–24.

Kim SY (2009). Effects of sample size on robustness and prediction accuracy of a prognostic gene signature. *BMC Bioinformatics* **10,** 4–7.

Kingma DP & Ba JL (2015). Adam: A method for stochastic optimization. *3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc*1–15.

Kossoff E (2004). Effect of an external responsive neurostimulator on seizures and electrographic discharges during subdural electrode monitoring. *Epilepsia* **45,** 1560–1567.

Kostopoulos GK (2000). Spike-and-wave discharges of absence seizures as a transformation of sleep spindles: the continuing development of a hypothesis. *Clin Neurophysiol* **111,** 27–38.

LeCun (1998). LeNet-5. *yann.lecun.com*. Available at: http://yann.lecun.com/exdb/lenet/.

Lecun Y, Bottou L, Bengio Y & Ha P (1998). LeNet. *Proc IEEE*1–46.

Li Y, Yu Z, Chen Y, Yang C, Li Y, Allen Li X & Li B (2020). Automatic seizure detection using fully convolutional nested LSTM. *Int J Neural Syst*; DOI: 10.1142/S0129065720500197.

Liu M, Wang M, Wang J & Li D (2013). Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. *Sensors Actuators B Chem* **177,** 970–980.

Loiseau P, Duché B & Pédespan J (1995). Absence epilepsies. *Epilepsia*1182–1186.

van Luijtelaar G, Hramov AE, Sitnikova E & Koronovskii A (2010). Spike-wave discharges in WAG/Rij rats are preceded by delta and theta precursor activity in cortex and thalamus. *Clin Neurophysiol* **122,** 687–695.

van Luijtelaar G, Lüttjohann A, Makarov V V, Maksimenko VA, Koronovskii AA & Hramov AE (2016). Methods of automated absence seizure detection, interference by stimulation, and possibilities for prediction in genetic absence models. *J Neurosci Methods* **260,** 144–158.

van Luijtelaar G, Mishra A, Edelbroek P, Coman D, Frankenmolen N, Schaapsmeerders P, Covolato G, Danielson N, Niermann H, Janeczko K, Kiemeneij A, Burinov J, Bashyal C, Coquillette M, Lüttjohann A, Hyder F, Blumenfeld H & Rijn C (2013). Anti-epileptogenesis: Electrophysiology, diffusion tensor imaging and behavior in a genetic absence model. *Neurobiol Dis*; DOI: 10.1016/j.nbd.2013.08.013.

van Luijtelaar G & Sitnikova E (2006). Global and focal aspects of absence epilepsy: The contribution of genetic models. *Neurosci Biobehav Rev* **30,** 983–1003.

van Luijtelaar G, Sitnikova E & Lüttjohann A (2011). On the origin and suddenness of absences in genetic absence models. *Clin EEG Neurosci* **42,** 83–97.

van Luijtelaar G, Zobeiri M, Lüttjohann A & Depaulis A (2017). Experimental treatment options in absence epilepsy. *Curr Pharm Des* **23,** 1–16.

Lüttjohann A & van Luijtelaar G (2012). The dynamics of cortico-thalamo-cortical interactions at the transition from pre-ictal to ictal LFPs in absence epilepsy. *Neurobiol Dis* **47,** 49–60.

Lüttjohann A & van Luijtelaar G (2013). Thalamic stimulation in absence epilepsy. *Epilepsy Res* **106,** 136–145.

Lüttjohann A & van Luijtelaar G (2015). Dynamics of networks during absence seizure's on- and offset in rodents and man. *Front Physiol* **6,** 1–17.

Lüttjohann A, Lüttjohann A, Makarov V V., Maksimenko VA, Koronovskii AA & Hramov AE (2016). Methods of automated absence seizure detection, interference by stimulation, and possibilities for prediction in genetic absence models. *J Neurosci Methods* **260,** 144–158.

Lüttjohann A, Schoffelen JM & van Luijtelaar G (2013). Peri-ictal network dynamics of spike-wave discharges: Phase and spectral characteristics. *Exp Neurol* **239,** 235–247.

Lüttjohann A, Zhang S, de Peijper R & van Luijtelaar G (2011). Electrical stimulation of the epileptic focus in absence epileptic WAG/Rij rats: Assessment of local and network excitability. *Neuroscience* **188,** 125–134.

Maksimenko VA, van Heukelum S, Makarov V V., Kelderhuis J, Lüttjohann A, Koronovskii AA, Hramov AE & van Luijtelaar G (2017). Absence seizure control by a brain computer interface. *Sci Rep* **7,** 2487.

Manzouri F, Heller S, Dümpelmann M, Woias P & Schulze-Bonhage A (2018). A comparison of machine learning classifiers for energy-efficient implementation of seizure detection. *Front Syst Neurosci* **12,** 1–11.

Matricardi S, Verrotti A, Chiarelli F, Cerminara C & Curatolo P (2014). Current advances in childhood absence epilepsy. *Pediatr Neurol* **50,** 205–212.

McCafferty C et al. (2018). Cortical drive and thalamic feed-forward inhibition control thalamic output synchrony during absence seizures. *Nat Neurosci* **21,** 744–756.

Mcculloch WS & Pitts W (1990). A logical calculus nervous activity. *Bull Math Biol* **52,** 99–115.

McCulloch WS & Pitts W (1943). A logical calculus of the ideas immanent in nervous activity. *Bull Math Biol*.

Meeren H, van Luijtelaar G, Lopes da Silva F & Coenen AM (2005). Evolving concepts on the pathophysiology of absence seizures. *Arch Neurol* **62,** 371.

Meeren H, Pijn JPM, van Luijtelaar EL, Coenen AM & Lopes da Silva F (2002). Cortical focus drives widespread corticothalamic networks during spontaneous absence seizures in rats. *J Neurosci* **22,** 1480–1495.

Meeren H, Veening JG, Möderscheim TAE, Coenen AML & van Luijtelaar G (2009). Thalamic lesions in a genetic rat model of absence epilepsy: dissociation between spike-wave discharges and sleep spindles. *Exp Neurol* **217,** 25–37.

Mirowski PW, LeCun Y, Madhavan D & Kuzniecky R (2008). Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. *Proc 2008 IEEE Work Mach Learn Signal Process MLSP 2008*244–249.

Missinglink.ai (2020). Backpropagation in neural networks: Process, example & code. *Missinglink.ai*. Available at: https://missinglink.ai/guides/neural-network-concepts/backpropagation-neural-networks-process-examples-code-minus-math/.

Motamedi-Fakhr S, Moshrefi-Torbati M, Hill M, Hill CM & White PR (2014). Signal processing techniques applied to human sleep EEG signals—A review. *Biomed Signal Process Control* **10,** 21–33.

Murakami S & Okada Y (2006). Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. *J Physiol* **575,** 925–936.

Murphy DP, Bai O, Gorgey AS, Fox J, Lovegreen WT, Burkhardt BW, Atri R, Marquez JS, Li Q & Fei DY (2017). Electroencephalogram-based brain-computer interface and lower-limb prosthesis control: A case study. *Front Neurol* **8,** 1–8.

Nguyen-Michel VH, Ourabah Z, Sebban C, Lavallard-Rousseau M-C & Adam C (2009). Idiopathic generalised epilepsies. *Rev Neurol (Paris)* **165,** 924–932.

Nielsen M (2019). *Neural networks and deep learning*.

Nitze I, Schulthess U & Asche H (2012). Comparison of machine learning algorithms random forest, artificial neuronal network and support vector machine to maximum likelihood for supervised crop type classification. *Proc 4th Conf Geogr Object-Based Image Anal – GEOBIA 2012*35–40.

Oda S, Kishi K, Yang J, Chen S, Yokofujita J, Igarashi H, Tanihata S & Kuroda M (2004). Thalamocortical projection from the ventral posteromedial nucleus sends its collaterals to layer I of the primary somatosensory cortex in rat. *Neurosci Lett* **367,** 394–398.

Osorio I (2005). Automated seizure abatement in humans using electrical stimulation. *Ann Neurol* **57,** 258–268.

Osterhagen L, Breteler M & van Luijtelaar G (2010). Does arousal interfere with operant conditioning of spike-wave discharges in genetic epileptic rats? *Epilepsy Res* **90,** 75–82.

Ovchinnikov A, Lüttjohann A, Hramov AE & Van Luijtelaar G (2010). An algorithm for real-time detection of spike-wave discharges in rodents. *J Neurosci Methods* **194,** 172–178.

Panayiotopoulos CP (1999). Typical absence seizures and their treatment. *Arch Dis Child* **81,** 351–355.

Pinault D (2003). Cellular interactions in the rat somatosensory thalamocortical system during normal epileptic 5-9 Hz oscillations. *J Physiol* **552,** 881–905.

Pinault D (2004). The thalamic reticular nucleus: structure, function and concept. *Brain Res Brain Res Rev* **46,** 1–31.

Pinault D & Deschênes M (1998). Projection and innervation patterns of individual thalamic reticular axons in the thalamus of the adult rat: a three-dimensional, graphic, and morphometric analysis. *J Comp Neurol* **391,** 180–203.

Pinault D, Vergnes M & Marescaux C (2001). Medium-voltage 5-9-Hz oscillations give rise to spike-and-wave discharges in a genetic model of absence epilepsy: In vivo dual extracellular recording of thalamic relay and reticular neurons. *Neuroscience* **105,** 181–201.

Pitkänen A, Schwartzkroin PA & Mosché SL (2006). *Models of seizures and epilepsy*. Elsevier Academic Press, San Diego.

Polack P-O, Guillemain I, Hu E, Deransart C, Depaulis A & Charpier S (2007). Deep layer somatosensory cortical neurons initiate spike-and-wave discharges in a genetic model of absence seizures. *J Neurosci* **27,** 6590–6599.

Polson NG, Willard BT & Heidari M (2015). A statistical theory of deep learning via proximal splitting. *arxiv.org*1–28.

Powell KL, Cain SM, Ng C, Sirdesai S, David LS, Kyi M, Garcia E, Tyson JR, Reid CA, Bahlo M, Foote SJ, Snutch TP, O'Brien TJ & Brien TJO (2009). A Cav3.2 T-type calcium channel point mutation has splice-variant-specific effects on function and segregates with seizure expression in a polygenic rat model of absence epilepsy. *J Neurosci* **29,** 371–380.

Puka K, Ferro MA, Camfield CS, Levin SD, Smith M Lou, Wiebe S, Zou G, Anderson KK & Speechley KN (2020). Trajectories of quality of life 10 years following a diagnosis of epilepsy in childhood. *Epilepsia*1–11.

Purves D, Augustine G & Fitzpatrick D (2001). *Neuroscience 2nd edition*. Sunderland. Available at: https://www.ncbi.nlm.nih.gov/books/NBK11154/.

Richard CD, Tanenbaum A, Audit B, Arneodo A, Khalil A & Frankel WN (2015). SWD reader: A wavelet-based algorithm using spectral phase to characterize spike-wave morphological variation in genetic models of absence epilepsy. *J Neurosci Methods* **242,** 127–140.

Rosenow F, Wyllie E, Kotagal P, Mascha E, Wolgamuth BR & Hamer H (1998). Staring spells in children: Descriptive features distinguishing epileptic and nonepileptic events. *J Pediatr* **133,** 660–663.

Sadleir LG, Farrell K, Smith S, Connolly MB & Scheffer IE (2006). Electroclinical features of absence seizures in childhood absence epilepsy. *Neurology* **67,** 413–418.

Sainath TN, Vinyals O, Senior A, York N & Sak H (2015). Convolutional, long short-term memory, fully connected deep neural networks. *IEEE Int Conf Acoust Speech Signal Process - Proc* **2015-Augus,** 4580–4584.

Schmidhuber J (2015). Deep Learning in neural networks: An overview. *Neural Networks* **61,** 85–117.

Schridde U & van Luijtelaar G (2004). The influence of strain and housing on two types of spike-wave discharges in rats. *Genes Brain behav* **3,** 1–7.

Scikit-learn (2020). Cross validation. *scikit-learn Dev*. Available at: https://scikit-learn.org/stable/modules/cross_validation.html.

Sharma A (2020). Decision tree vs. random forest – which algorithm should you use? *Anal Vidhya*. Available at: https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/.

Sherman S & Guillery R (2005). *Exploring the thalamus and its role in cortical function, second edition*. The MIT Press, Cambridge, Massachusetts.

Sherman SM & Guillery RW (2002). The role of the thalamus in the flow of information to the cortex. *Philos Trans R Soc Lond B Biol Sci* **357,** 1695–1708.

Shinnar R, Shinnar S, Liu C, Masur D, Dlugos D, Cnaan A, Hu F, Hirtz DG, Clark P, Weiss EF & Glauser TA (2017). Pretreatment behavior and subsequent medication effects in childhood absence epilepsy. *Neurology* **89,** 1698–1706.

Siddiqui MK, Morales-Menendez R, Huang X & Hussain N (2020). A review of epileptic seizure detection using machine learning classifiers. *Brain informatics* **7,** 5.

Sironia V (2011). Origin and evolution of deep brain stimulation. *Front Integr Neurosci* **5,** 42.

Sitnikova E, Hramov AE, Koronovsky AA, Luijtelaar G Van & van Luijtelaar G (2009). Sleep spindles and spike-wave discharges in EEG: Their generic features, similarities and distinctions disclosed with Fourier transform and continuous wavelet analysis. *J Neurosci Methods* **180,** 304–316.

Sitnikova E & van Luijtelaar G (2007). Electroencephalographic characterization of spike-wave discharges in cortex and thalamus in WAG/Rij rats. *Epilepsia* **48,** 2296–2311.

Sitnikova E & van Luijtelaar G (2009). Electroencephalographic precursors of spike-wave discharges in a genetic rat model of absence epilepsy: Power spectrum and coherence EEG analyses. *Epilepsy Res* **84,** 159–171.

Smyk MK & van Luijtelaar G (2020). Circadian rhythms and epilepsy: A suitable case for absence epilepsy. *Front Neurol* **11,** 1–13.

Sorokin JM, Paz JT & Huguenard JR (2016). Absence seizure susceptibility correlates with pre-ictal β oscillations. *J Physiol Paris* **110,** 372–381.

Srivastava N, Hinton G, Sutskever A, Ilya K & Salakhutdinov R (2014). Dropout: A simple way to prevent neural networks from overfitting. *J Mach Learn Res* **15,** 1929–1958.

Steriade M (2003). *Neuronal substrates of sleep and epilepsy*. Cambridge University Press, Cambridge. Available at: https://www.cambridge.org/core/books/neuronal-substrates-of-sleep-and-epilepsy/66F935DCB7ADA9F1D6E540018AE23F27.

Sun FT & Morrell MJ (2014). The rns system: responsive cortical stimulation for the treatment of refractory partial epilepsy. *Expert Rev Med Devices* **11,** 563–572.

Sysoeva M V, Lüttjohann A, van Luijtelaar G & Sysoev I V (2016). Dynamics of directional coupling underlying spike-wave discharges. *Neuroscience* **314,** 75–89.

Terlau J, Yang JW, Khastkhodaei Z, Seidenbecher T, Luhmann HJ, Pape HC & Lüttjohann A (2020). Spike-wave discharges in absence epilepsy: segregation of electrographic components reveals distinct pathways of seizure activity. *J Physiol* **598,** 2397–2414.

Vanhoecke J & Hariz M (2017). Deep brain stimulation for disorders of consciousness: Systematic review of cases and ethics. *Brain Stimul*; DOI: 10.1016/j.brs.2017.08.006.

Vercueil L, Benazzouz A, Deransart C, Bressand K, Marescaux C, Depaulis A & Benabid AL (1998). High-frequency stimulation of the sub-thalamic nucleus suppresses absence seizures in the rat: Comparison with neurotoxic lesions. *Epilepsy Res* **31,** 39–46.

Waaler PE, Blom BH, Skeidsvoll H & Mykletun A (2000). Prevalence, classification, and severity of epilepsy in children in Western Norway. *Epilepsia* **41,** 802–810.

Wang Y, Cao J, Lai X & Hu D (2019). Epileptic State Classification for Seizure Prediction with Wavelet Packet Features and Random Forest. *2019 Chinese Control Decis Conf*3983–3987.

Wang Z, Yan W & Oates T (2017). Time series classification from scratch with deep neural networks: A strong baseline. *Proc Int Jt Conf Neural Networks* **2017-May,** 1578–1585.

Westerhuis F, van Schaijk W & van Luijtelaar G (1996). Automatic detection of spike-wave discharges in the cortical EEG of rats. In: Proceedings of the measuring behavior. *Int Work Methods Tech Behav Res*.

Wheless JW, Clarke DF, Arzimanoglou A & Carpenter D (2007). Treatment of pediatric epilepsy: European expert opinion, 2007. *Epileptic Disord* **9,** 353–412.

Wirrell E, Camfield C, Camfield P, Gordon K & Dooley J (1997). Long-term prognosis of typical childhood absence epilepsy. *Neurology* **49,** 1187.

Zangiabadi N, Ladino LDi, Sina F, Orozco-Hernández JP, Carter A & Téllez-Zenteno JF (2019). Deep brain stimulation and drug-resistant epilepsy: A review of the literature. *Front Neurol* **10,** 1–18.

Zhang ZW & Deschênes M (1997). Intracortical axonal projections of lamina VI cells of the primary somatosensory cortex in the rat: a single-cell labeling study. *J Neurosci* **17,** 6365–6379.

Zheng TW, O'Brien TJ, Morris MJ, Reid CA, Jovanovska V, O'Brien P, Van Raay L, Gandrathi AK & Pinault D (2012). Rhythmic neuronal activity in S2 somatosensory and insular cortices contribute to the initiation of absence-related spike-and-wave discharges. *Epilepsia* **53,** 1948–1958.

9. Appendix

Appendix 1a: List of combinations used in predictions based on three channel.

TYPE	#	AREA 1	AREA 2	AREA 3	COMBINATION	N
3 CHANNEL	1	Ctx 4	Ctx 5	Ctx 6	CCC	15
	2	Ctx 4	Ctx 5	Po	CCT	15
	3	Ctx 4	Ctx 5	AN	CCT	12
	4	Ctx 4	Ctx 5	rRTN	CCT	6
	5	Ctx 4	Ctx 5	cRTN	CCT	12
	6	Ctx 4	Ctx 5	VPM	CCT	11
	7	Ctx 4	Ctx 6	Po	CCT	13
	8	Ctx 4	Ctx 6	AN	CCT	12
	9	Ctx 4	Ctx 6	rRTN	CCT	6
	10	Ctx 4	Ctx 6	cRTN	CCT	12
	11	Ctx 4	Ctx 6	VPM	CCT	11
	12	Ctx 5	Ctx 6	Po	CCT	10
	13	Ctx 5	Ctx 6	AN	CCT	6
	14	Ctx 5	Ctx 6	rRTN	CCT	12
	15	Ctx 5	Ctx 6	cRTN	CCT	9
	16	Ctx 5	Ctx 6	VPM	CCT	5
	17	Ctx 4	Po	AN	CTT	9
	18	Ctx 4	Po	rRTN	CTT	9
	19	Ctx 4	Po	cRTN	CTT	6
	20	Ctx 4	Po	VPM	CTT	5
	21	Ctx 4	AN	rRTN	CTT	8
	22	Ctx 4	AN	cRTN	CTT	13
	23	Ctx 4	AN	VPM	CTT	12
	24	Ctx 4	rRTN	cRTN	CTT	6
	25	Ctx 4	rRTN	VPM	CTT	12
	26	Ctx 4	cRTN	VPM	CTT	11
	27	Ctx 5	Po	AN	CTT	10
	28	Ctx 5	Po	rRTN	CTT	6
	29	Ctx 5	Po	cRTN	CTT	12
	30	Ctx 5	Po	VPM	CTT	9
	31	Ctx 5	AN	rRTN	CTT	5
	32	Ctx 5	AN	cRTN	CTT	9
	33	Ctx 5	AN	VPM	CTT	9
	34	Ctx 5	rRTN	cRTN	CTT	6
	35	Ctx 5	rRTN	VPM	CTT	5
	36	Ctx 5	cRTN	VPM	CTT	8
	37	Ctx 6	Po	AN	CTT	10
	38	Ctx 6	Po	rRTN	CTT	6
	39	Ctx 6	Po	cRTN	CTT	12
	40	Ctx 6	Po	VPM	CTT	9
	41	Ctx 6	AN	rRTN	CTT	5
	42	Ctx 6	AN	cRTN	CTT	9
	43	Ctx 6	AN	VPM	CTT	9

	44	Ctx 6	rRTN	cRTN	CTT	6
	45	Ctx 6	rRTN	VPM	CTT	5
	46	Ctx 6	cRTN	VPM	CTT	8
	47	Po	AN	rRTN	TTT	5
	48	Po	AN	cRTN	TTT	9
	49	Po	AN	VPM	TTT	7
	50	Po	rRTN	cRTN	TTT	6
	51	Po	rRTN	VPM	TTT	5
	52	Po	cRTN	VPM	TTT	8
	53	AN	rRTN	cRTN	TTT	5
	54	AN	rRTN	VPM	TTT	4
	55	AN	cRTN	VPM	TTT	6
	56	rRTN	cRTN	VPM	TTT	5
	57	Ctx 5a	Ctx 5b	Ctx 6	MCCC	6

Appendix 1b: List of combination within prediction based on two channel.

TYPE	#	AREA 1	AREA 2	COMBINATION	N
2 CHANNEL	1	Ctx 4	Ctx 5	CC	16
	2	Ctx 4	Ctx 6	CC	16
	3	Ctx 5	Ctx 6	CC	16
	4	Ctx 4	VPM	CT	11
	5	Ctx 4	AN	CT	13
	6	Ctx 4	Po	CT	14
	7	Ctx 4	cRTN	CT	13
	8	Ctx 4	rRTN	CT	6
	9	Ctx 5	VPM	CT	11
	10	Ctx 5	AN	CT	13
	11	Ctx 5	Po	CT	14
	12	Ctx 5	cRTN	CT	13
	13	Ctx 5	rRTN	CT	6
	14	Ctx 6	VPM	CT	11
	15	Ctx 6	AN	CT	13
	16	Ctx 6	Po	CT	14
	17	Ctx 6	cRTN	CT	13
	18	Ctx 6	rRTN	CT	6
	19	VPM	AN	TT	9
	20	VPM	Po	TT	9
	21	VPM	cRTN	TT	8
	22	VPM	rRTN	TT	5
	23	AN	Po	TT	11
	24	AN	cRTN	TT	10
	25	AN	rRTN	TT	5
	26	Po	cRTN	TT	13
	27	Po	rRTN	TT	6
	28	cRTN	rRTN	TT	6

Appendix 2: Historical verification of electrode position. X resembles properly assembled electrode. ctx4: somatosensory cortex layer 4, ctx5: somatosensory cortex layer 5, ctx6: somatosensory cortex layer 6, ATN: anterior thalamic nucleus, Po: posterior thalamic nucleus, VPM: ventral-postero-medial thalamic nucleus, cRTN: caudal reticular thalamic nucleus, rRTN: rostral reticular thalamic nucleus (Lüttjohann *et al.*, 2013)

	ctx4	ctx5	ctx6	ATN	Po	VPM	cRTN	rRTN
Rat1	X	x	x	x	x	x	x	X
Rat2	x	x	x	x	x		X	
Rat3	x	x	x	x	x		x	X
Rat4	x	x	x		x	x	X	
Rat5	x	x	x		x	x	x	X
Rat6	x	x	x	x	x	x	x	X
Rat7	x	x	x	x	x	x	x	X
Rat8	x	x	x	x		x		
Rat9	x	x	x	x	x	x	x	X
Rat10	x	x	x	x		X		
Rat40	x	x	x	x	x	X	x	X
Rat13	x	x	x	x	x	x		X
Rat14	x	x	x		x		x	X
Rat15	X	x	x	x	x	x	x	
Rat16	x	x	x	x	x		x	
Rat17	X	x	x		x	x	x	

Appendix 3: Video explaining preprocessing of raw EEG data for deep learning classification and data modulation by deep learning layers.

10. Curriculum vitae

11. Acknowledgements