Università degli Studi di Cagliari

PHD DEGREE

Industrial Engineering

Cycle XXX

Machine Learning Techniques for Detection of Nocturnal Epileptic Seizures from Electroencephalographic Signals

Scientific Disciplinary Sector
ING-IND/31

PhD Student:	Barbara Pisano
Coordinator of the PhD Programme	Prof. Francesco Aymerich
Supervisor	Prof. Alessandra Fanni

Final exam. Academic Year 2016 – 2017
Thesis defence: March 2018 Session

Università degli Studi di Cagliari

PHD DEGREE

Industrial Engineering

Cycle XXX

Machine Learning Techniques for Detection of Nocturnal Epileptic Seizures from Electroencephalographic Signals

Scientific Disciplinary Sector
ING-IND/31

PhD Student:	Barbara Pisano
Coordinator of the PhD Programme	Prof. Francesco Aymerich
Supervisor	Prof. Alessandra Fanni

Final exam. Academic Year 2016 – 2017
Thesis defence: March 2018 Session

Barbara Pisano gratefully acknowledges Sardinia Regional Government for the financial support of her PhD scholarship (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013 - Axis IV Human Resources, Objective l.3, Line of Activity l.3.1.)".

ACKNOWLEDGEMENTS

A few words to thank all the people who made this work possible.

First and foremost, I would like to express my sincere gratitude to my supervisor Professor Alessandra Fanni for the continuous support of my PhD study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me to grow as a research scientist.

I would also like to thank my colleagues of the Electrotechnics research group of the University of Cagliari, for the stimulating discussions on machine learning algorithms and the pleasant conversations during our downtime.

Thanks to Dr. Monica Puligheddu for generously sharing time and knowledge on nocturnal epilepsy. Her assistance has been fundamental for the development of this work.

I would also like to express my special thanks to Professor António Dourado, Professor César Teixeira and the colleagues of the Centre for Informatics and Systems of the University of Coimbra, who gave me a warm welcome to their beautiful country and provided me support during my internship in their group. Their collaboration has been essential for this thesis.

Last but not least, thanks go to my parents, my brothers and my sister for their encouragement, support and patience over the years. Thanks also to my friends, for always being there for me in all the happy and stressful moments.

ABSTRACT

Epilepsy is one of the major neurological disorders that affects more than 50 million people around the world; it is characterized by unpredictable seizures due to an abnormal electrical activity in the brain.

In this thesis nocturnal epilepsy has been investigated. In particular, Nocturnal Frontal Lobe Epilepsy (NFLE), that is a form of epilepsy in which seizures occur predominantly during sleep with symptoms including nocturnal awakenings, dystonic and tonic postures and clonic limb convulsions.

The electroencephalographic (EEG) signals, which record the electrical activity of the brain, are used by neurologists to diagnose epilepsy. However, in almost 50% of NFLE cases, the EEG does not show abnormality during seizures, making the neurologists work to identify the epileptic events very difficult, thereby requiring the support of video recording to verify the epileptic events, with a subsequent time-consuming procedure.

In literature few scientific contributions address the classification of nocturnal epileptic seizures. In this thesis, the automatic systems, both customized for single patient and generalized have been developed to find the best nocturnal epileptic seizure detection system from EEG signals. The combination of feature extraction and selection methods, associated to classification models based on Self Organizing Map (SOM), have been investigated following the classical machine learning approach.

The ability of SOM to represent data from a high-dimensional space in a low-dimensional space, preserving the topological properties of the original space, has been exploited to identify nocturnal epileptic seizures and track the temporal projection of the EEG signals on the map. The proposed methods allow the definition of maps capable of presenting meaningful information on the actual brain state, revealing the mapping potential of clustering data coming from seizure and non-seizure states.

The results obtained show that the patient-specific system achieves better performance than a patient-independent system. Moreover, comparing the performances with those of a binary classifier, widely used in epileptic seizure detection problems, the Support Vector Machine (SVM), the SOM model achieves good and, for some patients, higher performances.

In particular, the patient-customized system using SOM model, reaches an average value of sensitivity and specificity equal to 82.85% and 89.92%, respectively; whereas the SVM classifier achieved an average sensitivity and specificity equal to 82.11% and 82.85%, respectively, suggesting the use of SOM model as a good alternative for nocturnal epileptic seizure detection.

The discriminating power of SOM and the possibility to follow the temporal sequence of the EEG recordings on the map can provide information on an imminent epileptic seizure, highlighting the possibility to promote therapies aimed at rapid and targeted disarming the seizures.

Contents

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

An epileptic seizure describes a variety of neurological symptoms due to an abnormal, synchronized and prolonged electrical discharge in the brain. It affects approximately 1% of the world's population, without substantial differences of race or nationality. The identification of epileptic seizures is of decisive importance for the improvement of the quality of life and for the socio-psychological development of the patients.

The Nocturnal Frontal Lobe Epilepsy (NFLE) is a form of epilepsy in which seizures occur predominantly during sleep. The disease usually begins during childhood or adolescence, and then extends into adult life. Patients present different symptoms that include sudden nocturnal awakenings, placement of the arms or legs in bizarre postures, dystonic and tonic postures and clonic limb convulsions, leading to chronically disrupted sleep and daytime sleepiness.

The most powerful cure for NFLE is the assumption of carbamazepine. Nonetheless, not all patients have a successful response to the therapy. For this poor reaction to the therapy, several studies have been conducted on refractory epilepsy during these decades. The ability to early and accurately detect seizures can promote therapies aimed at rapid and targeted treatment of them.

Among the diagnostic measurements used by doctors to recognize the epilepsy, the electroencephalographic (EEG) recordings are widely used in machine learning techniques to identify the epileptic seizures. To date, several algorithms have been proposed with the aim to automatically detect different kinds of epileptic seizures. The combination of feature extraction and selection methods associated to classification models were studied with good results. Nonetheless, very few contributions are proposed to detect nocturnal epileptic seizures.

In the present thesis, different systems based on the Self Organizing Map (SOM) have been investigated for detection of nocturnal epileptic seizures using EEG signals. The goal is to define a patient-dependent system capable of early detecting nocturnal seizures of the subjects affected by nocturnal epilepsy with the aim to support the neurologists with seizures identification usually done manually by means of video and EEG recordings. Moreover, the early detection system would also be able to prevent accidents and minimize the occurrence of injuries [1].

The mapping method allows us to project multi-dimensional data into a 2D-dimensional space, maintaining the topological properties of the original space. This advantage has been used to display and identify the seizure (SZ) and no-seizure (NS) regions.

Additionally, the analysis of the trajectory has provided information on an eventual impending of the seizure events.

1.1 Outline of the thesis

The thesis is organized as follows:

- Chapter 2 reports an overview of the brain function and a classification of seizure types, focusing on the description of the Nocturnal Frontal Lobe Epilepsy.
- Chapter 3 describes the methodologies used by doctors to diagnose the epilepsy disease. In particular, the electroencephalography (EEG), electromyography (EMG), electrooculography (EOG), electrocardiography (ECG) and magnetoencephalography (MEG) are presented.
- Chapter 4 reports the sleep structure and the rules for scoring the different stages of the sleep, for both sleep macrostructure and sleep microstructure.
- In Chapter 5 an overview of the machine learning methods typically used in epileptic seizure detection systems are described based on the steps of the machine learning process.
- In Chapter 6 a summary of the Machine Learning approaches reported in literature for the detection of epileptic seizures, are described.
- Chapter 7 presents the patient-specific systems proposed for the detection of nocturnal epileptic seizures.
- Chapter 8 reports the patient--independent systems developed for the detection of Nocturnal Frontal Lobe Epileptic seizures.
- In Chapter 9 the conclusions are drawn.

2 EPILEPSY

Epilepsy is one of the major neurological disorders and, according to the World Health Organization (WHO), it affects more than 50 million people around the world.

Epilepsy is defined by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE) as the occurring of a sporadic epileptic seizure associated to an enduring disturbance of the brain sufficient to cause other seizures, where, the epileptic seizure is described as a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain [2].

In 2014, the ILAE extended the definition of epilepsy in an Official Report [3], establishing a practical guideline for clinical diagnosis. The disease is indicated as epilepsy if one of the following conditions is verified:

1. at least two unprovoked (or reflex) seizures occurring, separated by more than 24 h;
2. one unprovoked (or reflex) seizure and a probability of further seizures similar to the general recurrence risk (at least 60%) after two unprovoked seizures, occurring over the next 10 years;
3. diagnosis of an epilepsy syndrome.

Additionally, the medical diagnosis of epilepsy can be considered as concluded for *individuals who had an age-dependent epilepsy syndrome but are now past the applicable age or those who have remained seizure-free for the last 10 years, with no seizure medicines for the last 5 years* [3].

2.1 Brain and Communication

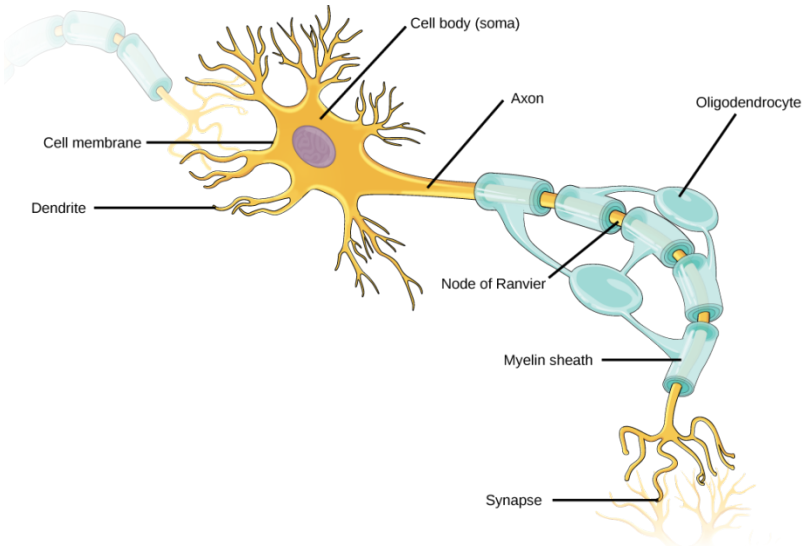


Fig. 2-1 Representation of the neuron [4].

A typical healthy human brain contains around 100 billion of *neurons*, more precisely, from a study carried out by Brazilian Universities [5], they are on average 86 billion. Each neuron may be connected to up to 10,000 other neurons, passing signals to each other via as many as 1,000 trillion *synaptic connections*.

The fundamental functional unit of the nervous system is devoted to receive and transmit information. Its morphological structure, described in Fig. 2-1, is made up of a cellular body, called also *soma*, from which *dendrites* and a single *axon* are branched.

Dendrites are characterized by an arborized structure capable to receive the nerve pulse, whereas the axon, which is a long cellular filament, conveys information away from the cell body.

Axons can be coated by a lipid substance, the myelin, which plays the role to protect and electrically isolate from their environment, increasing the velocity of information transmission.

Every neuron at rest is polarized, due to a difference concentration of ions across its membrane. The presence of a lower positive ions concentration, mostly given by potassium ions (K^+), inside the membrane, compared to the extracellular positive ions concentration, given in the majority part by sodium ions (Na^+), leads to a trans-membrane electrical potential, typically around -60 mV.

An alteration of the permeability membrane determines a perturbation of the electrical balance, permitting the temporary flux of ions inside the membrane and a consequent variation of trans-membrane electric potential. If the variation is higher enough to

overcome the threshold, an action potential, also called *nerve pulse*, is generated and transmitted along the axon.

The communication between the nerve cells takes place via synapses. Depending on the connection between the pre-synaptic and post-synaptic terminals, synapses can be defined as *chemical* or *electrical*.

In the case of chemical synapses, the action potential of pre-synaptic cell causes the release of a neurotransmitter, which spreads into the extracellular space (normally the distance is around 0.03 μm) by altering the cellular potential of the postsynaptic cell; dendrites receive the neurotransmitters and converts them back into an electrical signal.

The signal then travels through the neuron, to be converted back into a chemical signal, when it gets to neighboring neurons [6].

Turning to the electric synapses, the information is instead directly transmitted to the post-synaptic cell by direct contact, without the need of a chemical mediator. Cells are joined by ionic channels with very low resistance, so as to allow the passage of the ionic currents induced by the action potential.

During this communication, the changes of postsynaptic conductance vary the likelihood that an action potential will be produced in the postsynaptic cell. The postsynaptic potential (PSP) is called excitatory (EPSP) if it increases the probability that a PSP happens; whereas, it is called inhibitory (IPSP) if the probability decreases.

The mechanism that determines the excitement or inhibition of a postsynaptic cell depends on the type of channel that is coupled to the receptor, and on the concentration of permeant ions inside and outside the cell. In fact, the only factor that distinguishes postsynaptic excitation from inhibition is the reversal potential of the PSP in relation to the threshold voltage for generating action potentials in the postsynaptic cell [7].

In Fig. 2-2 an example of excitation and inhibition post-synaptic potentials is shown. In the first window (see Fig. 2-2 A), the case of EPSP is reported, a neurotransmitter and the receptor represented by the glutamate and the acetylcholine (ACh) respectively, determines the opening of ion channels and the subsequently flux of Na^+ and K^+ across the postsynaptic membrane, inducing a depolarization of the post-synaptic membrane potential surmounting the action potential threshold (in this example equal to -40 mV) and the reversal potential (E_{rev}), i.e., the potential at which the action of a given neurotransmitter causes no net current flow.

In Fig. 2-2 B and C, examples of IPSPs are illustrated. The synaptic activation of GABA-dependent ion channels, that are selectively permeable to anion Cl^-, establishes a variation of the post-synaptic potential. Fig. 2-2 B represents the case of E_{rev} -70 mV, assuming that the post-synaptic neuron has a resting potential of -60 mV, a flow of anions is generated, determining a hyperpolarization of the cell. In Fig. 2-2 C, instead, the case of E_{rev} equal to -

50mV is considered, determining a flow of Cl^- out of the cell and a consequent depolarization. In both cases the cell's membrane potential is more negative than the action potential threshold, reducing the probability that the post-synaptic cell will fire an action potential.

Fig. 2-2-D resumes the rule for the inhibition and excitation of the post-synaptic potential: the EPSP assumes a reversal potential higher than the action potential threshold, whereas, the IPSP has a reversal potential more negative than threshold.

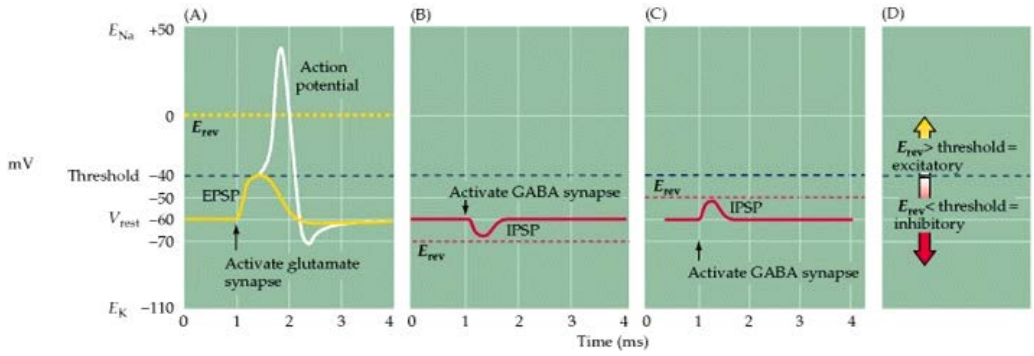


Fig. 2-2 Excitation and inhibition of postsynaptic potential [7].

2.2 Classification of Seizure Types

Neurons receive inputs from both excitatory and inhibitory synapses, whose resulting effect could be excitatory or inhibiting, preventing in the latter case the forwarding of information.

The balance between the two mechanisms determines the regular neuronal communication. Disrupting these behaviors that inhibit firing or promoting the excitation process can lead to seizures. The epileptic seizure is given by a hyper-excitability, due to a synchronism of depolarization of the neurons that generates a propagation of the discharge and the consequent behavioral changes. Generally, while a normal depolarization lasts from 10 to 16 ms, in a seizure the range increases between 100–200 ms [8].

This hyper-polarization depends on a malfunction of ionic channels that establishes a prolonged depolarization current. The hindrance to the repolarization potential causes the tonic phase behavior. In this phase, all the muscles stiffen, air being forced past the vocal cords causes a cry or groan and the person loses consciousness.

The consecutive phase, the *clonic phase*, expects large inhibitory potentials with the alternating of the depolarization potentials. In this stage, it happens that arms and usually legs begin to jerk rapidly and rhythmically bending and relaxing at the elbows, hips and knees with a duration of few minutes.

According to ILAE classification [9], that drew up the updated classification rules to identify the different kinds of seizures, seizures are distinguished in subcategories. This classification originates from the need to conform seizures and help doctors in diagnosis and the choice of therapies.

The basic classification, described in Fig. 2-3, is a simple version of the major categories of the seizures, distinguished on the basis of their origin in the brain. The seizures are divided into those that start focally, meaning involving circuits (networks) in one hemisphere of the brain, and those that engage networks in both sides of the brain at the onset. In case the onset may be missed or obscured, the seizure is of unknown onset category.

The type of seizure onset is important because it affects choice of seizure medication, possibilities for epilepsy surgery, outlook, and possible causes.

The updated classification distinguishes the focal seizures based on the level of awareness and the motor, and other symptoms may occur.

The level of awareness is of practical importance, as it is one of the main factors affecting a person's safety during a seizure. A focal aware seizure is when the person is aware of self and of environment during the seizure, even if immobile; whereas, the focal impaired awareness seizure is when the awareness is compromised at any time during the seizure. Furthermore, the focal seizures are categorized based on the involvement of movements.

The generalized onset seizures can be motor or non-motor (absence): the generalized tonic-clonic seizure category, describes seizures with stiffening (tonic) and jerking (clonic), whereas the category of generalized absence seizures involves brief changes in awareness, staring, and in some cases automatic or repeated movements like *lipsmacking*.

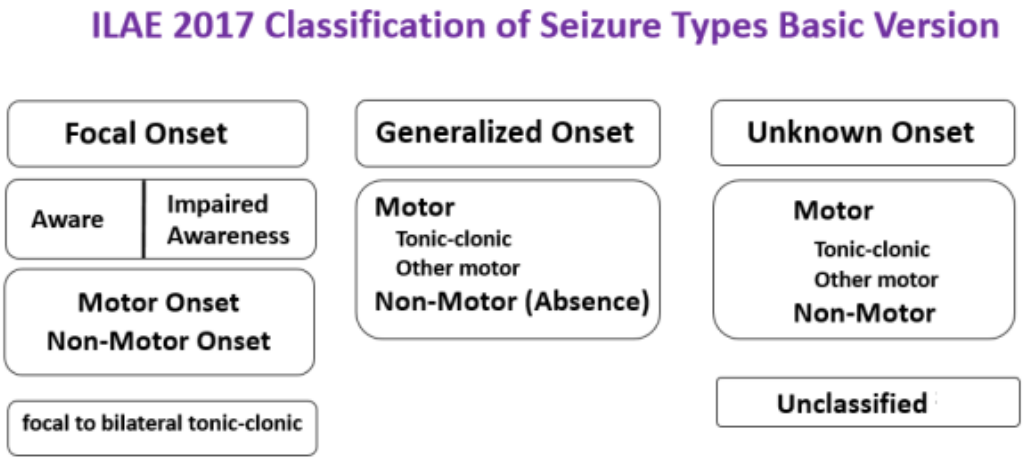


Fig. 2-3 Basic classification of seizure types, according to ILAE 2017 [9].

In order to provide more information about the seizures types for each category previously described, ILAE reported an extended version of the classification of seizure types. The expanded version, as shown in Fig. 2-4, keeps the framework of the basic classification, but adds more seizure types.

As an example, the focal motor onset seizure types include automatisms, atonic, clonic, epileptic spasm, hyperkinetic, myoclonic and tonic seizures, whereas the focal non-motor onset seizures include autonomic, behavior arrest, cognitive, emotional, and sensory seizures.

An example of *focal impaired awareness cognitive seizure* is a seizure that happens with sudden inability to understand language followed by impaired awareness and clonic left arm jerks [9].

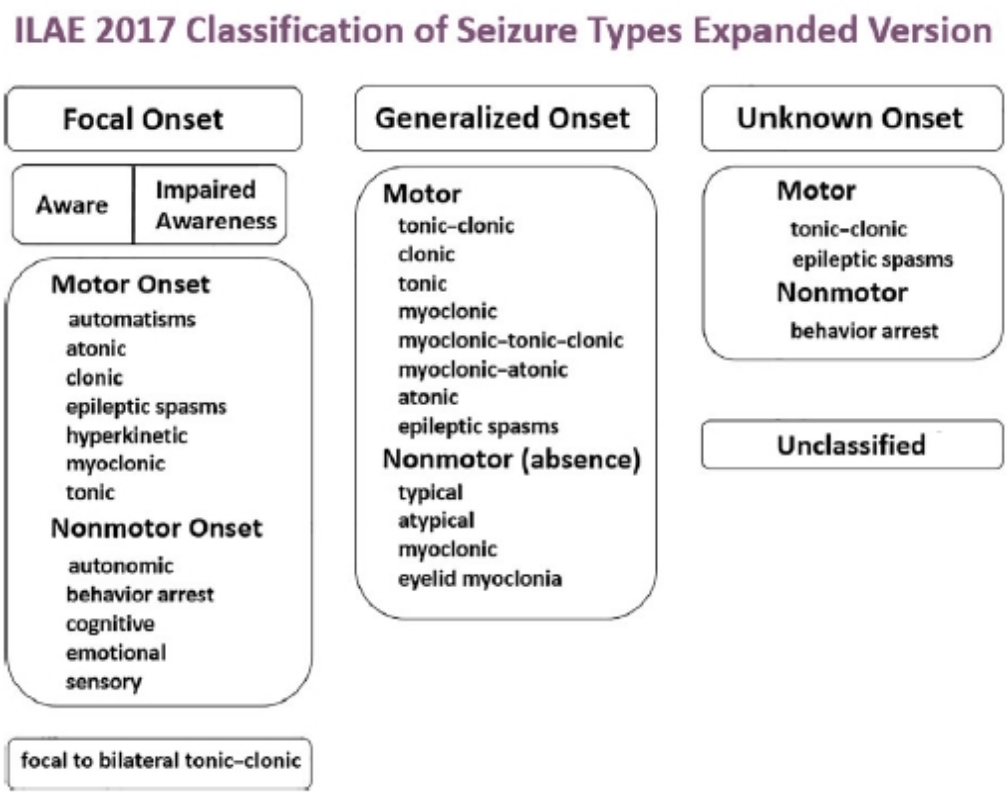


Fig. 2-4 Expanded version of seizure classification types, according to ILAE 2017 [9].

2.3 NFLE

In this thesis, the Nocturnal Frontal Lobe Epilepsy (NFLE) has been investigated. This type of epilepsy belongs to the focal seizure class, it is a form of focal epilepsy with predominant motor expressions, which originates from deep frontal regions (more specifically, from the orbitofrontal or mesial frontal regions [10]), and involves heterogeneous crises predominantly during the Non-Rapid Eye Movement (NREM) phase of sleep.

This form of epilepsy was originally considered as a sleep disorder, because of the similarity to parasomnias and the absence of the typical epileptic distinctive traits on the electroencephalogram (EEG). However, the short duration, the stereotypic features of the episodes and the response to antiepileptic drugs, sometimes at low dosages, suggested an epileptic origin of this syndrome [11].

The study reported in [11] examined the manifestations of 100 patients affected by this form of epilepsy, pointing out three major characteristics: Paroxysmal Arousal (PA), Nocturnal Paroxysmal Dystonia (NPD) and Episodic Nocturnal Wanderings (ENW).

Paroxysmal Arousal is identified by abrupt recurrent arousals from NREM sleep, associated with a stereotypic motor pattern, usually lasting <20 s. In Fig. 2-5 an example of a typical PA and the associated polysomnography is shown. The polysomnography is the set of biosignals of a subject when he is sleeping, among them EEG and electrocardiogram recordings. It will be explain in detail in chapter 4 (§ 4.1). In Fig. 2-5, the episode lasts 19 s, the patient opens his eyes and raises his head, trunk and limbs.

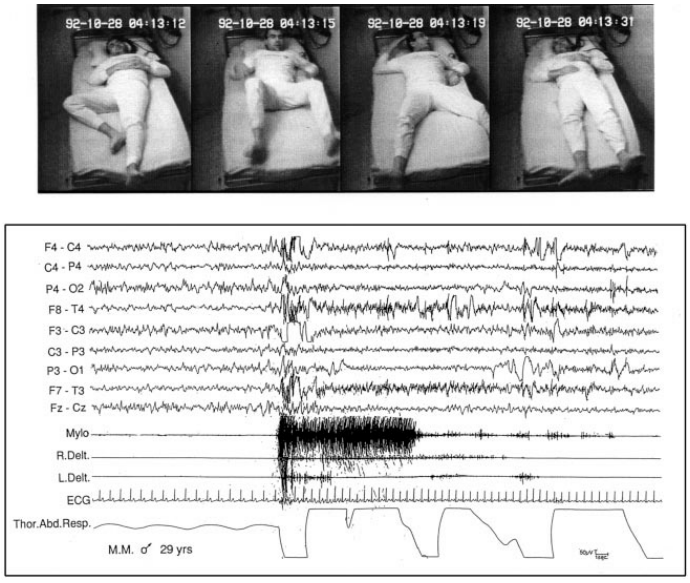


Fig. 2-5 Example of a typical paroxysmal arousal (PA) and the associated polysomnography [11].

Nocturnal Paroxysmal Dystonia is characterized by repeated motor attacks with dystonic-dyskinetic behaviors during the NREM sleep, which last less than 2 min. An example is reported in Fig. 2-6 with the associated polysomnography, where the patient violently rocks her legs, and the seizure lasts 40 s.

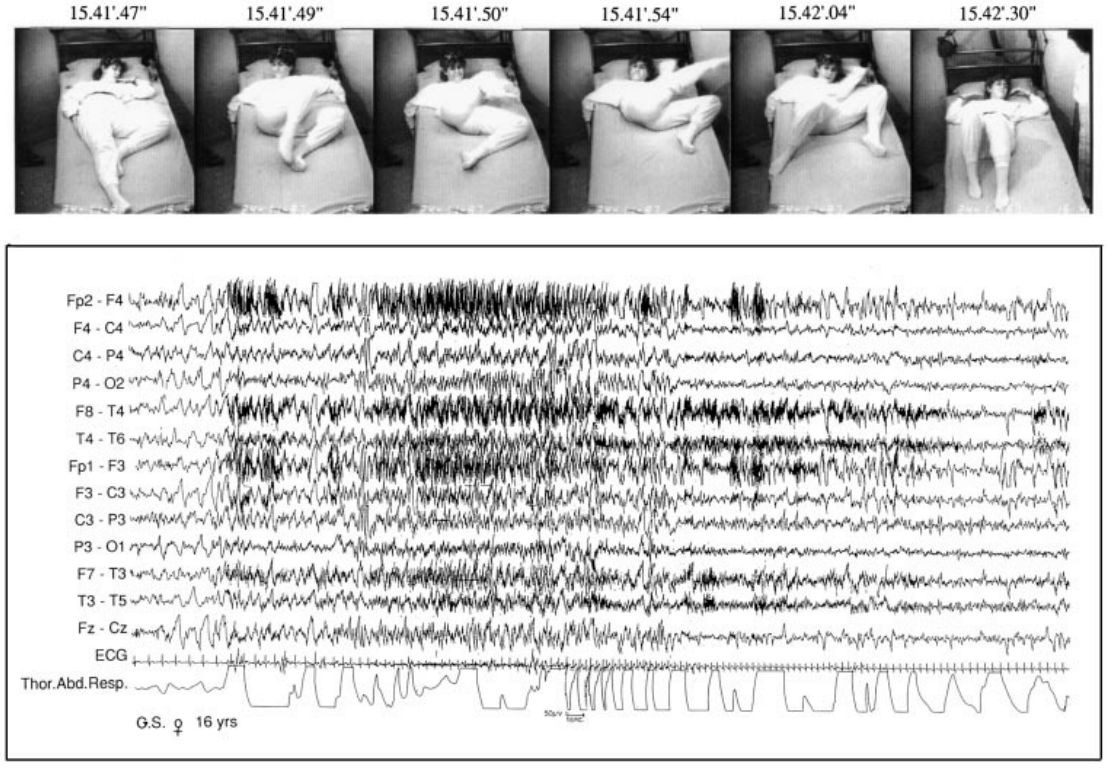
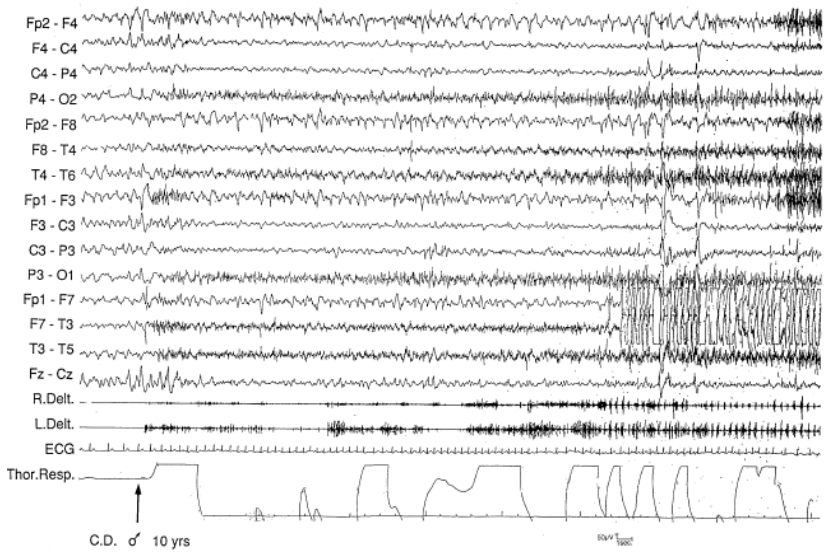


Fig. 2-6 Example of Nocturnal Paroxysmal Dystonia (NPD) and the associated polysomnography [11].

Lastly, Episodic Nocturnal Wanderings, that could include both episodes mentioned before, last up to 3min, present stereotypic paroxysmal ambulation during sleep, accompanied by screaming and bizarre movements, as shown in Fig. 2-7.


23:35
23. 4.1991
23:35
23. 4.1991
23:35
23. 4.1991
23:35
23. 4.1991
23:35
23. 4.1991
23:36
23. 4.1991
23:36
23. 4.1991
23:36
23. 4.1991
23:36
23. 4.1991
23:36
23. 4.1991
23:38
23. 4.1991


Fp2 - F4
F4 - C4
C4 - P4
P4 - O2
Fp2 - F8
F8 - T4
T4 - T6
Fp1 - F3
F3 - C3
C3 - P3
P3 - O1
Fp1 - F7
F7 - T3
T3 - T5
Fz - Cz
R.Delt.
L.Delt.
ECG
Thor.Resp.
C.D. ♂ 10 yrs

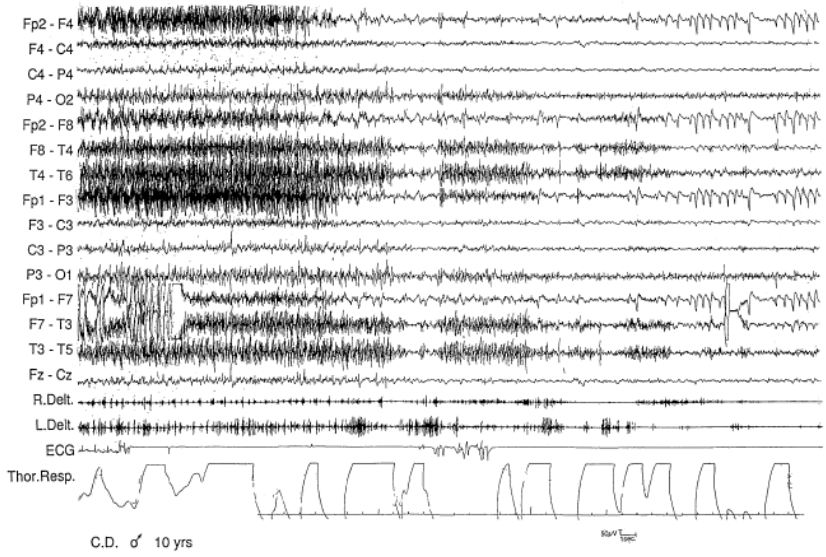


Fig. 2-7 Example of Episodic Nocturnal Wanderings (ENW) and the associated polysomnography [11].

NFLE can also originate from genetic mutations; in that case it is named Autosomal Dominant Nocturnal Frontal Lobe Epilepsy (ADNFLE). The onset typically occurs in childhood with a mean age of 11 years (a median of 8), characterized by hyperkinetic attack, with a short breakouts and sometimes a nonspecific aura of quiver and fear [12].

In 1995, a study on ADNFLE identified one of the gene responsible of idiopathic epilepsies, i.e., epilepsy with no identifiable causes [13]. Successively, other mutations have been found linked to the neuronal nicotinic Acetylcholine Receptor subunits (nAChRs), a ligand-gated channel.

Precisely, the genes identified as responsible for this form of epilepsy are: CHRNA4 (cholinergic receptor, neuronal nicotinic, alpha polypeptide 4); CHRNB2 (cholinergic receptor, neuronal nicotinic, beta polypeptide 2); CHRNA2 (cholinergic receptor, neuronal nicotinic, alpha polypeptide 2). The association of the activation of all these cholinergic receptors to the provenance of NFLE beginning is longer supported given that the cortico-subcortical networks regulate arousal from sleep [14].

The peculiarity of this type of epilepsy is that, excluding the possible artifacts due to body movements, in almost 50% of NFLE cases, the scalp EEG does not show abnormality during the seizures [15], which makes the neurologists work to recognize the epileptic events, only by using EEG very difficult.

Moreover, whilst the most powerful cure for NFLE (ADNFLE) is the assumption of carbamazepine, not all patients have a successful response to the therapy. In fact, as reported in [11], only in 20% of patients, seizures were totally extinguished; whereas, in

48% of cases, a reduction of seizures were partially obtained. This means that, in about one third of cases, epilepsy was resistant to antiepileptic drug treatments.

One of the therapy for refractory epilepsy is an implantable device [16], which provides a electrical stimulation that could be scheduled, as the vagus nerve stimulation therapy [17], or triggered upon detection of a seizure [18], as the system proposed by Osorio et al. [19].

The advantages of the latter approach, are the lower number of treatments applied and the modulation or modification of the therapy in response to physiological changes that may provide more effective and efficient therapy [20], compared to the first approach. However, it requires the identification of the epileptogenic area and a very accurate seizure detection system.

In this thesis the development of an automatic seizure detection system is investigated to support neurologists in nocturnal epileptic seizures identification, and contribute in brain stimulation therapies aimed at rapid and targeted treatment of seizures.

3 DIAGNOSTIC METHODS

The brain activity has been studied since the beginning of the twentieth century, with the aim to detect possible abnormalities. The combination of multiple tests associated to a careful assessment of symptoms and clinical history of the patient allows the doctors to identify the type of epilepsy.

The diagnosis of epilepsy includes the analysis of several physiological signals. The examination of the electrical activity of the brain, i.e., the electroencephalography, is an important test used to highlight patterns of normal or abnormal brain activities. Furthermore, it shows the specific area of the brain where the seizures occur in case of focal epilepsy.

The recordings of muscles activities, i.e., the electromyography, contribute to better understand and recognize tonic-clonic epileptic seizures. In addition to the previous electrophysiological measures, the recording of the electric potential difference between cornea and retina, the electrooculography, is used by neurologists in sleep disorders and nocturnal epilepsy investigations, for staging human sleep.

Electrocardiography is another important tool in diagnosing epilepsy, which reveals the electrical activity of the heart. It provides valuable information about cardiac rhythm helping to identify possible abnormalities due to epileptic events.

Furthermore, the representation of magnetic fields produced by neurons, i.e., magnetoencephalography, is a non invasive neuroimaging technique mostly used by doctors in the pre-surgical evaluation of patients with epilepsy, which allows them to localize the epileptic focus.

In this chapter, the above techniques, principally used in the diagnosis of epilepsy, with a particular focus on the electroencephalography, are presented.

3.1 Electroencephalography

The principal instrument used to analyze the electrical activity of the brain is the Electroencephalography (EEG).

The excitations of neurons establish an electrical potential that is measured by electrodes placed on the scalp or directly on the cortical surface. In the first case, the typical amplitude is around 100 μV, whereas for intracranial EEG the amplitude is about 1-2 mV [21].

The analysis of the cortical pyramidal neurons from the scalp, provides important information regarding the characteristics and location of seizures, contributes to the definition of the type of crisis and seizure syndrome.

The rhythmic activity of the brain is characterized by typical frequencies, as reported in Table 3-1.

- *Delta rhythm* has frequencies lower than 4 Hz. It tends to be the highest one in amplitude and the slowest waves. It is normal as the dominant rhythm in infants up to one year and in stages 3 of NREM sleep. Although delta waves are generally prominent during sleep, there are cases when they are recorded from awake individuals. Delta waves may increase during difficult mental activities requiring concentration and other continuous-attention tasks.
- *Theta rhythm* is in the range from 4 to 8Hz. It is classified as "slow" activity. Theta rhythm is involved in memory and emotional regulation; it has been found to temporarily increase when a person is actively trying to repress a behavioral response. It is perfectly normal in children up to 13 years and in sleep, but it is abnormal in awake adults that could highlight a manifestation of focal subcortical lesions.
- *Alpha rhythm*, usually seen in the occipital regions of the head, has frequencies between 8 and 15 Hz. It appears when closing the eyes and relaxing, and disappears when opening the eyes or alerting by any mechanism (thinking, calculating). It is the major rhythm seen in normal relaxed adults. It is present during most of life especially after the thirteenth year. Alpha rhythms are correlated with inhibitory processes in the brain. Reduced alpha wave occurrence and amplitude, associated to delta-theta activity, determines reduced inhibitory control over behavior.
- *Beta rhythm* is in the range from 15 to 30 Hz. It is the "fast" activity. It is generally regarded as a normal rhythm; it is the dominant rhythm in subjects who are alert, attentive to external stimuli, or anxious or have their eyes open. It may be absent or reduced in areas of cortical damage. Beta rhythms also occur during rapid eye movement (REM) sleep. In this situation, the typical alpha rhythm is suppressed and supplanted by beta waves. This replacement of alpha rhythm is called desynchronization, because it represents a change in the synchronized activity of neural systems in the brain. It is thought that beta waves represents arousal of the cortex to a higher state of alertness and may also be associated with memory retrieval.

Lastly, all frequencies higher than 30 Hz belong to the *gamma band*. Gamma waves are present during mixed sensory processing, such as perceptual tasks that combine hearing and seeing. Gamma waves are also evident during short-term memory matching of recognized sights, sounds or sensations. A reduction in gamma wave activity might be associated with cognitive decline, especially when compared to theta wave activity levels.

Table 3-1 EEG rhythms

Rhythms	Frequencies [Hz]
Delta	<4
Theta	4 – 8
Alpha	8 – 15
Beta	15 – 30
Gamma	>30

3.1.1 Brain Function

EEG records the electrical activity of the cerebral cortex, originates especially from cortical neurons.

Cortical neurons can be classified into two principal categories: *pyramidal* and *non-pyramidal cells*. The pyramidal neurons constitute the predominant part of the cerebral cortex and play a fundamental role in EEG generation. In particular, the arrangement of pyramidal neurons is oriented perpendicular to the surface, they are composed by a conic shaped soma, a long dendrite (named apical dendrite) emerges from the soma and crosses the various cortical layers toward the cortical surface and branches into different terminations. Hence, the pyramidal cells are oriented vertically in the cortex, with their apical dendrites arranged parallel to each other. The measured potentials are originated from averages in the space of dendritic electric fields in the superficial cortical layers.

Conversely, the short axons of non-pyramidal neurons remain into the cortex, ending up on nearby neurons. The orientation between them and in relation to the cortex surface is not well defined, leading a minimal contribute to the measure of surface potential.

Head model

The brain is composed of excitable neural tissues that not only are the location of electric sources but also constitutes part of the volume conductor which includes also the skull and scalp.

The EEG measured on the scalp surface has the great advantage of being a non invasive procedure; however, between electrode and neuronal layers the current goes through different tissues, whose thickness and inhomogeneities influence the scalp surface potentials [22].

Several studies investigated different models of the human head [23]–[25] to estimate the resistivity of the head layers. One of the simplest, that has been successfully considered, is represented by a series of three concentric regions: scalp, skull and brain.

As shown in Fig. 3-1, the model, studied in [26], presents a different average value of the radium and a different resistivity for each layer. The brain has an average radium of $8\ \mathrm{cm}$ with a resistivity of $2.22\ \Omega\mathrm{m}$; whereas the skull is thick $0.5\ \mathrm{cm}$ and has a resistivity around $177\ \Omega\mathrm{m}$; this value is obtained multiplying by a factor of 80 the brain (or scalp) resistivity. Finally, the outer layer, the scalp, is considered with an average radium of $9.2\ \mathrm{cm}$ and a resistivity equal to the brain, i.e., $2.2\ \Omega\mathrm{m}$.

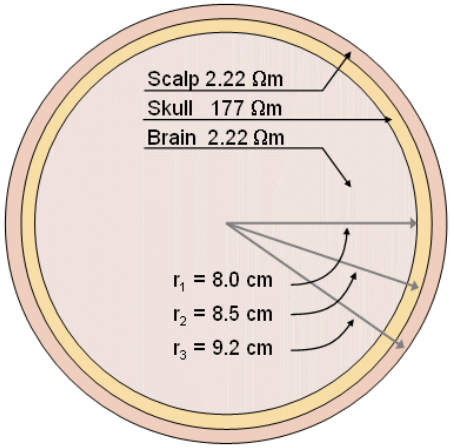


Fig. 3-1 Concentric spherical head model [21]

Despite the simplified model does not consider the anisotropies and inhomogeneities, the results are found to be in accordance with the practical analysis; in [27] the authors obtained a mean resistivity of the cortical bone around $157\ \Omega\mathrm{m}$ and $215\ \Omega\mathrm{m}$ for circumferential and radial directions, respectively.

The resistivity of skull tissue strongly influences the scalp EEG; hence, a more realistic model considers the skull constituted by three bone layers, i.e., the inner and the outer tables (cortical bones) and the spongy layer (*diploe or cancellous bone*), located in the middle, as depicted in Fig. 3-2.

The local resistance of the human skull flaps in this case, considering the direction normal to its local surface, is:

$$R = \frac{\eta_1 d_1 + \eta_2 d_2 + \eta_3 d_3}{A} \tag{3.1}$$

Where μ and d are the resistivity and the thickness respectively, and A is the cross-sectional area.

The three indexes (1, 2, and 3) are referred to the inner, spongy and outer layers, respectively. The inhomogeneities of these three layers are studied in [28], that achieved conductivities ranging from 0.76 ± 0.14 to 11.5 ± 1.8 mS/m.

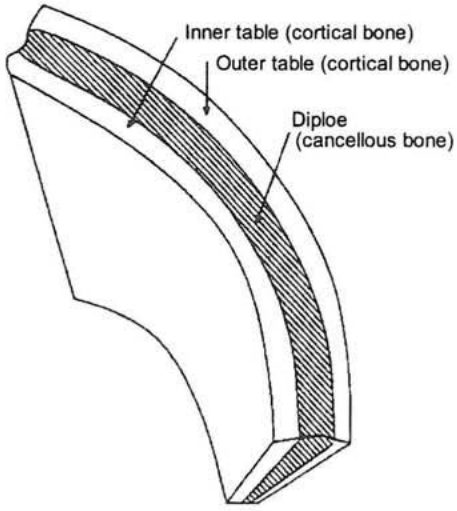


Fig. 3-2 Three layer human skull [27].

3.1.2 EEG electrodes

The measurement of the brain electric activity derives by a proper placement of electrodes. The electrodes, as mentioned above, can be placed on the scalp or on the brain tissue.

The requirements are the stability of the devices and the low susceptibility to external interferences, in order to allow the correct recording of the signals. The non-polarized electrodes are in gold-plated silver, gold, or sintered metal. However, the most commonly used material for EEG electrodes is silver, precisely Ag/AgCl, which has to be chlorinated to prevent the polarization that causes signal distortion. This type of electrodes quickly establishes and maintains consistent and stable electrochemical potentials against biological tissues. Furthermore, they are free from potential allergenic compounds and have excellent long-term electrical stability [29].

Electrodes can be held in a cap, a possible design is Ag/AgCl electrodes fitted with cellulose sponges, and, after the cap is placed on the head, a small amount of a special electrolyte solution is injected into each electrode in order to wet the sponge and allow long recordings (of up to 3 hours).

Electrodes can be also attached to the skin by adhesive and conductive paste, the preparation of the scalp to the EEG measurements being an important step to obtain good

results and few artifacts. According to the guideline [30], the skin-electrode impedance should be less than 5 kΩ, to reduce the noise, artefacts, and other interference.

Other types of electrodes are the *needle electrodes*, very useful in coma patients, i.e. in subjects with minimal or absent response to the pain. They are easy to apply and do not require fixing systems; on the other hand, they smooth the slower frequencies and need precautions regarding the possibility of infections transmission.

Routine recordings also often use *prebuilt cap*, in which the disk electrodes are positioned in a textile shell. It is a quick and comfortable electrodes application system for the patient, but scarcely editable and often a source of artifacts.

3.1.3 EEG placement

The positioning of the EEG electrodes is defined by the International Standard System 10-20, introduced by Dr. Jasper that presented some guidelines in 1949 during the Second International EEG Congress. In the guidelines, recommendations on the position, coverage and designations of electrodes are expressed.

The placement is based on specific ratio of 10% and 20% of the skull perimeters measured in the median and transverse planes. The evaluation is performed taking four reference points: the nasion, the inion, the left and right preauricular, as shown in Fig. 3-3. The nasion is the point located at the top of the nose, at the level with the eyes, whereas the inion is located at the base of the skull. The preauricular points (A1 and A2 in Fig. 3-3B) are felt as depressions at the root of the zygoma, just anterior to the prominence in front of the external opening of the ear (*tragus*). The measure of the imaginary line between nasion and inion, in the median plane, is used to define 5 separated regions, as depicted in Fig. 3-3A. The first area is placed at 10% of the total line and labeled Fp, representing the fronto-polar region. The other sections are located with intervals of 20% of the total measurement and labeled F, C, P, and O, indicating the frontal, central, parietal and occipital areas, respectively. Therefore, the occipital region is located at 10% of the measurement above the inion.

As reported in Fig. 3-3B, the line of the central coronal plane, from left to the right preauricular points intersecting the C vertex in the median line, defines other marks: at 10% of this perimeter measure, over the preauricolar points, the temporal regions, named T, are defined; whereas, at intervals of 20%, starting from the temporal marks other locations are pinpointed, named with odd and even numbers according to the left and right hemisphere respectively.

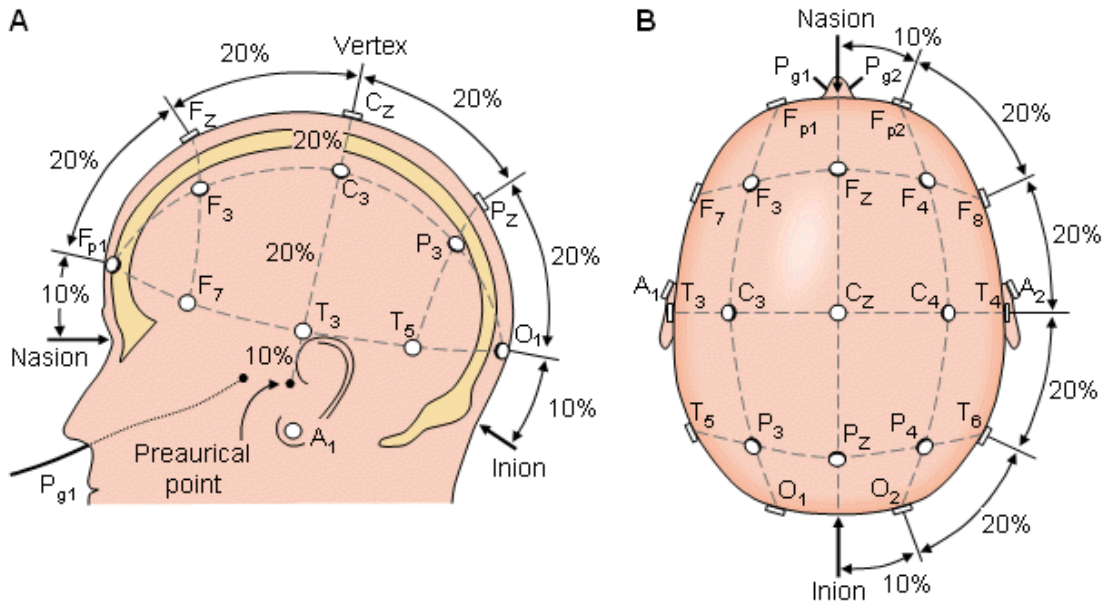


Fig. 3-3 The International 10-20 System seen from (A) left and (B) above the head [21]

The International Standard System 10-20 provides measures of 19 of the 21 electrodes; the remaining two electrodes are placed on the ear lobes and labeled auricular electrodes A1 and A2.

As can be seen from the Fig. 3-3, the electrode positions are named matching the anatomical terms of the cortical area (with the exception of the *C*, that is referred to locations over the central sulcus), with the numbering system based on the hemisphere, that have been added to differentiate between left and right homologous regions. While on the left hemisphere Fp1, F3, F7, C3, T3, P3, T5 and O1 are designated; on right hemisphere, the same symmetrical points are described with even numbers: Fp2, F4, F8, C4, T4, P4, T6, and O2. The electrodes placed in the coronal plane, originally matched with the zero number, are instead named with z letter, becoming Fz, Cz and Pz [31].

Additionally, the American Clinical Neurophysiology Society (ACNS), in the Guideline 1 [32], provided a new system to reveal the areas containing normal and abnormal EEG patterns.

The 10-10 System considers the possibility to add the number of scalp electrodes in order to decrease the likelihood of interpretative errors, particularly true for transient activity [32]. Fig. 3-4 reports the 10-10 System; four electrodes, drawn black with white text, are renamed: the T3, T4, T5 and T6 electrodes of the 10-20 System to T7, T8, P7, and P8, respectively.

The new system allows the identification of 75 positions on the scalp for as many electrodes placed along 11 sagittal and 9 coronal planes.

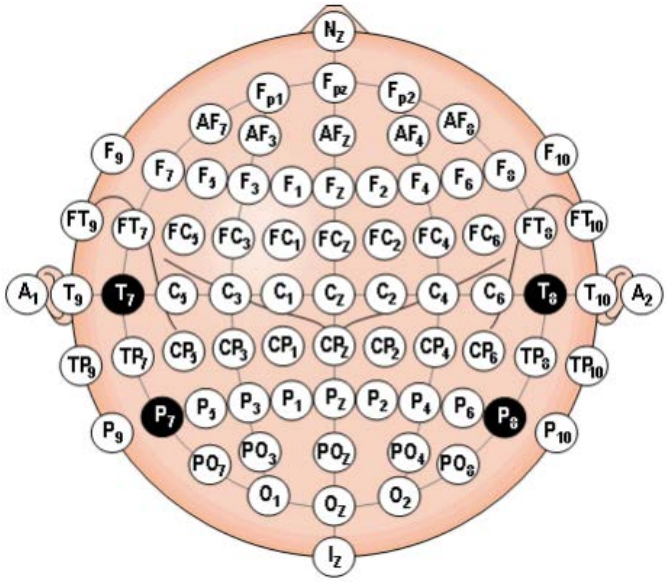


Fig. 3-4 The 10-10 System [21].

The EEG recordings can be acquired based using the unipolar or bipolar derivation. The unipolar derivation is based on a reference potential that could be the A1 or A2 electrodes, placed on the earlobes of the left and right hemispheres, respectively, or to the average potential coming from all electrodes. In the bipolar derivation, instead, both electrodes are placed on active sites in the area of interest and the detected signal corresponds to the difference that emerges between the activities of the two sites.

3.1.4 EEG machine

In addition to the electrode, the EEG recording system is equipped with differential amplifiers, filters, A/D converter and data storage.

In order to be compatible with the other components, EEG signals need to be amplified. The requirement for the amplifiers is that the measure should not be distorted and should not influence the physiological process. The amplifier should provide the best possible separation of signals and interferences.[33]

Differential amplifiers are sensitive to potential difference between electrode pairs and insensitive to the (generally much larger) spatially constant potentials over the scalp (common modes). Modern EEG systems record simultaneously from about 32 to 131 scalp locations.

According to the guideline [32], which establishes the instruction for the EEG measurements, the typical gain setting of the differential amplifiers with common mode

rejection is 7 μV/mm leading to deflections of 3–20 mm for input voltages of 20–140 μV. In digital systems, instead, as computer monitor sizes may vary, a reference marker should always be available in the display area.

Additionally, since scalp EEG signal frequency lies between 1 to 70 Hz, the bandwidth of the recording channels should correspond to this frequency range. The low frequency filter should not be higher than 1 Hz and the high frequency filter should be 70 Hz. If a narrower bandwidth is used, information will be lost and, with a wider bandwidth, noise in the recorded data will contain irrelevant information. High pass analogue EEG filter typically remove substantial power below about 0.5 Hz, depending on application and filter roll-off characteristics. A notch filter is used to remove the power line frequencies (50 Hz or 60 Hz).

Then, A/D conversion involves two processes: the sampling and the quantization. In the sampling process, the analogue signals are sampled at regular time intervals obtaining discrete signals. The number of samples per second defines the sampling rate. The minimum acceptable sampling rate, required to accurately reconstruct the analog signals, according to the sampling theorem, is 2 times greater than the highest frequency of interest. However, most digital EEG systems sample at 256 Hz. Quantization is a process aimed at reducing to a finite number each sample value measured during the sampling process. In a digital system, the resolution of the converter is equal to the minimum difference of amplitude that yields a change in the measure. This is obtained by dividing the voltage range of the A/D converter by 2 raised to the power of the number of bits of the A/D converter.

Currently, on the market, the number of bits used for quantization is typically 16 (with a recommended minimum of 12 bits, discerning 2^12=4,096 value levels) on a signal maximum input EEG of at least 1 mV. Ability to resolve 0.5 μV is recommended.

The guideline provides also information about the time base of the recordings in both cases of analogue and digital printings. The paper speed should be 15 or 30 mm/s with the option of 60 mm/s, considering that, the lower is useful to detect slow-wave abnormalities, whereas the higher, the asynchronicity in bilateral discharges.

3.2 Electromyography

Electromyography (EMG) is a diagnostic procedure to assess the health of muscles and the nerve cells that control them (motor neurons). Motor neurons, activated by the central nervous system, transmit electrical signals that cause muscles to contract. EMG measures this muscle electrical activity that occurs during muscle contraction and relaxation cycles.

Depending on the pathology, the conditions of the muscle can be analyzed in case of rest, during a voluntary contraction, or if necessary, subjecting it to artificial stimulation. The results can reveal nerve dysfunction, muscle dysfunction or problems with nerve-to-muscle signal transmission.

EMG can be considered as a non invasive (surface EMG), or an invasive (intramuscular EMG), method.

In the surface EMG, the electrodes, placed on the skin, are used to measure the speed and strength of signals travelling between two or more points. These surface electrodes are usually made of silver/silver chloride (Ag/AgCl), silver chloride (AgCl), silver (Ag) or gold (Au); however, as well as the scalp EEG, the electrodes made of Ag/AgCl are preferred over the others, since they are almost non-polarizable electrodes [34].

In the intramuscular EMG, needle or fine-wire electrodes are inserted directly into a specific muscle to record its electrical activity.

3.3 Electrooculography

The *electrooculography* (EOG) is the method used to record the eye movements. Through the electrodes located around the eyes, the voltage caused by the motion of the ocular globes is detected.

Specifically, EOG registration reveals the electric potential difference between cornea and retina. Since the eye can be described as a fix dipole where the positive pole is the cornea and the negative is the retina, the eye movements establish a rotation of the dipole, thereby causing a recordable signal from the surface.

The advantages of the EOG are related to the non invasive application of surface electrodes, that does not create discomfort for the patient or limit his field of view. Additionally, it is possible to record eye movements with eyes closed.

However, the disadvantages are associated with the amplitude of the corneo-retinal potential, that changes with the amount of ambient light; this means that illumination has to be kept constant as much as possible [35].

3.4 Electrocardiography

The *electrocardiography* (ECG) is a simple, painless procedure that detects and records the heart electrical activity.

The electrical activity is associated to the development of electrical potentials generated by cell groups (pacemakers) placed in well defined areas of the myocardium. The electrical signals originated in the sinoatrial (SA) node, propagate along the heart muscle through particular fibers (bundle of His) determining the rhythmic contraction.

Under normal conditions, the contraction (systole phase) and the consequent relaxation (diastole phase) of the cardiac muscles are repeated over time, defining the characteristic ECG rhythm reported in Fig. 3-5. The ECG is recorded by surface electrodes placed on prefixed position of limbs and torax in order to have a particular orientation in space. The standard application is the 12-lead ECG, obtained using the wrist and ankle placements.

The first deflection is the P wave associated with right and left atrial depolarization: from the SA node, the signal travels through the right and left atria, causing their contraction of the atria, which moves blood into the ventricles. Normal P lasts less than 120 ms.

The second wave is the QRS complex. Typically, this complex has a series of 3 deflections that reflect the current associated with right and left ventricular depolarization. By convention, the first deflection in the complex, if it is negative, is called a Q wave, the first positive deflection in the complex is called R wave and a negative deflection after an R wave is called S wave. Normally, it lasts less than 120 ms.

The flat line between the end of the P wave and the beginning of the Q wave corresponds to the signal propagation along the atria and the ventricles through a group of cells called the atrioventricular (AV) node. The signal slows down as it passes through the AV node, allowing the ventricles enough time to finish filling with blood.

Finally, the T wave represents the ventricles repolarization, the muscle stops contracting to allow the heart to refill with blood.

The PR interval measures the time from the initial depolarization of the atria to the initial depolarization of the ventricles and reflects a physiological delay in AV conduction imposed by the AV node; the normal range is 120 – 200 ms. Furthermore, the QT interval represents the time in which the ventricles depolarize and repolarize. QT values are heart-rate dependent and can vary from 270 ms at a heart rate of 150 beats/min to 500 ms at a heart rate of 40 beats/min.

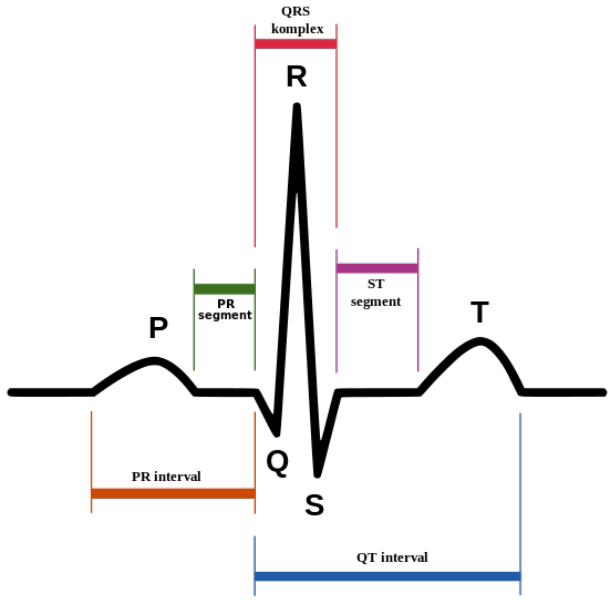


Fig. 3-5 Typical ECG rhythm.

3.5 Magnetoencephalography

The bioelectric activity originating from the brain generates not only electric fields but, of course, also magnetic fields associated with them. The technique, that indirectly analyzes brain activity through the measurement of biomagnetic fields, is the *magnetoencephalography* (MEG). MEG is a direct measure of brain function with good temporal and spatial resolution; it is usually combined with a Magnetic Resonance Imaging (MRI), which is a scanning technique for creating detailed images at high resolution using magnetic fields and radio wave frequencies, to get a good structural perspective. This combination of MEG and MRI is called Magnetic Source Imaging (MSI).

The usual amplitude of magnetic fields created by the brain are extremely small, around 0.1 pT, the recording requires superconducting sensors, as the Superconducting Quantum Interference Device (SQUID), and shielded room to attenuate the external magnetic noise. The SQUID sensors are bathed in a large liquid helium cooling unit at approximately -269°C; due to low impedance at this temperature, the SQUID device can detect and amplify magnetic fields generated by neurons a few centimeters away from the sensors.

Since the decay of magnetic fields as a function of distance is more pronounced than for electric fields, MEG is more sensitive to superficial cortical activity, which should be useful for the study of neocortical epilepsy. In the surgical treatment of epilepsy, the most critical

piece of information is the location of the seizure focus and MEG can substantially contribute to this localization, identifying the eloquent cortex and helping to determine the area to remove during epilepsy surgery [36].

Comparing MEG to scalp EEG, MEG does not require a contact with the head of the patient and detects the magnetic fields induced by intracellular currents without being subjective by anisotropic inhomogeneities of the tissues. On the other hand, MEG is an expensive technique that, normally, has to be paired with an MRI.

4 SLEEP STRUCTURE

The NFLE, investigated in this thesis, is a form of epilepsy that principally occurs during sleep; in particular, most of the crises take place during the NREM phase of sleep.

In this chapter, the sleep structure and the techniques used by neurologists to detect and diagnose the nocturnal epileptic seizures are presented.

Sleep and wake can be considered as two states, continuous among themselves, integrated within a single rhythm: sleep-wake rhythm.

The alternation between sleep and wake is governed by various mechanisms including a homeostatic process that tracks the need for sleep in proportion to the duration of the previous wake state, and a circadian process that oversees the temporal distribution of wakefulness and sleep, with the determination of permissive windows of greater propensity to one or the other state.

In mammals, including man, the center of the circadian control is located in a cerebral structure defined as the suprachiasmatic nucleus (SNC) of the hypothalamus, which acts as a pacemaker, regulating sleep and other biological rhythms.

Under normal conditions, the activity of this nucleus is influenced by the light stimulation from the retina (during the day) and melatonin hypophysar secretion (during the dark period), making the internal biological clock in tune with light/dark cycle of the external environment. Thus, the endogenous circadian component, estimated for the human beings at around 25 hours, adapts to the exogenous circadian component which is instead, around 24 hours.

The circadian rhythm problems originate from a synchronization request between the endogenous component and the exogenous component, to which the pacemaker, represented by the SNC of the hypothalamus, cannot cope. An example is given by the jet lag, whose shift in time and light signals force the body to alter its normal pattern, causing frequent nocturnal awakenings and daily sleepiness; travelers show more difficulty in thinking and doing well. Furthermore, these symptoms can also occur in everyday life, when the circadian rhythm is disrupted by irregular sleep schedule, which could be due to shift work type, that is, when the work programme takes place during normal sleep period.

The analysis of sleep structure allows the neurologists to evaluate and differentiate nocturnal seizures from sleep disorders. In fact, by analyzing sleep macrostructure and microstructure, patients affected by NFLE present an increase of sleep instability parameters and a percentage of the first stages of NREM sleep with a decrease of REM sleep duration.

4.1 Polysomnography

To discover and diagnose sleep disorders, but, mostly, the NFLE/ADNFLE, the polysomnography (PSG), supported by video, is the efficient test used by doctors in sleep medicine. In the video-polysomnography (VPSG), the PSG and the monitoring with a closed circuit camera, help the doctors to identify the starting time of the motor events, which cannot be only detected with the PSG.

The analysis is usually carried out at a Sleep Center, recordings take place in a "sleeping room", where the patient is connected to the equipment and monitored during the night. Additionally, the measurements could also be taken at home by means of a portable device (Holter); the patient, previously prepared by doctors with the application of the sensors, can sleep at home and return the day after for removing the electrodes and consigning the recordings that will be analyzed.

PSG involves a set of analysis that evaluate the bioparameters of a subject when is sleeping. It consists of scalp EEG, the electrooculography, the electromyogram of the submentalis muscle, EMG of the right and left tibialis anterior muscle, the electrocardiography, sensors placed at nares and mouth to evaluate the airflow parameters and sensors to measure the oxygen saturation.

The movements measured by the EMG assist the doctors in the case of motor events, whereas EOG contribute to distinguish the sleep stages, from REM to NREM stages.

An EOG waveform analyzed to distinguish the sleep stages is the **Slow Eye Movements** (SEM). SEM waveforms do not manifest abrupt movements, they are smooth sinusoidal eye movements with amplitudes of 100 μV or more, and 10 s or less in duration.

The evaluation of the electrical functions of the heart, by means of the ECG, has been revealed useful in the detection of epilepsy and sleep microstructure [37]. In [38], a study in patients with focal epilepsies evaluated the arising of tachycardia for the 86% of all seizures.

The EEG recordings are used to reveal the sleep patterns, as well as highlighting the characteristic EEG rhythms.

The following sleep patterns are explained:

- **Vertex sharp wave** is a sharp waveform distinguished from background activities at or near Cz (C3, C4), with an amplitude of 75 μV or greater, and with a frequency of at least 5 Hz but no more than 14 Hz.
- **K complex** is a sharp, negative EEG wave followed by a high-voltage slow wave. The complex duration is at least 0.5 seconds, its peak-to-peak amplitude must be greater than 200 μV.
- **Sleep spindles** are defined as trains of 12–16 Hz waves of 10 μV or greater amplitude, composed of at least six consecutive waves, or a train duration longer

than 0.5 s. The spindle shape is not a requirement for identification as a sleep spindle; however, the mean frequency of a single train of waves can be used as a single descriptor for labeling as sleep spindles.

- ***Delta burst*** is a sleep pattern represented by a sequence of at least two waves in the frequency bandwidth ranging from 0.5 to 4 Hz and with an amplitude 1/3 higher, or more, than the background activity.
- ***Vertex sharp transients*** are EEG potentials of 50–200 ms duration and variable amplitude (up to 250 mV) expressed maximally on derivations at the central vertex areas. Sequences of vertex sharp transients composed of two or more repetitive potentials lasting 2 s or more, often appear at the transition from stage 1 to stage 2 of the NREM sleep.
- ***Polyphasic bursts*** are clusters of high-voltage delta waves, intermixed with theta, alpha or beta rhythms.
- ***K-alpha*** is composed of a K-complex followed immediately by an alpha burst. It has an overall duration of 2 s or more.
- ***EEG arousals*** are sudden frequency shifts toward faster rhythms, that shortly interrupt sleep continuity for interval greater than or equal to 3 s.

4.2 Sleep Macrostructure

According to Rechtschaffen and Kales (R&K) [39] and the supplements proposed by the American Academy of Sleep Medicine [40], it is possible to recognize a macrostructure of the sleep, consisting of two general stages: Rapid Eye Movement (REM) and Non-Rapid Eye Movement (NREM). NREM is, in turn, subdivided into three stages. Each stage is characterized by particular brain activities and EEG frequency content.

Sleep Stage 1 (S1) includes the presence of alpha waves, vertex sharp wave (V-wave) and Slow Eye Movements (SEM). It is characterized by a marked decrease in alpha activity, typical of the physiological wake activity (W), and appearance of theta activity.

The theta activity must exceed 50% of a recording period (usually 20-30 seconds), so as that period be named S1. The first stage of sleep is also characterized by an even higher electromyogram, albeit lower than that of wakefulness, from the absence of rapid ocular movements and the presence of slow ocular movements, especially in the first transition phase from wakefulness to S1.

Sleep Stage 2 (S2) is characterized by K-complex and sleep spindles, with a relatively low amplitude of EEG path and a prevalence of theta rhythms. EMG is further reduced and eye movements are absent. The K complex may be preceded by polyphasic negative-positive waves that must be visually distinguished from background EEG activities.

An example of S2 is shown in Fig. 4-1. The sleep spindles are followed by the K-complex waveform.

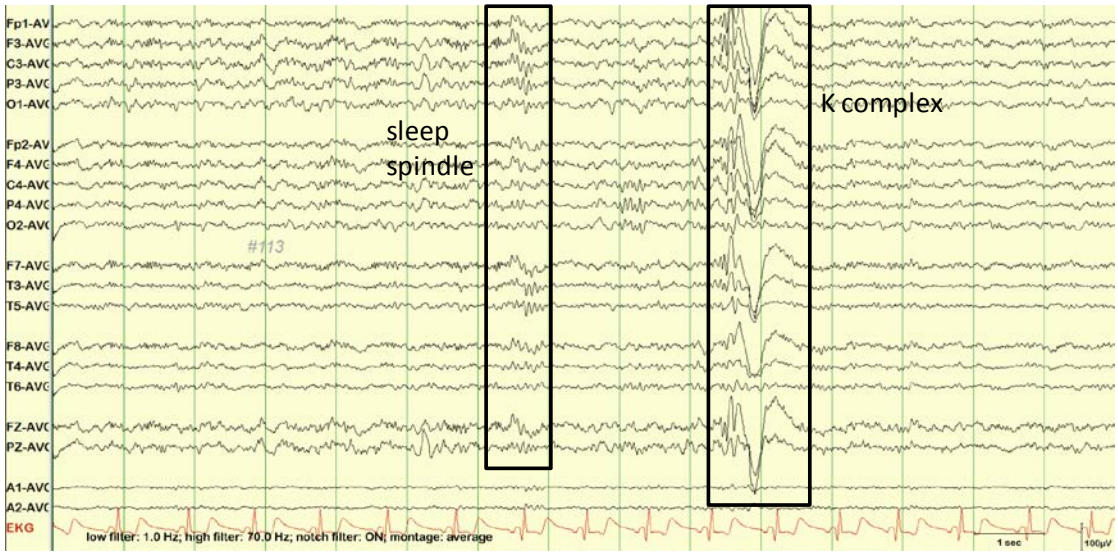


Fig. 4-1 Example of sleep spindle and K-complex [41]

Sleep Stages 3 and 4, defined as Slow Waves Sleep (SWS) are characterized by a further net slowing of the EEG path, which exhibits delta dominant rhythms of high amplitude. An EEG epoch, usually a window of 15 s or 30 s, belongs to sleep stage 3 (S3) if slow waves with frequencies below 2 Hz and amplitudes greater than 75 µV occupy between 20 and 50% of the entire epoch, whereas if the percentage is higher than 50%, the epoch is judged belonging to Sleep Stage 4 (S4).

In S3, the EMG looks decreased with lack of eye movements. In S4 the recording shows intense synchronization, a further slowdown and a higher voltage. EMG activity is even smaller, and no eye movements appear. Generally, the spindles are absent, which, though rarely, may appear in S3. The vegetative functions (breath, heart rate) appear very regular during synchronous sleep phases.

The **REM** stage usually appears after about 90 minutes of NREM sleep, and it is characterized by a complete flattening of the EMG, intense EEG desynchronization, rapid eye movements, cardiac arrhythmias and respiratory changes.

To recognize this stage, the REM waveform must appear as a saccadic movement (i.e., a quick movement of the eyes by which the gaze is transferred from one fixation point to another), with rapid changes in angular velocity at the onset and termination of eye movement. Since small eye movements often appear during the transitions to and from REM sleep, a low amplitude criterion of 40 µV or greater is recommended. In recordings at 15 mm/s (or the equivalent digital display), REM activity should be identifiable as a 45° or greater angular departure from the baseline.

In general, REM episodes are 4 or 5 during 7-8 hours of night sleep, they tend to become longer in the early hours of the morning.

In Fig. 4-2, an example of the hypnogram, the chart used by doctors to represent the different stages of the sleep macrostructure during the night, is shown. In this case, seven episodes of REM stages are observed.

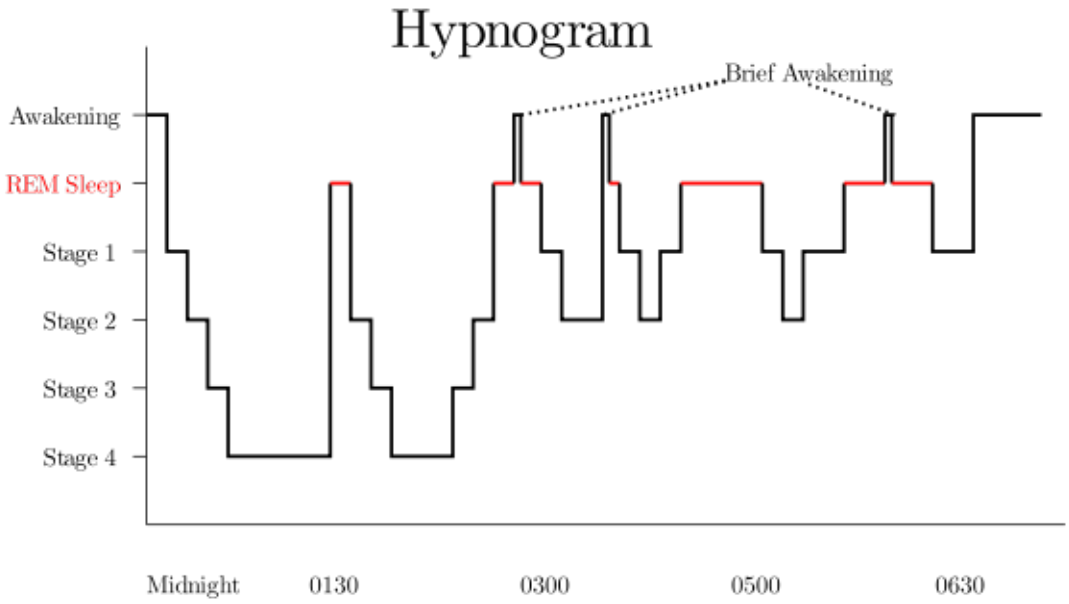


Fig. 4-2 Example of a hypnogram* between midnight and 6.30 am *[42]

4.3 Sleep Microstructure

Within the sleep macrostructure, a periodic pattern can be identified in the different stages of NREM sleep. The microstructure is defined by the presence of a periodic EEG activity, called Cyclic Alternating Pattern (CAP).

CAP is characterized by sequences of transient electrocortical events that are distinct from background EEG activity and recur at up to 1 min intervals [43].

CAP can appear spontaneously in NREM sleep, but it can occur also in association with identifiable sleep pathophysiologies (e.g. sleep-disordered breathing and periodic leg movement activity). It is characterized by cyclic sequences of cerebral activation (phase A) followed by periods of deactivation (phase B), both range between 2 and 60 s. Each phase B separates two successive phase A periods with an interval less than one minute. At least two CAP cycles (phase A + phase B) are required to form a CAP sequence [44].

CAP sequence onset must be preceded by non-CAP (NCAP), which is a continuous NREM sleep EEG pattern with duration greater than 60 s.

This consideration on the time period is not valid in these following cases:

- before the first CAP sequence arising in NREM sleep;
- after a wake to sleep transition;
- after a REM to NREM sleep transition.

An example of hypnogram and CAP sequence is shown in Fig. 4-3. At the top of the figure is reported the hypnogram, representing the sleep macrostructure; whereas at the bottom of the figure, the CAP sequence constituted by a series of CAP cycles is depicted. Each CAP cycle includes a phase A and a phase B.

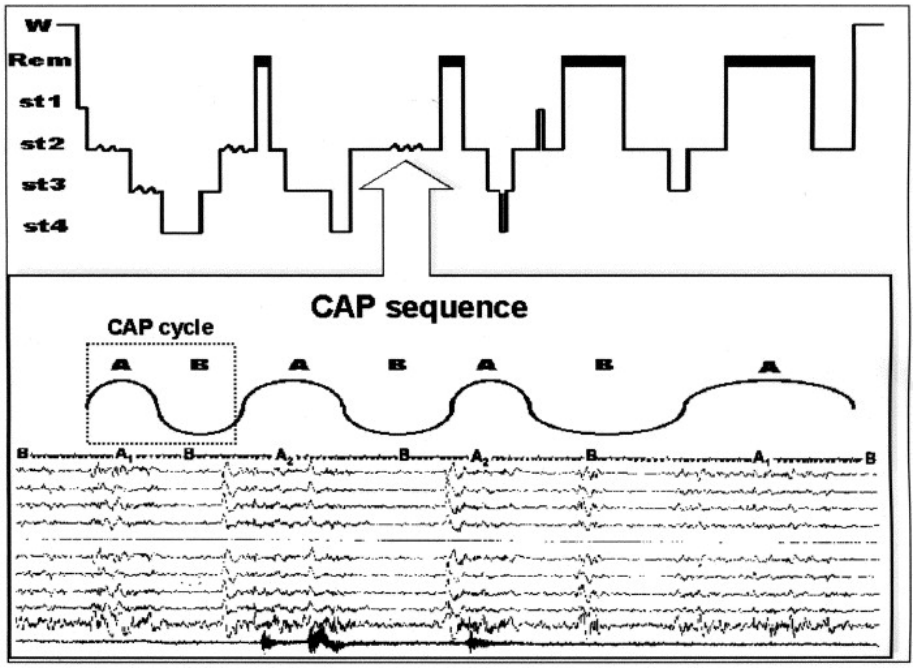


Fig. 4-3 Example of CAP sequence. At the top of the figure, the hypnogram are represented. At the bottom, the CAP sequence [45].

Each phase A is characterized by frequency/amplitude components related to different transient cerebral activation.

In [44], the rules for the scoring of CAP are established. Phase A activities can be classified into three subtypes, based on the reciprocal proportion of high-voltage slow waves (EEG synchrony) and low-amplitude fast rhythms (EEG desynchrony) throughout the entire phase A duration. In Fig. 4-4 the three subtypes of the A phase are reported.

- Subtype A1 is characterized by strong delta waves. EEG desynchrony occupies less than 20% of the entire phase A duration. It includes delta bursts, K-complex sequences, vertex sharp transients, polyphasic bursts with less than 20% of EEG desynchrony. Events are synchronized with low impact on autonomic and somatomotor activities.
- Subtype A2 is characterized by a mixture of slow and fast rhythms with 20–50% of A phase occupied by EEG desynchrony. Subtype A2 specimens include polyphasic bursts with more than 20% but less than 50% of EEG desynchrony. An intermediate influence on the autonomic and somatomotor activities is present.
- Subtype A3 is predominantly identified by rapid low-voltage rhythms with EEG activity greater than 50% of the entire phase A occupied by EEG desynchrony.

Subtype A3 specimens include K-alpha, EEG arousals, and polyphasic bursts with greater than 50% of EEG desynchrony. A movement artifact within a CAP sequence is also classified as subtype A3. The desynchronized EEG events present heavy effects on the autonomic and somatomotor activities.

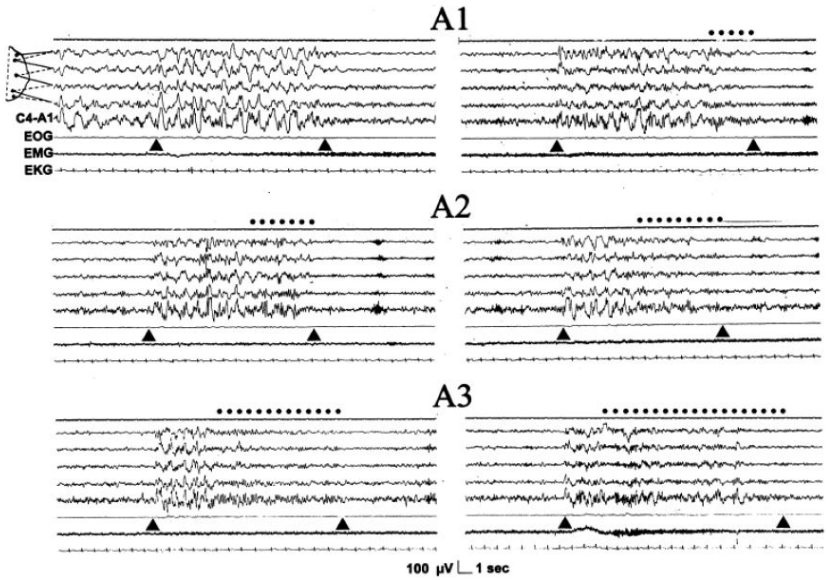


Fig. 4-4 Three Subtypes of the phase A of CAP. The dotted spots indicate the fast low-amplitude portion of the phase A [44].

Despite being a physiological phenomenon, CAP is also a marker of sleep instability and can be correlated with several sleep-related pathologies. In fact, the phase A of CAP has been interpreted as a kind of gate through which pathological events more easily occur [46]. Furthermore, in [37]. A rise in heart rate has been observed 4s after the starting of the phase A, for subjects both healthy and affected by NFLE.

In particular, increased amounts of CAP sequences are often observed in epileptic diseases such as NFLE\ADNFLE [47] [48]. It was noted that, the majority of nocturnal partial motor seizures and temporal lobe seizures occur during NREM sleep, especially in the phase A of CAP with a predominance of the subtype A3, given that it is associated to an increase of fast rhythms with voltage attenuation and strong autonomic activation.

Fig. 4-5 shows an example of sleep structure of a patient affected by ADNFLE. The data have been provided by the Sleep Center of Cagliari. The black signal describes the REM stage, whereas the remainder is the NREM signal and its microstructure, according to the neurologists. The cyan and green signals illustrate the A and B phases of the CAP cycle, respectively; in blue the NCAP phases.

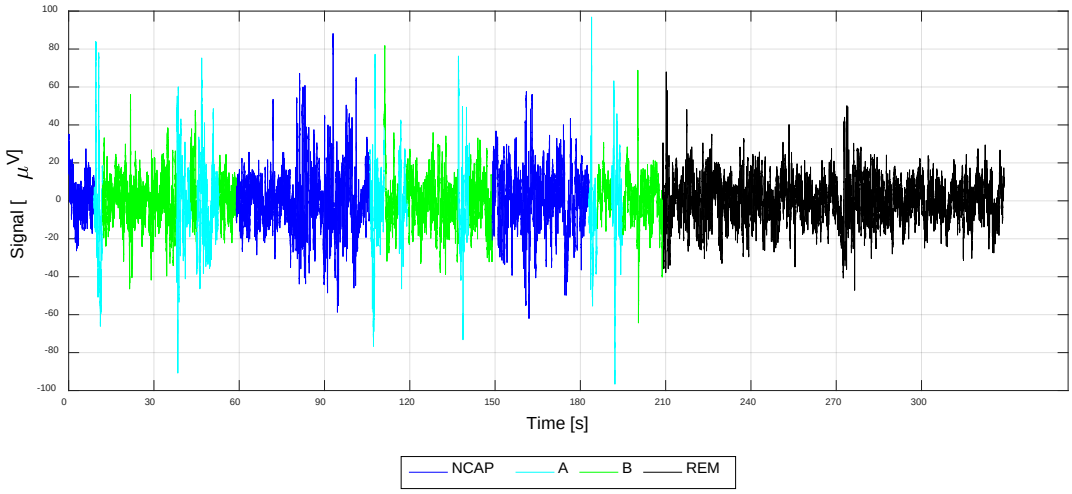


Fig. 4-5 Sleep macrostructure and sleep microstructure of an EEG signal.

5 MACHINE LEARNING

The pioneer of machine learning was Arthur Samuel, who, in 1959, in his study on machine learning applied to the checkers game, investigated on the capability of a computer, that, programmed with incomplete variables, was able to play a better game than the person who wrote the program [49]. The definition of *learning* was given later in a wider vision by Tom Mitchell, in 1997, in which [50]:

> *A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E*.

Consequently, a well-defined learning problem requires a well-specified task, performance metric, and source of training experience.

In general, the machine learning process, as reported in Fig. 5-1, can be described into five main steps:

- PREPROCESSING
- FEATURE EXTRACTION
- FEATURE SELECTION
- MODEL BUILDING
- PERFORMANCE EVALUATION

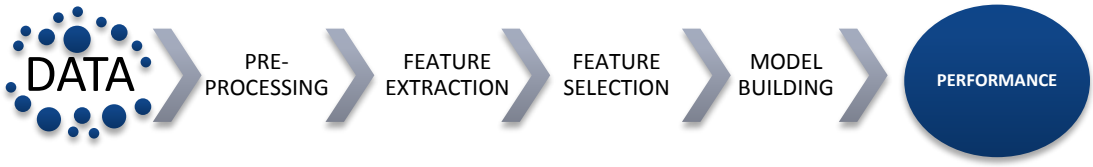


Fig. 5-1 Workflow of the machine learning process.

In the preprocessing step, the raw signals are filtered and cleaned from events that do not contain significant information, thereby reducing data to process. Successively, the variables, named features, that characterize the original dataset are extracted, and a feature subset can be selected with the aim to avoid redundant or not relevant information.

After these steps, the information can be provided to learning algorithms to solve the desired task. In the learning phase, a training set is used to tune the parameters of an adaptive model. Once the model is trained, new data, called test set, can be identified. The ability to correctly identify new examples that differ from those used for training phase is

called *generalization*. Two main categories of learning algorithms can be distinguished: supervised and unsupervised.

In supervised learning, input vectors together with their corresponding desired output vectors are fed into the model. The outputs of the model are compared with the desired outputs and the parameters of the model are changed according to their difference. This error measure is used to guide the learning process.

Examples of supervised learning algorithms include Support Vector Machine (SVM) [51], Decision Tree (DT) [52], Artificial Neural Network (ANN) [53] and K-Nearest Neighbor (KNN) [54].

On the other hand, unsupervised learning, also known as clustering, can identify differences between the input patterns based solely on characteristics of such patterns. The learning system forms groups, cluster, of similar examples within the data, detecting or categorizing persistent features without any feedback from the environment [55]. The fundamental difference from the supervised learning algorithm is that it is not necessary to know the form of the output and therefore, no effective, correct, output is needed. It is useful in cases where the challenge is to discover implicit relationships in a dataset not pre-assigned with labels; its applications can be found in clustering, feature extraction and data reduction purposes [56].

Self Organizing Map (SOM) [57], Principal Component Analysis (PCA) [58], Linear Discriminant Analysis (LDA) [59] and Gaussian Mixture Model (GMM) [60] are some popular examples of the Unsupervised Machine Learning Methods.

Finally, the last step of the machine learning process is the evaluation of the performance. New examples are tested on the trained model and its generalization performance is estimated.

The application of machine learning techniques ranges in several research fields Over the past decades, its employment in seizures classification and prediction, that are extensively presented in reviews [1], [61], [62], demonstrated interesting results.

In the following paragraphs, an overview of the principal techniques and algorithms applied in epileptic seizures detection and prediction are described for each machine learning step.

5.1 Preprocessing

To ensure accurate, efficient and meaningful analysis, data are prepared by means of preprocessing techniques.

General information is collected and, successively, the signals are explored and cleaned by the presence of noise, outliers and artifacts that compromise the success of the results.

5.1.1 Smoothing and Filtering Data

Smoothing is the process of removing noise from data. The simpler technique for smoothing signals consists of a moving average filter, that is a type of Finite Impulse Response (FIR) filter [63], frequently used for time series analysis. It replaces each data point with the average of the neighboring samples included in a considered span.

Let us consider a vector $\boldsymbol{x} \in \mathbb{R}$, each i-th smoothed sample $x_{i_{smooth}}$ is defined by (5.1):

$$x_{i_{smooth}} = \frac{1}{H} \sum_{j=-\frac{H-1}{2}}^{\frac{H-1}{2}} x_{i+j} \tag{5.1}$$

where H indicates the number of samples within the span [64]. This technique contributes to eliminate insignificant variations from a point to the next.

At the same time, the moving average filter shows good performance in time domain, but poor in frequency domain, since it is not able to separate one band of frequencies from another. Thus, a digital filter is designed to extract only the interesting range of frequencies. FIR and Infinite-duration Impulse Response (IIR) filters [63] are the mostly applied; the first is preferred when a linear phase is desired, whereas the second requires a lower filter order to achieve the same given level of performance, with a consequent lower number of operations and computational time.

The selection of the frequency range must take account of the sampling frequency f_s of the signal; in particular, to satisfy the Nyquist criterion (1932):

The sampling frequency should be at least twice the highest frequency contained in the signal.

$$f_s \geq 2f_{max}$$

For example, the scalp EEG signals are generally sampled at 256 Hz, and filtered with a low-pass filter at 0-120 Hz; additionally, a notch filter is applied to remove the power line interference, in order to avoid artifacts that may distort original signals.

5.1.2 Outliers Detection and Normalization Approach

The presence of outliers in the original data can significantly alter the output performance of the model. The process of removing these contributions is performed by grouping or binning methods that identify relationships among the data variables. In the first case, data are clustered and outliers are automatically or manually detected and removed; in the second case, data are sorted and partitioned into bins and successively, smoothed by bin means or bin median.

Subsequently to the steps previously described, data are usually normalized. The normalization can be performed based on the minimum-maximum, scaling the attributes in a range [0-1], or through mean and standard deviation carrying them to a normal distribution, as reported in the equations (5.2) and (5.3), respectively, for a random vector $\boldsymbol{x} \in \mathbb{R}$:

$$x_{i_{norm}} = \frac{x_i - min(\boldsymbol{x})}{max(\boldsymbol{x})}, \tag{5.2}$$

$$x_{i_{norm}} = \frac{x_i - mean(\boldsymbol{x})}{std(\boldsymbol{x})}, \tag{5.3}$$

where $x_{i_{norm}}$ represents the normalized i-th sample of the vector $\boldsymbol{x}$.

5.2 Feature Extraction

In order to choose the best features that can detect epileptic seizures, several approaches have been proposed in literature [65]–[68]. In the following paragraphs, the most used methods in the frequency and time-frequency domains and some features from nonlinear analysis are introduced.

5.2.1 Fourier Transform

The Fourier transform is a useful mathematical instrument for the signal spectral analysis. It is generally applied, with excellent results, on periodic signals as the decomposition is into sinusoidal basis at different frequencies.

The definition of the Fourier transform $\mathcal{F}$ of a continuous function $y(t) \in \mathcal{L}^2(\mathbb{R})$ is given by (5.4):

$$\mathcal{F}(y(t)) = \hat{y}(\omega) = \int_{-\infty}^{\infty} y(t)\, e^{-i\omega t} dt \tag{5.4}$$

The main problem of the Fourier transform is that, while sinusoidal and cosine functions are perfectly located in frequency, they lose information about the timing localization of a given frequency.

For a periodic signal, the transform is a set of discrete values, i.e. a discrete spectrum. The lowest frequency of the sinusoid, representing the signal, is called *fundamental harmonic* and it is of greater importance for the signal reconstruction; the other frequencies are multiple of the fundamental one and are often called secondary harmonics.

However, for transient signals it is more efficient to represent them by localized finite energy bases. The deficiency of the time localization, determines the necessity to apply the *Short Time Fourier Transform* (STFT). It consists of segmenting the signal into time-localization windows and evaluating the spectrum for each of them. One of the choices, even if it is not the only one, is to have a rectangular function g, and translated it in time through a parameter b. The STFT is defined as:

$$STFT(b,\omega) = \int_{-\infty}^{\infty} y(t)g(t-b)\, e^{-i\omega t}dt$$

5.2.2 Wavelet Transform

The wavelet transform provides information from a transient signal in both time and frequency domains, simultaneously. It overcomes the issue of the Fourier analysis of losing temporal localization of the events, through a linear combination of a scaled and translated version of the wavelet basis.

A wavelet basis defines a sparse representation of piecewise regular signals, which may include transients and singularities.

By means of the wavelet basis, the signal is represented by the amplitude of coefficients and the wavelet structure leading to a fast computational algorithm. Wavelets, also called as *small waves*, are finite-energy functions that are suitable for analyzing and processing signals.

The wavelet transforms provide information on both time and frequency of a signal, time-frequency localization capacity and being easily accomplished through a filter banks, makes them an excellent instrument to extract information.

The propriety of the wavelet $\psi(t)$ is that, the area under the graph is zero:

$$\int_{-\infty}^{\infty} \psi(t) = 0,$$

this means that the spectrum of the wavelet has a value of zero at $\omega = 0$.

This spectral-domain behavior makes the wavelet a band-pass filter.

The localization properties are based on the concentration of the wavelet energy in a certain region of both the t- and ω- axes. High resolution and low coefficients in the

representation of the signal in the time-frequency (or time-scale) plane indicate that the energy of the wavelet is concentrated in a small region and a wavelet is localized.

For a given wavelet $\psi(t)$, a scaled and translated version is formulated in

$$\psi_{ba}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$$

Where $a \in \mathbb{R}^+$ and $b \in \mathbb{R}$ are the *scaling* and the *translation* parameters, respectively. The first one captures the local frequency content, whereas the latter, localizes the wavelet basis function around time $t = b$.

The factor $\frac{1}{\sqrt{a}}$ is used to normalize the energy so that it stays at the same level for different values of a and b [69].

For $a = 1$ and $b = 0$, the wavelet results:$\psi_{01}(t) = \psi(t)$; this is called *basic wavelet* or *mother wavelet*. Despite varying the translation and scaling parameters, the shape of the wavelet remains the same.

High frequency wavelets correspond to $a < 1$ or narrow width, while low frequency wavelets have $a > 1$ or wider width, as shown in Fig. 5-2.

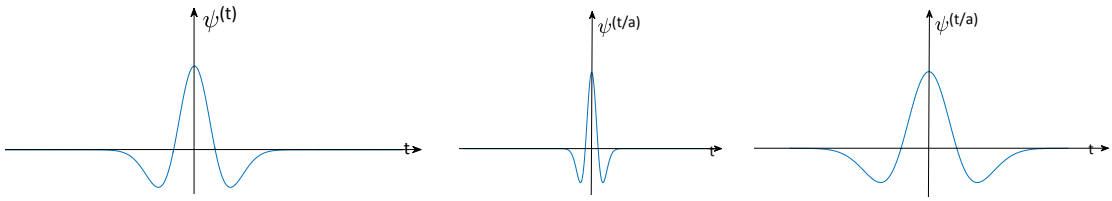


Fig. 5-2 Examples of wavelet at varying the scaling parameter.

A set of orthonormal wavelets $\psi_{ba}(t)$, with the proper choice of the two parameters, constitutes a basis in the $\mathcal{L}^2(\mathbb{R})$ space [70].

While in the Fourier analysis the basis functions are sinusoids, in the wavelet transforms different families of basis functions are employed, from Haar to Morlet wavelets [69]. The most appropriate for the analysis of epileptic EEG data was found to be the Daubechies order 4 (db4) wavelet. This family of wavelets is really appreciated for its orthogonal property and efficient filter implementation [69]. Some examples of Daubechies mother wavelets are shown in Fig. 5-3.

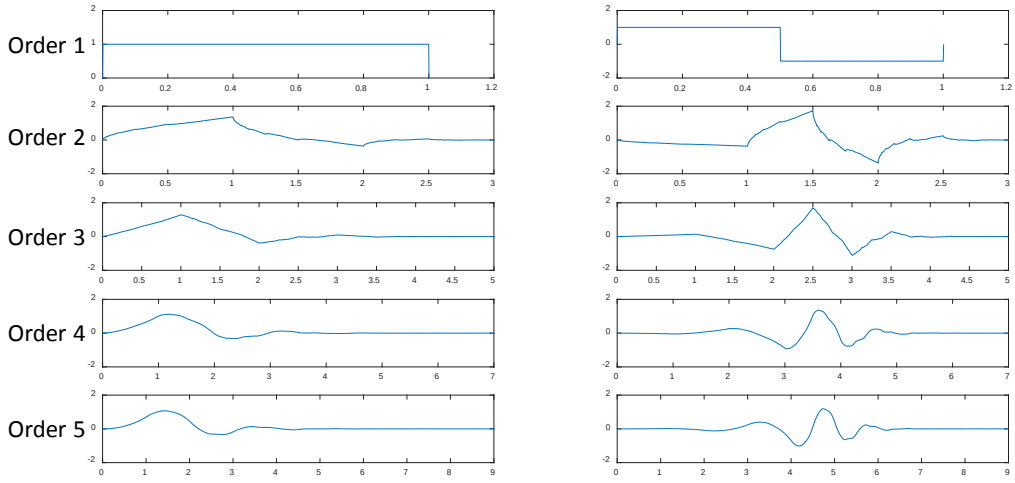


Fig. 5-3 Daubechies scaling and wavelet function at different order.

From the mother wavelet, it is possible to define the Continuous Wavelet Transform (CWT) of the signal $y(t)$:

$$CWT(b,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} y(t)\psi^* \left(\frac{t-b}{a}\right) dt = \langle y(t), \psi_{ba}(t) \rangle, \tag{5.5}$$

where the star (*) denotes its complex conjugate operator.

The linear transform defined by (5.5), computes, via the inner product formula, the wavelet coefficient of $y(t)$ associated with the wavelet $\psi_{ba}(t)$. This coefficient indicates the correlation of the function $y(t)$ to the family of wavelet $\psi_{ba}(t)$. The higher the correlation, the larger the coefficient.

In the CWT, the window widths can be only governed by the scale parameter a, nonetheless, the evaluation of all characteristic of the signal involves a heavy computational load with a consequent generation of redundant information for the signal reconstruction.

The problem of redundancy and the need to apply the wavelet transform to discrete signals led to the application that restrict the scaling and the translation parameters to powers of 2, that is, $a = 2^{-j}$ and $b = k2^{-j}$, where the integer values j and k are the level of resolution and the translation location, respectively.

The multi-resolution analysis (MRA) resolves the function space $\mathcal{L}^2(\mathbb{R})$ into a sequence of subspaces $V^{(j)}$, each of which includes all the subspaces related to the levels lower than j.

Assuming $c_n^{(j)}$ the approximation of a sequence y_n, $n = 1, \dots, N$ at level j, the difference $d_n^{(j)}$ between the two approximations $c_n^{(j+1)}$ and $c_n^{(j)}$ of y_n, is an additional information, called *detail* of $\mathrm{y_n}$.

$$d_n^{(j)} = c_n^{(j+1)} - c_n^{(j)}, \quad n = 1, \dots, N$$

Thus, the subspaces $V^{(j+1)}$ can be resolved as follows:

$$V^{(j+1)} = V^{(j)} \oplus W^{(j)},$$

where the space $W^{(j)}$, orthogonal to $V^{(j)}$, is called *detail space* and $\oplus$ is the orthogonal sum operator. It means that, the subspace $W^{(j)}$ is the orthogonal complementary subspace of $V^{(j)}$ in the subspace $V^{(j+1)}$.

At the limit for $j \to \infty$, $V^{(j)}$ should include all the functions of the domain of interest, that is:

$$\mathrm{V}^{(\infty)} = \{\mathcal{L}^2(\mathbb{R})\}.$$

For each subspaces $V^{(j)}$, the bases $\phi_k^{(j)}$ are generated by a scaling function ϕ, also called *father function*:

$$\phi_k^{(j)}(n) = 2^{\frac{j}{2}}\phi\left(2^j n - k\right).$$

Similarly, by means of the mother wavelet ψ, the bases $\psi_k^{(j)}$ are generated for each subspace $W^{(j)}$.

$$\psi_k^{(j)}(n) = 2^{\frac{j}{2}}\psi\left(2^j n - k\right).$$

Hence, a signal y_n can be written as:

$$y_n = \sum_k c_k^{(j_0)}\phi_k^{(j_0)}(n) + \sum_j \sum_k d_k^{(j)}\psi_k^{(j)}(n) = c_n^{(j_0)} + \sum_j d_n^{(j)}$$

where j_0 is the level of decomposition, $c_k^{(j_0)}$ are the approximation coefficients at resolution j_0, $d_k^{(j)}$ are the detail coefficients at level j.

The Discrete Wavelet Transform (DWT) is based on the relations between subspaces previously described. The algorithm uses two digital filters, a high-pass filter h_n and a low-pass filter g_n, which give rise to a bank of filters able to decompose the signal into components, by means of undersampling operations by a factor of two, as shown in Fig. 5-4. The high frequency coefficients information is drawn out from the analysis and the scaling function components can be further decomposed into two components at resolution j-1.

This process may be repeated until there is no more information left in the signal. The equations of the process expressed in terms of the digital filters g_n and h_n for approximate and detail components are reported in (5.6) and (5.7), respectively.

$$c_n^{(j-1)} = \sum_k g_{2n-k} c_k^{(j)} \tag{5.6}$$

$$d_n^{(j-1)} = \sum_k h_{2n-k}\, c_k^{(j)} \tag{5.7}$$

The signal reconstruction is achieved by inverting the decomposition procedure, as reported in Fig. 5-5. The approximation and the detail at each level are oversampled by a factor of two, then filtered by a high-pass filter and a low-pass filter respectively, and finally summed, to give the approximation at higher level.

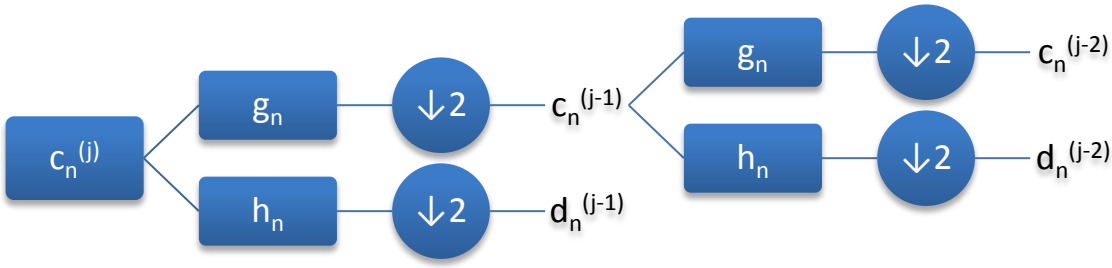


Fig. 5-4 Example of wavelet decomposition.

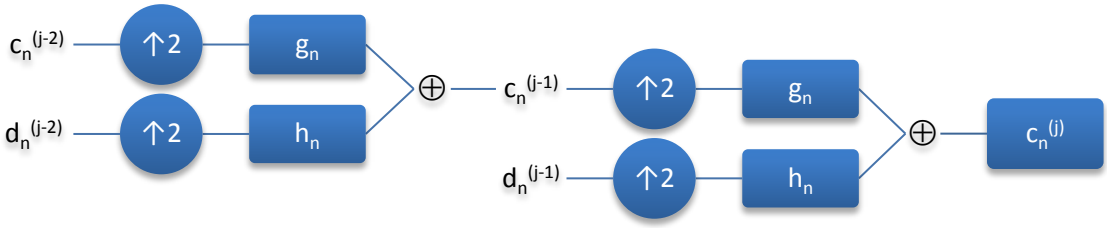


Fig. 5-5 Example of wavelet reconstruction.

The filtering of both approximate and detail components are applied by the Wavelet Packet Transform (WPT) [71]. The wavelet packet coefficients are obtained with a filter bank of conjugate mirror filters that splits the frequency axis in several frequency intervals. Different frequency segmentations correspond to different wavelet packet bases.

5.2.3 Non Linear Analysis

Entropy

Entropy represents the degree of information that a random variable can provide. The more the variable is unpredictable, the higher is the entropy.

Let us consider a set $\boldsymbol{x}$, which contains N events, i.e., $\boldsymbol{x} = \{x_1, x_2, \dots, x_N\}$, three typical formulas of entropy, used in literature, are defined below:

- The *Shannon Entropy*, whose concept of information entropy was firstly introduced in information theory, is defined in (5.8) for the set $\boldsymbol{x} = \{x_1, \dots, x_N\}$:

$$S(\boldsymbol{x}) = -\sum_{i=1}^{N} p(x_i) \log p(x_i) \tag{5.8}$$

  where $p(x_i)$ is the probability of occurrence of the event x_i.
- The *Spectral Entropy*, is obtained by substituting the probability defined in (5.8) with the normalized components of the power spectrum of a signal to describe the variability within the signal [68].
- The *Approximate Entropy* (ApEn), proposed by Pincus [72] is a measure of data regularity, and evaluates the predictability of the current amplitude data, based on the previous amplitude. It is the likelihood that proximal data trends tend to be close together in the next pattern.

Considering a sequence of $\boldsymbol{x}$, $\boldsymbol{x}_i^{(m)}$, with length m , such that $\boldsymbol{x}_i^{(m)} = [x_i, x_{i+1}, \dots, x_{i+m-1}]$, for $1 \le i \le N - m + 1$.

Let $\{\boldsymbol{x}_j^{(m)}\}$ represent a set of sub-sequences obtained from $\boldsymbol{x}_j^{(m)}$ for $1 \le j \le N - m + 1$; each sequence $\boldsymbol{x}_j^{(m)}$ in the set $\{\boldsymbol{x}_j^{(m)}\}$ is compared with $\boldsymbol{x}_i^{(m)}$ and the evaluation of similarity is assessed. The two patterns are similar if the difference between any pair of corresponding samples in the patterns is less than a parameter r, which defines the criterion of similarity.

The fraction of similar patterns is calculated by the parameter $C_i^{(m)}$, defined as follows:

$$C_i^{(m)}(r) = \frac{\tau_i^{(m)}}{N - m + 1}$$

where $\tau_i^{(m)}$ is the number of similar patterns.

Then, the approximate entropy is computed by using $C_i^{(m)}$ and $C_i^{(m+1)}$ as described in (5.9).

$$ApEn(\boldsymbol{x}, m, r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1}\log C_i^{(m)}(r) \; - \frac{1}{N-m}\sum_{i=1}^{N-m}\log C_i^{(m+1)}(r) \tag{5.9}$$

Higher Order Spectra

The *Higher Order Spectra* (HOS) analysis can detect nonlinearity, deviations from Gaussianity and phase relationships between harmonic components of the signal.

It is the spectral representation of statistics of third and higher order. It can also be defined for deterministic sampled signals as the Fast Fourier Transform (FFT) of the third order correlation of a signal:

$$B(f_1, f_2) = E[\hat{y}(f_1)\hat{y}(f_2)\hat{y}(f_1 + f_2)] \tag{5.10}$$

where $E[\cdot]$ is the expectation operator, $\hat{y}(f)$ is the FFT of the signal $y(nT)$, n is an integer index, T is the sampling interval.

The triple product, showed in (5.10) of the two frequencies and their sum, determines the *bispectrum*, which provides information of cross correlation between frequency components in a two-dimensional frequency plot.

5.3 Feature Reduction

After the extraction of the features, the huge dimension of the data set becomes a very important issue for the classification goal. The large volume and complexity of data lead to a high computational time, which could be avoided selecting the attributes that allow us to clearly define a problem.

Hence, feature reduction methods solve the multidimensionality problem retaining only the features with the most informative contents, in order to avoid redundancy instances and classification mistakes during the training phase of the model.

Two feature reduction methods are generally applied in machine learning, the *Principal Component Analysis* (PCA) and the *minimum Redundancy Maximum Relevance* (mRMR) method. In this paragraph, an overview of the principal feature reduction techniques used, are presented.

5.3.1 Principal Component Analysis

The Principal Component Analysis (PCA) [58] is a statistical technique. It aims to reduce the dimensionality of a data set consisting of a large number of interrelated variables *p*, while retaining as much as possible of the variation present in the data set. In particular, it is performed by transforming the original variables to a new set of variables, the principal components (PCs), which are uncorrelated, and sorted in decreasing order so that the first PCs retain most of the variation present in all the original variables.

The PCA can also be considered as a rotation of the axes of the original variable coordinate system to new orthogonal axes, such that the new axes coincide with directions of maximum variation of the original observations [73].

Let us define a $[n \times p]$ matrix $\widetilde{\mathbf{X}}$, that has n independent observations on the p variables. The aim is to obtain a new matrix $\boldsymbol{Z}$ composed of uncorrelated variables and that are linear combination of the original data.

Assuming that each of the original variable is subtracted from their average, leading to a data set whose mean is zero, let $\boldsymbol{X}$ be a matrix whose elements are defined as:

$$\mathrm{x_{i,j}} = \tilde{\mathrm{x}}_{\mathrm{i,j}} - \bar{\mathrm{x}}_{\mathrm{j}} \qquad \text{for } i = 1, 2, \dots, n; \quad j = 1, 2, \dots, p$$

where $\tilde{\mathrm{x}}_{\mathrm{i,j}}$ is the (*i*, *j*)-th element of the matrix $\widetilde{\mathbf{X}}$ and the average $\bar{\mathrm{x}}_{\mathrm{j}}$ is calculated for each variable *j* as:

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} \tilde{x}_{i,j}$$

The matrix $\boldsymbol{Z}$ of PC scores writes:

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{A} \tag{5.11}$$

where $\boldsymbol{A}$ is the orthogonal matrix $[p \times p]$ whose k-th column $\boldsymbol{a}_k$is the k-th eigenvector of the *covariance matrix* $\boldsymbol{S}$.

The resulting $\boldsymbol{Z}$ matrix in (5.11) has dimension $[n \times p]$, each element $z_{i,\mathrm{k}}$ is the score of the i-th observation on the k-th PC

The covariance matrix $\boldsymbol{S}$ can be described as:

$$\boldsymbol{S} = \frac{1}{n-1}\boldsymbol{X}'\boldsymbol{X}$$

The eigenvectors of $\boldsymbol{S}$ are perpendicular to each other and provide the information about the patterns in the data. At each eigenvector an eigenvalue l is associated corresponding to the variance of the associated PC. After the evaluation of the eigenvectors, the components are ordered based on their eigenvalues l_k, from the highest to the lowest, that gives the order of significance. The eigenvector with the largest eigenvalue is the first PC of the data set.

If the original variables were partially correlated, some eigenvalues would have a negligible value.

This means that the corresponding eigenvectors can be neglected and it is possible to limit the representation only to eigenvectors with the greatest eigenvalues.

Additionally, the covariance matrix on the PCs basis is diagonal, it means that the total variance is the sum of the variances of each PC.

Components selection

To decide how many PCs should be used in order to account for most of the variation of $\boldsymbol{X}$, different rules are developed.

The most commonly applied empirical methods are three: the cumulative percentage of total variation, the size of variance of PCs and the scree graph and the log-eigenvalue diagram.

In the *cumulative percentage of total variation* method, the first *m*-components of $\boldsymbol{Z}$ whose contribution is around 70% - 90% are considered.

Since, as stated earlier, the variance of the k-th PC is represented by the eigenvalue l_k, the evaluation of the percentage variation, accounted for the first m PCs, is:

$$\rho_m = 100\frac{\sum_{k=1}^{m} l_k}{\sum_{k=1}^{p} l_k}$$

The best choice for a cut-off value ρ^*, to retain the m PCs, where m is the smallest integer for which $\rho_m > \rho^*$, is in the range 70% to 90% [58]. However, it can sometimes be higher or lower depending on the practical details of a particular data set. A value greater than 90% is indicated when one or two PCs represent very dominant sources of variation. Conversely, when p is very large, the selection of m PCs corresponding to 70% might still be higher for the further analyses. One solution is to decrease the threshold.

The method of *the size of variance of PCs,* consists in choosing only the m PCs with $l_k > 1$, according to the Kaiser's rule [74], in which any PC with variance less than 1 contains less information than one of the original variables, and so it is not worth retaining.

This rule is specifically constructed for use with the correlation matrix to evaluate the PCs, in view of the fact that, if all elements of $\boldsymbol{X}$ are independent, the PCs are the same as the original variables and will have unit variances. However, the method could be adapted in case of covariance matrix application.

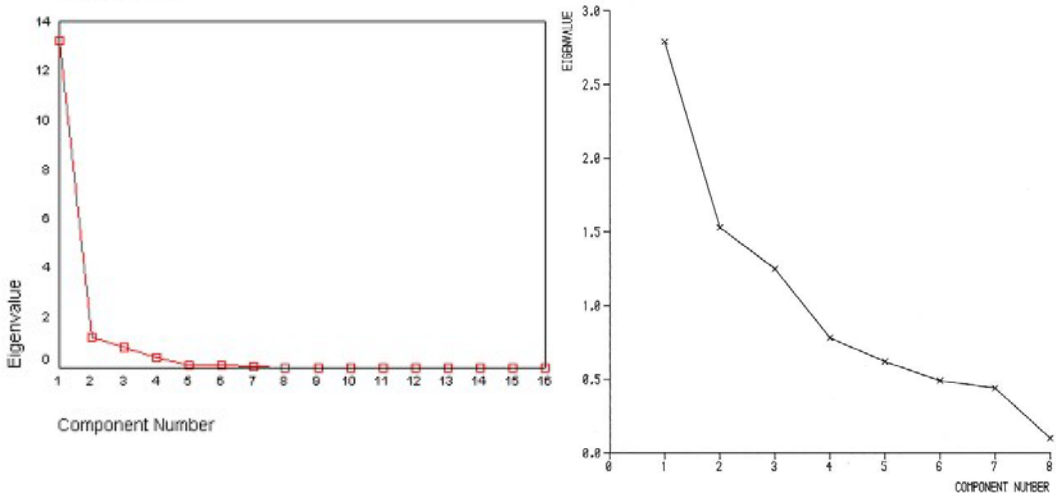
Despite the unit threshold, cut-off values $l^* < 1$ have been explored. According to simulation studies, in [75] a value of 0.7 was considered appropriate.

The *scree graph and the log-eigenvalue diagram* are based on choosing the components by graphical representation of l_k against k.

The *scree* graph, named because of the similarity of the typical shape to that of the accumulation of scree, at the foot of a mountain slope, consists in selecting the k value based on the slope alteration of the curve. The *k* in which is present an 'elbow', a changing of steepness, defines the m PCs that should be kept.

The alternative to this graph is the log-eigenvalue diagram, where the *y*-axis is represented by the logarithm of the eigenvalues. However, this procedure is not indicated when the slope gradually becomes less steep, with no clear elbow.

In Fig. 5-6 two examples of scree graphs are depicted, Fig. 5-6a shows a sharp slope of the curve, meaning that the first three PCs give significant information, whereas the components on the shallow slope may add little contribution to the solution. Conversely, in Fig. 5-6b no clear elbows can be observed, meaning that, the method is not suitable for PCs selection.



a. b.

Fig. 5-6 Example of scree graphs: a) represents sharp slope variations [76], b) represents gradual slope variations [58].

5.3.2 Independent Component Analysis

While PCA identifies the new dimensional space by finding the major axis of variation, the Independent Component Analysis (ICA) identifies the new dimensional space by finding the components that are independent from each other. The main computational difference between PCA and ICA is that, while PCA uses variance which is a second order statistic, ICA uses higher order statistics for computation of independent variables [77].

ICA assumes that each observed variable is a linear combination of the independent components or sources. The requirements to identify the sources are that the independent components are statistically independent and have non-gaussian distributions.

An approach for ICA estimation, inspired by information theory, is the minimization of *mutual information* [78] of the components, that is a measure of the dependence between random variables, and it is null and not negative only when each variable is independent from the other.

5.3.3 Linear Discriminant Analysis

The Linear Discriminant Analysis (LDA), proposed in 1936 by Fisher [79], used both in feature extraction and building classification model, provides a dimensional reduction through linear combinations of the observed variables, so as to separate two or more classes without altering the location of the initial data. It is obtained by maximizing the difference between the predefined groups with respect to the new variable.

Let us define a set $\mathbf{X}$ of p-dimensional vectors $\boldsymbol{x}$ with n data points belonging to M classes, in which $\boldsymbol{x}_j^{(\varsigma)}$ is defined as the j-th vector belonging to the ς-th class.

To perform dimensionality reduction, each sample vector is mapped into a lower q-dimensional space by a linear transformation:

$$\boldsymbol{y}_j^{(\varsigma)} = \boldsymbol{W}^T \boldsymbol{x}_j^{(\varsigma)}$$

$\boldsymbol{W}$ is determined in order to reduce data variation in the same class and increase the separation between classes.

Considering the mean for each cluster as $\boldsymbol{\mu}_\varsigma$ and for the entire data set as $\boldsymbol{\mu}$, the aim of the Fisher-LDA is the maximization of the objective function expressed as:

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T S_B \boldsymbol{w}}{\boldsymbol{w}^T S_w \boldsymbol{w}} \tag{5.12}$$

where $\boldsymbol{w}$ is the transformation vector; whereas $\mathbf{S}_\mathrm{B}$, defined in (5.13) and $\mathbf{S}_\mathrm{W}$, defined in (5.14), are the between-class and within-class scatter matrices, respectively.

$$\boldsymbol{S}_B = \frac{1}{n}\sum_{\varsigma=1}^{M} n_\varsigma (\boldsymbol{\mu}_\varsigma - \boldsymbol{\mu})(\boldsymbol{\mu}_\varsigma - \boldsymbol{\mu})^T \tag{5.13}$$

where n_ς represents the number of data point belonging to the ς-th class.

$$\begin{gathered} \boldsymbol{S}_W = \frac{1}{n}\sum_{\varsigma=1}^{M} S_\varsigma \\ \text{where } S_\varsigma = \sum_j \left(\boldsymbol{x}_j^{(\varsigma)} - \boldsymbol{\mu}_\varsigma\right)\left(\boldsymbol{x}_j^{(\varsigma)} - \boldsymbol{\mu}_\varsigma\right)^T \end{gathered} \tag{5.14}$$

The objective function in (5.12) can be optimized by evaluating eigenvectors and eigenvalues, whose equation to establish the transformation vector $\boldsymbol{w}$ is reported in (5.15).

$$S_B \boldsymbol{w} = l S_W \boldsymbol{w} \tag{5.15}$$

Generally, the q generalized eigenvectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_q$, associated with the first q largest eigenvalues l, compose the column vectors of the transformation matrix $\boldsymbol{W}$. These generalized eigenvectors form the projection vectors of LDA [59].

5.3.4 Minimum Redundancy Maximum Relevance

The minimum Redundancy Maximum Relevance (mRMR) approach, developed by Ding and Peng [80], extracts a set of feature that has the minimal redundancy among them and

that are more representative for the classification. To evaluate the redundancy and relevance, for discrete variables, the *mutual information* can be applied.

Mutual information, a measure widely used in communication systems [81], represents the amount of information that a variable X contains about another variable Y:

$$I(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}$$

where $p(x_i, y_j)$ is the *joint probability density* for the two variables and $p(x_i)$, $p(y_j)$ are the individual probabilities. The minimal redundancy, calculated for a subset S, is defined by the equation (5.16):

$$\min W_I \qquad W_I = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \tag{5.16}$$

where x_i, x_j are two features belonging to a subset S.

The aim is to choose the set of features S, that best characterizes the whole data information. On the other hand, the relevance takes into account the information between features and their different target classes h; the maximization of their mutual information, reported in the equation (5.17) determines the extraction of the features with high discriminant power.

$$\max V_I, \qquad V_I = \frac{1}{|S|} \sum_{i \in S} I(h, x_i) \tag{5.17}$$

The combination of the two quantities can be evaluated maximizing their difference (5.18) or their quotient (5.19). The last one is considered as the best choice when applied to discrete variables [80].

$$\max (V_I - W_I) \tag{5.18}$$

$$\max (V_I / W_I) \tag{5.19}$$

5.4 Building Model

5.4.1 Artificial Neural Network

An Artificial Neural Network (ANN) is a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs [82].

Its inspiration comes from the biological nervous system, where the activation or inhibition of the neurons depends on the sum of input signals and the threshold potential.

In the artificial neuron, represented in Fig. 5-7, the inputs are multiplied by the connection weights, summed, and subsequently passed through an activation function to produce the output. The weighed sum of the neuron inputs x, including a bias, usually represented by b is defined as v:

$$v = \sum_{i=1}^{n} w_i x_i + b$$

The activation *function* φ associates the sum v to the output value y.

$$y = \varphi(v)$$

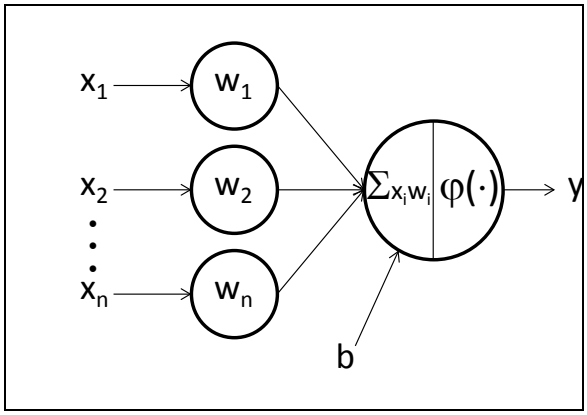


Fig. 5-7 Representation of the artificial neuron.

The most commonly used activation function φ is the sigmoid function [83]:

$$\varphi(v) = \frac{1}{1 + \exp(-v)}$$

Other activation functions are described in Table 5-1.

Table 5-1 Activation functions

Threshold	$\varphi(v) = \begin{cases} 0, & v < 0 \\ 1, & v \geq 0 \end{cases}$
Hyperbolic tangent	$\varphi(v) = \tanh(v)$
Linear	$\varphi(v) = v$

The artificial neural networks are typically organized into layers; they can be classified in two main categories based on their topologies: feed-forward and recurrent networks.

In feed-forward neural networks, the neurons in a layer get input from the previous layer and feed their output to the next layer, not permitting connections to the neurons of the same or previous layers and not keeping a record of the previous output values. The feed-forward networks can be further subdivided in static or dynamic networks. In the static network, the output of the network at a given moment depends only on the input at the same time instant, whereas in dynamic network, the output depends on the initial and past states of the neurons.

In Fig. 5-8, an example of feed-forward network is represented. The network is composed by an input, an output layer, and a certain number of layers between them, named *hidden layers*. A network constituted by an input layer and an output layer of neurons define the *single layer neural network*; whereas the presence of one or more hidden layers describes the *multilayer networks*. A classical example of a feed-forward network is the *Multi-Layer Perceptron*, firstly described by Rosenblatt in 1961 [84].

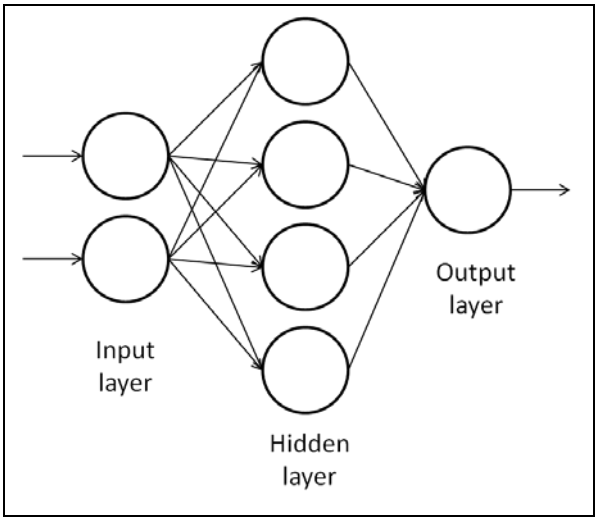


Fig. 5-8 Architecture of a feed-forward neural network.

In the *Recurrent Neural Networks* (RNNs), represented in Fig. 5-9, the network contains, instead, at least one feed-back connection, the activations can therefore flow round in a loop, enabling the networks to do temporal processing and learn sequences. These networks are dynamics, the composite input at time t has some historical information about the happenings at time $T < t$. An example of recurrent network is the Elman Networks (EN) [85].

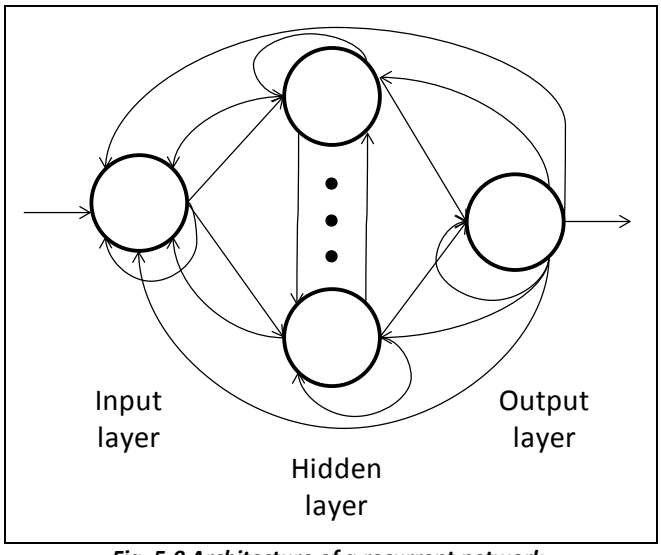


Fig. 5-9 Architecture of a recurrent network.

Perceptron

The perceptron, proposed by Rosenblatt in 1956, is the simplest form of feed-forward neural network. It consists of a single neuron with adjustable synaptic weight and bias as depicted in Fig. 5-7. Normally, the output is evaluated based on the Heaviside function. Perceptron, represented by a single layer of process nodes, can be used to classify linearly separable patterns that lie on opposite sides of a hyperplane.

Multi-Layer Perceptrons

Perceptrons with at least three layers, i.e., a minimum of one hidden layer, and non-linear activation functions, are named Multi-Layer Perceptrons (MLP). Each node in the network, with the exception of the input layer nodes, can perform computing and transformation operations. Each of these nodes calculates a weighing aggregate function of the confluent signals to the node itself, and an activation function processes the input to output signal.

In Fig. 5-10, a MLP with a single hidden layer is shown. It presents a structure with *d* inputs x, n_H neurons in the hidden layer and *s* neurons in the output layer.

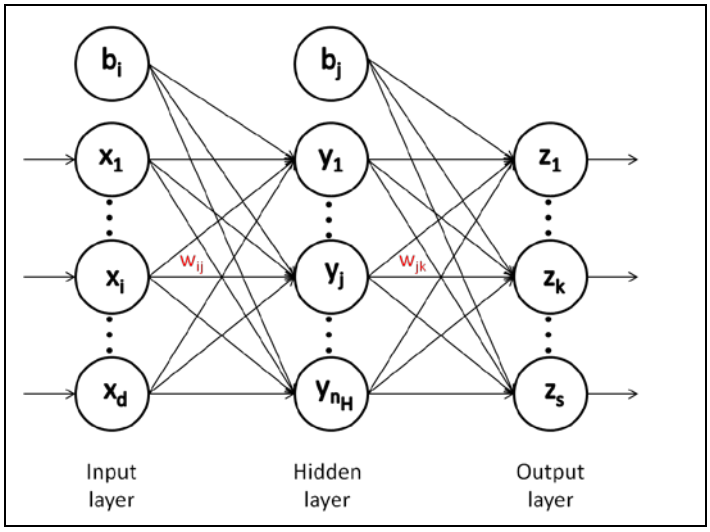


Fig. 5-10 Example of a Multi Layer Perceptron neural network.

The total number of parameters (weights) results: $d \times n_H + n_H \times s + n_H + s$, where the last two are the weights of the *biases*.

The outputs $\mathbf{z}$ of the network are given by transforming with the activation function φ the weighted sum of the outputs of the prior hidden layer and the bias. The evaluation of the k-th output z, is given by:

$$z_k = \varphi\left(\sum_{j=1}^{n_H} w_{j,k}^{(1)} y_j + b_k^{(1)}\right) \tag{5.20}$$

where $w_{j,k}^{(1)}$ denotes the weight in the single hidden layer, of the connection from the *j*-th unit of the hidden layer to the *k*-th output unit, whereas y_j and $b_k^{(1)}$ represent the activation of the hidden unit *j*, reported in (5.21), and the bias for the output unit *k*, respectively.

The output y_j of the *j*-th hidden unit is calculated by the nonlinear transformation of the weighted sum of the inputs and the bias, as described in the following equation:

$$y_j = \varphi\left(\sum_{i=1}^{d} w_{i,j} x_i + b_j\right) \tag{5.21}$$

Hence, the equation (5.20) can be written as:

$$z_k = \varphi\left(\sum_{j=1}^{n_H} w_{j,k}^{(1)} \varphi\left(\sum_{i=1}^{d} w_{i,j} x_i + b_j\right) + b_j^{(1)}\right) \tag{5.22}$$

Note that, the equation (5.22) can be extended to the case of more than one hidden layer with the same iterative process.

To evaluate the contribution of each weight, different learning algorithms can be applied. Basically, they estimate the local error at each neuron and systematically update the network weights. The combination of weights which minimizes the error function is considered to be a solution of the learning problem. Some of the algorithms currently used to minimize the error functions are the steep gradient descent that is a method of the first order, and the Newton's method, a second order method. Additionally, another method applied to minimize a non linear function, is the Levenberg-Marquardt algorithm that converges to the solution faster than the steep gradient descent algorithm [86].

Probabilistic Neural Network

The probabilistic neural network (PNN), based on statistical Bayesian decision rule, is a feed-forward neural network with two hidden layers, named *pattern* and *summation* layers.

Each pattern neuron performs a dot product of the input pattern vector with a weight vector and then, employs the Radial Basis Function (RBF) as activation function, instead of the sigmoid function, commonly used for back-propagation. Then, each neuron of the

summation layer, that represents an output class, computes the sum of the contribution only from the pattern neurons associated with the output class.

The PNN is commonly used in classification purposes; it offers a speed advantage for problems in which the incremental adaptation time of back propagation is a significant fraction of the total computation time. One of the advantages is the capability to operate completely in parallel without a need for feedback from the individual neurons to the inputs. In addition, the toleration of the erroneous samples and the possibility to choose the parameter of RBF based on the complexity of the system, make the PNN an advantageous classification model [87].

5.4.2 K-Nearest Neighbor

The K-Nearest Neighbor (KNN) classifier is a supervised learning model that classifies new points based on a similarity measure.

Assuming a p-dimensional training feature space and an unseen p-dimensional instance $\boldsymbol{x}$, let us consider a positive integer K. The KNN classifier evaluates the distances between x and each training observations; after that, the K points in the training data closest to x are selected. Finally, x gets assigned to the class with the largest conditional probability, that is the fraction of points of a given class that are in the subspace defined by the K value. An example of KNN classification for $K = 5$ is shown in Fig. 5-11. The instance x will belong to the negative class.

The distance between two instances can be measured in many possible ways. The most employed metric is the *Euclidean distance*. Given two data points in p-dimensional feature space: $\boldsymbol{x}_1 = (x_{11}, \dots, x_{1p})$ and $\boldsymbol{x}_2 = (x_{21}, \dots, x_{2p})$, the Euclidean distance between these points is given by:

$$d(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sqrt{\sum_{j=1}^{p} \left(x_{1j} - x_{2j}\right)^2}$$

Other distances, that can be used, are:

- Manhattan distance $d(x_1, x_2) = \sum_{j=1}^{p} \left|x_{1j} - x_{2j}\right|$.
- Minkowski distance $d(x_1, x_2) = \left(\sum_{j=1}^{p} \left|x_{1j} - x_{2j}\right|^{\xi}\right)^{1/\xi}$,

where ξ denotes the order of the norm. For $\xi = 1$ or $\xi = 2$ the Minkowski distances correspond to the Manhattan and the Euclidean distances, respectively.

The KNN classifier is easy to implement and interpret, the classification is based solely on local information. However, KNN classification is time consuming and storage intensive [88], additionally, it is sensitive to the correct choice of K: a small value of K gives less stability,

whereas, a large value of K causes lower precision. An empirical rule-of-thumb is choosing K equal to the square root of the number of instances [89].

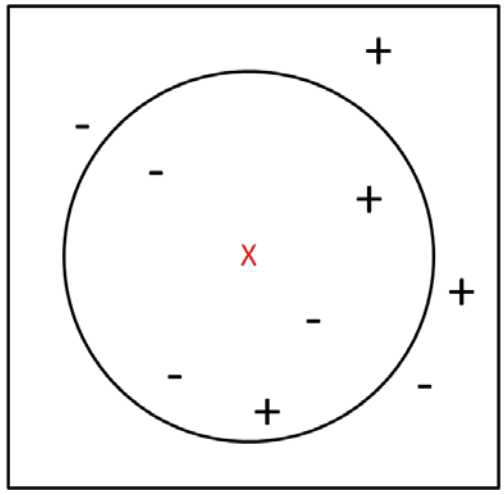


Fig. 5-11 Example of KNN classification for K=5.

5.4.3 Decision Trees

A decision tree (DT) is a tree-like structural supervised model. It is designed with leaves, which represent classifications or decisions, and branches, which represent the conjunctions of features that lead to those classifications.

The classification of a particular pattern begins at the root node and ends at leaf nodes. Each leaf specifies a test of some attribute of the input instance and each branch descending from that node corresponds to one of the possible values for this attribute, as shown in Fig. 5-12. The example, reported by Mitchell [50], illustrates the classification of Saturday mornings according to whether or not they are suitable for playing tennis.

DT learning is generally best suited to problems whose instances are describable by attribute–value pairs and target function has discrete output values [50], as in Fig. 5-12 with the Boolean classification (i.e., yes or no) to each example.

A variety of DT algorithms can be applied to solve the problem, three popular algorithms are: CART (Classification and Regression Tree), ID3 (third in a series of identification, “ID”, procedures) and C4.5, the classification successor algorithm of ID3.

The C4.5 algorithm is the most popular in classification tree methods. It constructs the tree, employing a top-down, greedy search through the space of possible branches using entropy and the information gain. The *information gain* is defined as the expected reduction in entropy caused by partitioning the examples according to this attribute.

Assuming S as the training set and A as the attribute, the information gain is:

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

where $Values(A)$ is the set of all possible values for attribute A, and S_v is the subset of S for which attribute A has value v. For example, in the case depicted in Fig. 5-12, the attribute *Wind*, can have the values *Weak* or *Strong*.

After the construction of the tree, the algorithm uses a heuristics formula for the pruning phase, where some pairs of neighboring leaf nodes are removed in order to have leaf nodes with minimum impurity (i.e., heterogeneity of composition).

This simple and interpretable structure allows decision trees to solve multi-type attribute problems and manage missing values or noise data. The method is robust to errors, both errors in classifications of the training examples and errors in the attribute values. However, the method presents difficulties of interpretation when trees have large dimension and sometimes, it cannot guarantee the optimal accuracy compared to other machine learning methods [88].

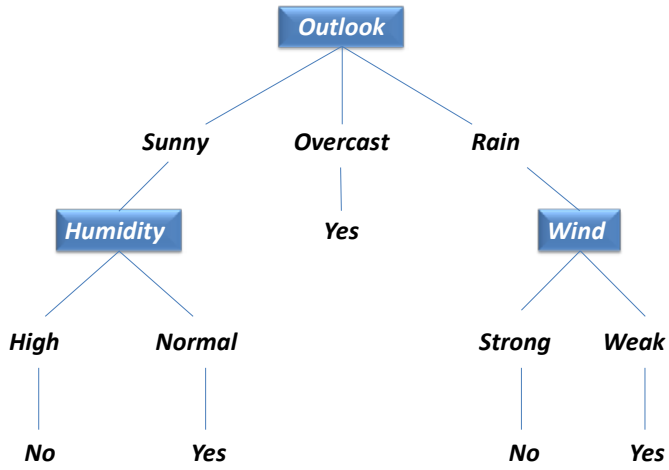


Fig. 5-12 A decision tree for the concept Play Tennis.

5.4.4 Gaussian Mixture Model

The *Gaussian Mixture Model* (GMM) is an unsupervised machine learning technique based on the expectation maximization (EM) method. It assumes that all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. It is used as clustering method, where clusters are considered as Gaussian distributions centered on their barycenters.

Given an observed set $\mathfrak{I} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, where $\boldsymbol{x}_i$ is a d-dimensional vector, the multivariate Gaussian density, for each of the k Gaussian components, is defined with the parameters: mean $\boldsymbol{\mu}_\mathrm{j}$ and covariance Σ_j.

$$p_\mathrm{j}(\mathbf{x}_\mathrm{i}|\theta_\mathrm{j}) = \frac{1}{\sqrt{(2\pi)^d|\Sigma_\mathrm{j}|}} \exp[-\frac{1}{2}(\mathbf{x}_\mathrm{i} - \boldsymbol{\mu}_\mathrm{j})^\mathrm{T}\Sigma_\mathrm{j}^{-1}(\mathbf{x}_\mathrm{i} - \boldsymbol{\mu}_\mathrm{j})], \qquad 1 \le i \le n, 1 \le j \le k$$

where $\theta_\mathrm{j} = \{\boldsymbol{\mu}_\mathrm{j}, \Sigma_\mathrm{j}\}$

Let us define$p(\boldsymbol{x_i})$ as a finite mixture model with k components:

$$p(\mathbf{x}_\mathrm{i}|\Theta) = \sum_{j=1}^{\mathrm{k}} \alpha_j \, p_j(\mathbf{x}_\mathrm{i}|\mathrm{z}_\mathrm{j}, \theta_\mathrm{j})$$

where:

$p_j(\mathbf{x}_\mathrm{i}|\mathrm{z}_\mathrm{j}, \theta_\mathrm{j})$ are mixture components, $1 \le j \le k$. Each is a density or distribution defined over $p(\boldsymbol{x}_i)$, with parameters θ_j.

$\boldsymbol{z} = (z_1, \ldots z_k)$ is a vector of k binary indicator variables that are mutually exclusive and exhaustive, it represents the identity of the mixture component that generated $\boldsymbol{x}_i$.

$a_j = p(z_j)$ are the mixture weights, representing the probability that a randomly selected $\boldsymbol{x}_i$ was generated by component k, where$\sum_{j=1}^{k} \alpha_j = 1$

$\Theta = \{\alpha_1, \ldots, \alpha_k, \theta_1, \ldots, \theta_k\}$ is the complete set of parameters for a mixture model with k components.

The EM method consists of the expectation (E) and the maximization (M) steps performed iteratively. The first step computes the expectation of the log likelihood with respect to the current estimate of the distribution, evaluating the responsibilities (or membership weights) using the current parameter values:

$$\omega_{ij} = p(z_{ij} = 1|\boldsymbol{x}_i, \Theta^{(\mathrm{t})}) = \frac{p_j(\boldsymbol{x}_i|z_j, \Theta^{(\mathrm{t})})\alpha_j}{\sum_{m=1}^{k} p_m(\boldsymbol{x}_i|z_m, \Theta^{(\mathrm{t})})\alpha_\mathrm{m}}, \qquad 1 \le i \le n, \qquad 1 \le j \le k$$

The second step computes the updated parameters $\boldsymbol{\theta}^{(t+1)}$ that maximize the expected log likelihood.

$$\mu_j^{(t+1)} = \frac{1}{n_j}\sum_{i=1}^{n} \omega_{ij}\boldsymbol{x}_i \qquad 1 \le j \le k$$

$$\Sigma_j^{(t+1)} = \frac{1}{n_j}\sum_{i=1}^{n} \omega_{ij}(\boldsymbol{x}_i - \boldsymbol{\mu}_j^{(t+1)})(\boldsymbol{x}_i - \boldsymbol{\mu}_j^{(t+1)})^T, \qquad 1 \le j \le k$$

$$a_j^{(t+1)} = \frac{n_j}{n}$$

where $n_j = \sum_{i=1}^{n} \omega_{ij}$, it is the effective number of data points assigned to component j, and T denotes the transpose operation.

The entire iterative process repeats until the algorithm converges. The main advantages of this model are the flexibility in choosing the components distribution and in obtaining a density estimation for each cluster. However, it fails with problems having a large dimension of data.

5.4.5 Support Vector Machine

The *Support Vector Machine* (SVM) is one of the most used supervised machine learning method for binary classification [51]. The geometrical interpretation of *Support Vector machine Classification* (SVC) is that the algorithm projects the points to be classified into a larger space, called *feature space*, where the classes are linearly separable from each other. In this space, a separating hyperplane is determined, maximizing the margin between the classes, defined as the distance between the hyperplane and the points of the two classes closest to it.

Let us first consider the problem of separating two linearly separable sets. Let n training vectors belonging to two separate classes defined in y:

$$\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}, \qquad \boldsymbol{x}_i \in \mathcal{R}^p, y_i \in \{-1, 1\},$$

Let a linear hyperplane that divides the negative from positive training samples, and that is defined by the following equation:

$$\boldsymbol{w} \cdot \boldsymbol{x} + b = 0 \tag{5.23}$$

where $\boldsymbol{w}$ is the weight vector normal to the hyperplane and b is the bias. $\frac{b}{\|\boldsymbol{w}\|}$ is the perpendicular distance from the hyperplane to the origin, and $\|\cdot\|$ is the Euclidean norm.

The margin is obtained by the sum of the minimum distances evaluated from the separating hyperplane to both the closest positive and negative samples, as depicted in Fig. 5-13. The points are called *support vectors*, and are illustrated in the Fig. 5-13 with extra circle. The parallel hyperplanes, where the support vectors lie, are indicated as in (5.24) and, (5.25) for the positive and negative samples, respectively:

$$\boldsymbol{x_i} \cdot \boldsymbol{w} + b = 1 \qquad \text{for } y_i = 1 \tag{5.24}$$

$$\boldsymbol{x_i} \cdot \boldsymbol{w} + b = -1 \qquad \text{for } \mathrm{y_i} = -1 \tag{5.25}$$

Considering all the vectors, the constraints for the training samples can be written by the inequality (5.26):

$$y_i(\boldsymbol{x_i} \cdot \boldsymbol{w} + b) - 1 \geq 0, \qquad \forall i \tag{5.26}$$

The aim of the SVM is to maximize the margin in order to achieve the optimal separation from the two categories. Since the distance is inversely proportional to $\|\boldsymbol{w}\|$, the problem is solved minimizing $\|\boldsymbol{w}\|^2$, under the constraint (5.26), applying the Lagrangian formulation. Thus, the problem can be indifferent solved in two ways, due to the particular dual expression of the problem. One solution is reported in (5.27), under the constrains $\alpha_i \geq 0$, where α_i are Lagrange multipliers.

$$\max_{\alpha} \left(\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j < \boldsymbol{x}_i, \boldsymbol{x}_j > \right), \qquad \alpha_i \geq 0, i = 1, \dots, n \tag{5.27}$$

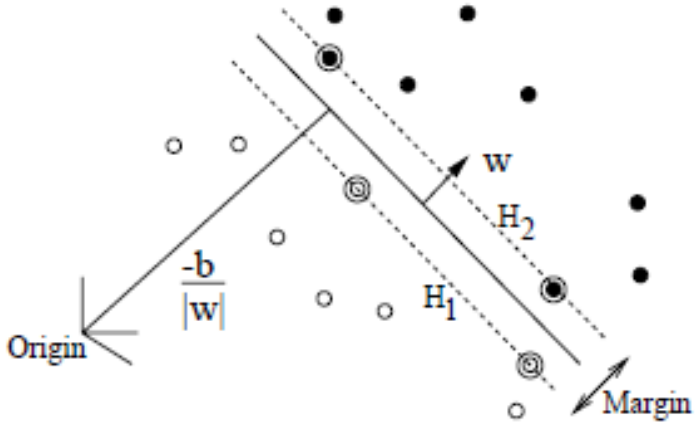


Fig. 5-13 Example of the hyperplanes positioning in the case of separable vectors [51].

Considering that, in most cases, training data cannot be separable, a feasible solution is to introduce a slack variable ξ_i in (5.23), which takes into account the training error. Thus, the minimization problem becomes the (5.28), subject to the constraint reported in (5.29):

$$\min_{(\mathbf{w},b,\xi)} \left(\frac{1}{2} \boldsymbol{w}^T \boldsymbol{w} + C \sum_{i=1}^{n} \xi_i \right), \qquad \xi_i \geq 0, i = 1, \dots, n \tag{5.28}$$

$$y_i(\boldsymbol{w}^T \phi(\boldsymbol{x_i}) + b) \geq 1 - \xi_i, \qquad \xi_i \geq 0, i = 1, \dots, n \tag{5.29}$$

where, *C>0* is the regularization parameter, that bounds the training error.

The optimization problem in (5.28) has two objectives: the maximization of the margin between the input data classes and the minimization of the number of misclassifications. The parameter *C* controls the trade-off between these two goals.

To generalize the problem, training data can be classified by means of non linear separation surfaces. The non-linear SVM maps the input vectors $\boldsymbol{x}$ into a higher dimensional space using a non-linear mapping function ϕ in order to find a better separation of the classes. While in the linear SVM the optimization problem depends on the dot product of the training data $\boldsymbol{x}_i \cdot \boldsymbol{x}_j$, in non linear SVM, a kernel function $\mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i) \cdot \phi(\boldsymbol{x}_j)$ is introduced [90].

Among the kernel functions, whose expressions are reported in Table 5-2, the most applied in non linear SVM, is the Radial Basis Function (RBF):

$$\mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma \left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2\right) \tag{5.30}$$

where *γ* can be seen as the inverse of the radius of influence of samples selected by the model as support vectors.

Hence, the optimal separating hyperplane (5.23) and the optimization problem (5.27) become the (5.31) and the (5.32), respectively.

$$\boldsymbol{w} \cdot \phi(\boldsymbol{x}) + b = 0 \tag{5.31}$$

$$\max_{\alpha} \left(\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) \right), \tag{5.32}$$

subjected to $\sum_{i=1}^{n} y_i \alpha_i = 0,\ 0 \leq \alpha_i \leq C,\ i = 1, \dots, n.$

Table 5-2 Some of the kernel functions most used in SVM.

Linear	$\mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^T \boldsymbol{x}_j$
Radial Basis Function	$\mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma \left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2\right)$
Polynomial	$\mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left(\boldsymbol{x}_i^T \boldsymbol{x}_j + 1\right)^{\nu}$, ν is the degree of polynomial kernel
Sigmoid	$\mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh\left(\eta \boldsymbol{x}_i^T \boldsymbol{x}_j + r\right)$, η and r are kernel parameters

5.4.6 Self Organizing Map

A *Self Organizing Map* (SOM) [57] is an unsupervised neural network, developed by Kohonen, that demonstrated very useful to convert complex nonlinear relationships between data items into a low-dimensional space, preserving the topological properties of the input space.

Its applications range from the speech recognition to image analysis.

The SOM defines a mapping from a *p*-dimensional input space onto a regular (usually two-dimensional) array of neurons, where, each neuron is represented by a *p*-dimensional weight vector $\boldsymbol{w}$.

The SOM applies a competitive learning algorithm adapting and updating, in each training step, the weights vectors of the SOM, in order to obtain the minimum distance with the input vectors.

Let us consider a vector $\boldsymbol{x} \in \mathbb{R}^p$ randomly chosen from the training set N, and $\boldsymbol{w}_i(t)$ as the weight vector of the *i*-th neuron, at time t, that was first initialized. The neuron, whose weight is closest to the input vector, is called the Best Match Unit (BMU).

$$\|\boldsymbol{x} - \boldsymbol{w}_c\| = \min_i \|\boldsymbol{x} - \boldsymbol{w}_i(t)\| \tag{5.33}$$

Where $\boldsymbol{w}_c$ represents the weight vector of the BMU. In (5.33) the Euclidean distance is considered.

According to the competitive Learning Rule, the weight vectors are subsequently updated,

$$\boldsymbol{w}_i(t+1) = \boldsymbol{w}_i(t) + \alpha(t) h_{ci}(t)[\boldsymbol{x}(t) - \boldsymbol{w}_i(t)]$$

Where:

- $h_{ci}(t)$ is a decreasing neighborhood function of time, it is centered around the neuron *c*, with a radius σ*;*
- $\alpha(t)$ is the learning rate at time t whose value can be $0 \leq \alpha(t) \leq 1$;
- $\boldsymbol{x}(t)$ is the vector, randomly chosen, at time t.

The learning step includes the weight vectors update of all neurons around the neuron *c*, that are contained within a time-variable radius. The procedure normally consists of taking a wide radius and a large initial learning rate at the beginning, and shrinking with time.

The neighborhood function determines how strongly the neurons are connected to each other, influencing the training result of SOM procedure. Four examples of neighborhood function commonly applied in the SOM algorithm are shown in Fig. 5-14 are displayed [91]:

- *Bubble function* is a constant function expressed by:

$$h_{ci}(t) = \mathbf{1}(\sigma_t - d_{ci})$$

Where σ_t is the neighborhood radius at time *t;* $d_{ci} = \|r_c - r_i\|$ is the distance between the neuron *c* and *i* on the map grid and $\mathbf{1}(\rho)$ is the unit step function, that is $\mathbf{1}(\rho) = 0$ if ρ <0 and $\mathbf{1}(\rho) = 1$ if $\rho \geq 0$.

- *Gaussian function* is expressed by:

$$h_{ci}(t) = exp(-d_{ci}^2/2\sigma_t^2)$$

- *Cut-Gaussian function*, is obtained multiplying Gaussian function by the bubble function, expressed by:

$$h_{ci}(t) = exp(-d_{ci}^2/2\sigma_t^2)\ \mathbf{1}(\sigma_t - d_{ci})$$

- *Epanechnikov function* is a piecewise polynomial function defined as:

$$h_{ci}(t) = \max\{0, 1 - (\sigma_t - d_{ci})^2\}$$

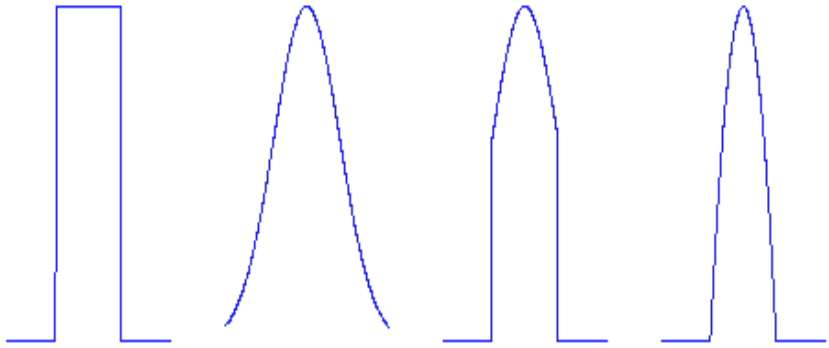


Fig. 5-14 Representation of neighbourhood functions. From the left: Bubble function, Gaussian function, Cut-Gaussian function and Epanechnikov function [91].

The map size K, that is the number of neurons constituent the map, and the map shape, are two of the training parameters that influence the system output. K is correlated to the size of the training set; a high dimensional map causes greater flexibility, allowing a better cluster separation, but elevated computational time. In [91], the hexagonal shape grid and the empirical value of K are suggested. The calculation of the map size is reported in the equation (5.34), in which N is the number of the training set samples.

$$K = 5 \cdot N^{0.54321} \tag{5.34}$$

5.5 Performance

To quantify the performance quality of a classification, or prediction, system, different metrics can be evaluated.

Given a binary classification (prediction) problem, and calling positive points, or samples, those associate to the positive label in the training set, and negative samples those associated to a negative one, the output of the classifier can be represented by the confusion matrix, a table that presents the following four categories:

- *True positive* (TP): number of positive samples correctly classified (predicted) as positive by the model.
- *False negative* (FN): number of positive samples wrongly classified (predicted) as negative by the model.
- *True negative* (TN): number of negative samples correctly classified (predicted) as negative by the model.
- *False positive* (FP): number of negative samples wrongly classified (predicted) as positive by the model.

Based on these definitions, the statistical performance indexes, typically used in epileptic seizure detection and prediction, are described below:

- *Sensitivity* (SS) or *Recall*: is the ratio between the positive samples correctly classified (predicted) by the model and the sum of all positive samples, it defines how good a test detects (or predicts) the positive class.

$$SS = \frac{TP}{TP + FN}$$

- *Specificity* (SP): is the ratio of negative samples correctly classified (predicted) by the model and the sum of all negative samples, it defines how good a test detects (predicts) the negative class:

$$SP = \frac{TN}{TN + FP}.$$

- *Accuracy* (ACC): is the ratio between the samples correctly classified (predicted) and all the samples, it defines the closeness of the test to known values.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

- *Precision* (PR) or *Positive Predicted Value*: is the ratio between the positive samples correctly classified (predicted) by the model and the samples classified

(predicted) as positive, it defines how many of the positively classified (predicted) are relevant.

$$PR = \frac{TP}{TP + FP}$$

- *False Positive Rate* (FPR): is the ratio between the negative samples wrongly classified (predicted) and the total negative samples. It defines the rate of incorrect results of the negative class.

$$FPR = \frac{FP}{FP + TN}$$

- *False Negative Rate* (FNR): is the ratio between the positive samples wrongly classified (predicted) and the total positive samples. It defines the rate of incorrect results of the positive class.

$$FNR = \frac{FN}{FN + TP}$$

6 MACHINE LEARNING IN EPILEPSY

In order to automatically detect and predict epileptic seizures all the previously described steps of the machine learning approach must be followed. First of all, the selection of the more suitable diagnostic signals has to be performed from, e.g., scalp and intracranial EEG, electrocardiography, accelerometer and motion sensors, electrodermal activity, and audio/video captures. Then, feature extraction techniques and subsequent classification models have to be tested in order to maximize the performance in detecting and/or predicting epileptic seizures.

6.1 Epilepsy Detection using EEG

Several methods have been proposed in literature where the scalp or intracranial EEG recordings is largely used to automatically detect several kinds of epileptic seizures with good results. In particular, different features extraction and selection methods associated with as many classification models have been investigated with the aim to find the best performance. Table 6-1 reports a selected collection of seizure detection methods proposed in the recent literature referring to the feature extraction methods and the classification models proposed. Moreover, for the sake of comparison, the declared performance in terms of sensitivity, specificity, and accuracy have been reported in the same table, when available.

Polat and Gunes [92] proposed the classification between normal and epileptic classes applying the Welch FFT as feature extraction method. In their study the decision tree was used as classification model, reaching an accuracy of 98.72%. However, the same authors, by means of PCA as feature reduction method, and a hybrid model based on Artificial Immune Recognition System (AIRS) [93] with fuzzy resource allocation as classification model, reported an accuracy about 100% [94].

To take into account the information given by the EEG signals, that are typically non-stationary time series, the feature extraction method in the time-frequency domain has been widely applied in [95] [96] [97] [98].

Berdakh and Don in [99], used the Continuous Wavelet Transform to detect pre-ictal and ictal epochs. Four different wavelet functions (Biorthogonal 1.3, Biorthogonal 1.5, db2, db5) were used to find the best performance in terms of sensitivity, specificity and accuracy obtaining an average of 95.49%, 93.8% and 94.69% respectively.

The application of wavelet packet transform, associated with a combined seizure index (CSI), in order to detect seizures in patients affected in the majority (11 of 14 patients) by temporal lobe epilepsy, was instead proposed by Zandi et al. [67]. The distinction between

seizure and non-seizure states by the patient-specific algorithm reached a mean value of sensitivity of 90.5%.

A greater value of sensitivity was reached by Gabor et al [100] with a method based on DWT, who proposed a combination of DWT and SOM to identify 181 seizures from 65 patients, achieving a sensitivity of 92.8% with 1.35 ± 1.35 false positive errors per hour. The SOM has been also applied in combination to a fuzzy logic, with the aim to incorporate spatial contextual information in the spike detection problem, by James et al. [101]. The authors obtained a low value of sensitivity, equal to 55.3%, with a selectivity of 88%, because of the inclusion of questionable events in the data set that contributed to incorrect estimation.

Orhan et al [102] combined the DWT with the multilayer perceptron neural network (MLP) to classify healthy segments from epileptic seizure-free segments and epileptic seizure segments extracted by the database described in [103]. The algorithm achieved a 100% of accuracy in the distinction from healthy to epileptic seizures, higher than the results obtained by Subasi et al [96] in which a modular neural network named as Mixture of Experts (ME) and MLP reached accuracies of 94.5% and 93.2%, respectively. Also Kocyigit et al. [104] classified healthy segments and epileptic seizure segments with MLP model, extracting features from EEG raw data by means of fast independent component analysis (ICA). The authors were able to achieve interesting values of sensitivity and specificity, equal to 98% and 90.5%, respectively.

The same database were investigated by Lima et al. [97], associating the DWT through a variant of the Support Vector Machine (SVM), the least square SVM (LS-SVM), obtaining the performance of sensitivity, specificity and accuracy around 100%, both for RBF (Radial Basis Function) and ERBF (exponential RBF) kernels configuration model. Moreover, the comparison with the SVM revealed that the sensitivity profiles were qualitatively similar to LS-SVM.

Mahajan et al. [105] classified between healthy segments and segments from patients affected by absence seizures, comparing two feature reduction methods: the PCA and ICA. The authors, after the decomposition of the EEG signals through the DWT, evaluated the performance in terms of sensitivity, specificity and accuracy for both the SVM and ANN models investigated. Overall, the SVM associated with the ICA method, showed better accuracy (99.50%) than applying the PCA method (96.15%) or combining ANN with both the feature reduction methods, that resulted 62.93% and 96.75% for PCA and ICA methods, respectively. However, Subasi and Gursoy [106], investigated on the comparison between three different feature reduction methods: PCA, ICA and Linear Discriminant Analysis (LDA) combined with the SVM, reaching values of sensitivity, specificity and accuracy around the 100% for both ICA and LDA methods.

Given the higher accuracies, the application of time-frequency techniques, such as DWT, and SVM appears to be promising methods for diagnosis systems.

However, nonlinear feature extraction methods proved to be very useful for the detection of epilepsy. Acharya et al. [107] extracted the nonlinear features: Higher Order Spectra (HOS), approximation entropy and sample entropy from the EEG segments, and different classifiers were then used to evaluate the classification performance among normal, inter-ictal and ictal EEG signals, reaching an accuracy of 98.5%. The same classification was investigated by Chua et al. [108], comparing two different classification models: the Gaussian Mixture Model (GMM) and the SVM, using the features derived from HOS and from the power spectrum. HOS showed better performance than the power spectral feature. The application of HOS as features, reached an average accuracy of 93.11%, in the case of the GMM model, where 92.00%, 89.00% and 97.67% were the accuracies obtained for normal, pre-ictal and epileptic EEG phases, respectively. Conversely, for the case of the SVM classifier with HOS based features, the average classification accuracy was 92.56%, achieving accuracies of 96.67%, 83.67% and 97.33% for normal, pre-ictal and epileptic EEG phases, respectively.

The use of the Approximate Entropy (ApEn) associated with two types of neural networks (ANNs) as detection models, was proposed by Srinivasan et al. [109]. The authors employed the Elman Network (EN) and the Probabilistic Neural Network (PNN) to distinguish between healthy segments, measured with the scalp EEG, and epileptic segments, measured with intracranial EEG, achieving high values of overall accuracies around 100%. The application of entropy as features can be also observed in Acharya et al. [110], in which the authors evaluated the detection performance using four types of entropy: ApEn, Sample Entropy (SampEn), and two phase entropies. The features were fed into seven classifiers: Fuzzy [111] , SVM, K-Nearest Neighbor (KNN), PNN, Decision Tree (DT), GMM and Naive Bayes Classifier (NBC) [112], in order to identify normal, pre-ictal and ictal EEG segments.

As can be noted in Table 6-1, Fuzzy classifier achieved the higher performance, with a sensitivity of 99.4%, specificity of 100% and accuracy of 98.1%.

The combination of features extracted in multi-domain was presented by different authors [68] [113] [114] [115] [116].

Temko et al. [113] proposed a seizure detection system using scalp EEG signals, demonstrating the capability of the SVM to discriminate data using the entire dataset coming from 17 newborns affected by Hypoxic Ischemic Encephalopathy. The extraction of 55 features in time and frequency domains and in information theory, associated with a smart post-processing approach, due to the presence of unbalance data, reached good performance. The percentage of seizure events detected were around 89%, with a cost of 1

FD (False Detection)/h up to 100% of seizure detected with a cost of 4 FD/h. Additionally, some improvement reported in [117], suggested how to tune a parameter in the post processing, achieving, in this way, a considerably decrease of FD/h from 0.45 to 0.25.

As well as detection problems, the multi-domain and non linear analysis associated with SVM or ANN models, reached good performance in the prediction of epileptic seizures using EEG signals [118] [119][120] [121].

Rasekhi et al. [122] combined 22 multiple univariate features, extracted from scalp EEG, to predict refractory epilepsy. Ten patients were studied, varying the pre-ictal time, using the SVM, obtaining 48 different models for each patient. The 73.9% of epileptic seizures were predicted, with a False Prediction Rate (FPR) of 0.15 per hour.

In addition, the prediction of epileptic seizures of 21 patients affected by intractable focal epilepsy, using intracranial EEG (iEEG) signals, was investigated by Park et al. [123]. Spectral power, calculated in several band frequencies, were used as features to train an SVM model, performing a sensitivity equal to 97.5%, false alarm rate of 0.27 per hour and total false prediction times of 13.0%.

Table 6-1 Selected seizure detection systems using EEG signals.

Reference	Feature Extraction method	Model	Sensitivity	Specificity	Accuracy
Gabor (1998) [100]	DWT	SOM	92.8%		
Khan and Gotman. (2003) [124]	DWT	Heuristic rules	87.7%		
Kannhatal et al. (2005) [125]	Entropy	Adaptive Neuro-Fuzzy Inference System (ANFIS) [93]			92.22%
Subasi and Erçelebi (2005) [126]	DWT	MLP	92.8%	92.3%	
Wilson (2005) [127]	FFT	PNN	89%		
Hopfengärtner et al. (2007) [128]	FFT	Thresholding technique	90.9%		
Polat and Gunes (2007) [92]	FFT	Decision tree			98.72%
Srinivasan et al. (2007) [109]	ApEn	ANN (EN, PNN)			99.35-100% (EN) 98%-100% (PNN)
Subasi (2007) [96]	DWT	ME ANN			94.5% 93.2%
Ghosh-Dastidar et al. (2008) [129]	Mixed-band feature space	Radial Basis Function Neural Network (RBFNN) [130]			96.60%
Kocyigit et al. (2008) [104]	Fast ICA	MLP	98%	90.5%	
Guo et al.(2009) [131]	DWT	ANN			95.20%
Ocak (2009) [95]	DWT + ApEn	ANN	96%	96.00%	96.00%
Guo et al. (2010) [132]	ApEn +Wavelet	ANN			99.85%
Lima et al. (2010) [97]	DWT	SVM	100%	100%	100%

Minasyan et al. (2010) [133]	Multi-domain features	RNN	100% (Post-onset)		
Subasi and Gursoy (2010) [106]	DWT	SVM	99%(PCA)	98.5%(PCA)	99%(PCA)
			100% (ICA)	99% (ICA)	100%(ICA)
			100% (LDA)	100% (LDA)	100%(LDA)
Zandi et al. (2010) [67]	WPT	Combined Seizure Index (CSI)	90.5%		
Acharya et al. (2011) [107]	HOS+ Entropy	SVM			98.5%
Chua et al. (2011) [108]	HOS	GMM SVM			93.11% 92.56%
Orhan et al [102]	DWT	MLP			100%
Shen et al. (2011) [134]	Entropy	SVM			98.91%
Temko et al. (2011) [113]	55 features in different domains	SVM	100% (4FD/h)		
Wang et al. (2011) [135]	DWT	KNN			100%
Uzzaman Khan et al. (2012) [136]	DWT	LDA	83.6%	100%	91.8%
Acharya et al. (2012) [110]	Entropies	Fuzzy [111]	99.4 %	100%	98.1%
		SVM	97.2%	100%	95.9%
		KNN	97.8%	97.8%	93%
		PNN	97.8%	97.8%	93%
		DT	98.3%	91.1%	88.5%
		GMM	98.3%	95.6%	95.9%
		NBC [112]	94.4%	97.8%	88.1%
Mahajan et al. [105]	DWT	ANN	62.93%(PCA) 96.75%(ICA)	98.83%(PCA) 96.75%(ICA)	93.63%(PCA) 96.75%(ICA)
		SVM	96.15%(PCA) 99.50%(ICA)	99.48%(PCA) 97.05% (ICA)	97.75%(PCA) 99.50%(ICA)
Parvez et al. (2014) [137]	DWT+ Empirical Mode Decomposition (EMD)	LS-SVM	91.36%		

Shoaib et al. (2014) [138]	DWT	SVM	91% -96%		
Fergus et al. (2015) [115]	PSD, Entropy and other statistical features	KNN	93%	94%	
Kavitha et al. (2016) [139]	DWT	SVM	64%-83%	88%-92%	81%-89%
Alsharabi et al [114] (2016)	DWT+ Entropy	ANN	100%	100%	100%
Chen et al. (2017) [65]	DWT	SVM	97%	100%	99.33%
Liu et al. (2017) [140]	Increment entropy	SVM			97.32%
Wang et al. (2017) [68]	Entropy + wavelet	SVM	97.98%	99.56%	99.25%

6.2 Epilepsy Detection using ECG

It has to be highlighted that, some studies proposed the ECG recordings to detect epileptic seizures revealing interesting results.

Leutmezer et al. in [38] found that, in 85% of 58 patients, tachycardia occurred during the seizures. Übeyli, in [141], used the DWT as feature extraction method and SVM as classification model to identify the ECG beats between normal and partial epilepsy, evaluating the post ictal oscillations and obtaining an accuracy of 99.44%.

Furthermore, Malarvili et al. [142] proposed an automatic seizure detection system for newborns affected by epilepsy, using the heart rate variability to classify seizures from non-seizure events. The algorithm, tested on eight newborns, achieved a sensitivity of 85.7% and a specificity of 84.6%.

6.3 Nocturnal Epilepsy Detection

Despite the good results published in these last decades, very few contributions are available to detect nocturnal epileptic seizures.

A detection system based on accelerometers (ACM) in order to detect motor seizures, was proposed by Nijsen et al. [143] with a percentage of seizure detected around the 48%. ACM assesses the acceleration along the three orthogonal direction axes, giving optimal information about the emergence of motor events. Another system based on ACM attached to the proximal limbs of seven pediatric patients affected by NFLE, was investigated by Van de Vel et al. in [144]. The authors reached a good percentage of seizure detection equal to 95.7%, but a low percentage of positive predictive value (57.8%), given by a high number of false positive outputs.

Although the use of the accelerometers can give a significant contribution in the motor events detection, it is subjected to false alarms, due to movements that are not always associated to epilepsy and cannot be readily distinguished from convulsions.

From the review previously reported, EEG signals provided interesting performances in seizure detection systems using wavelet transform, because of the non-stationary nature of the signals. Among the presented models, SVM achieved good results in seizure and no-seizure identification.

However, the advantage to visualize and analyze complex multidimensional data in a 2-dimensional map, preserving the topology of the input patterns has been investigated in the present thesis.

Two systems [145], [146], based on the scalp EEG signals, for detection of nocturnal frontal lobe epileptic seizures, have been proposed. The approaches, applying the Self Organizing Map (SOM), revealed the mapping potential to recognize the nocturnal epileptic seizures obtaining good performances. The results have been then compared with the SVM system, showing similar, and, in some patients, higher performances than those obtained with the SVM model.

The proposed systems are described in detail in the rest of the thesis.

7 PATIENT-CUSTOMIZED SYSTEM FOR NOCTURNAL EPILEPTIC SEIZURE DETECTION

As every patient possesses different physiological characteristics [11], in this thesis a detection system for nocturnal epileptic seizures is firstly proposed, customized for the specific patient.

The diagnostic data comes from the European Epilepsy Database (http://epilepsy-database.eu) and a specific detection system has been developed for each patient, using the EEG signals. The Self Organizing Maps (SOM) have been used as classification model to identify the seizures; that networks are able to investigate over not straightforward relations in the complex input space.

This chapter is organized in three sections: the first one describes the European Epilepsy Database; in the second section, following the steps of the machine learning procedure, a system to automatically detect the Nocturnal Frontal Lobe Epilepsy seizures is proposed. Finally, in the third section, the use of the SOM has been extended to identify seizures from patients affected by frontal and/or temporal nocturnal epileptic seizures. In order to evaluate the suitability of the proposed approach, the obtained results have been compared to another classification model, which has been developed using SVM technique, widely used, in literature, as seizure detection classifier.

7.1 European Epilepsy Database

Data, for developing nocturnal epileptic seizure detection models, have been extracted by the European Epilepsy Database, that is a product of the EU-funded project *EPILEPSIAE* [147].

The database includes high-quality, long-term continuous EEG data of 300 epilepsy patients, enriched with clinical metadata; recorded by three different epileptic units: Unidade de Monitorização em Epilepsia e Sono, Centro Hospitalar e Universitário de Coimbra, Portugal, Unitéd'Épilepsie of the Pitié-là-Salpêtrière Hospital, Paris, France and Epilepsy Center, and Epilepsiezentrum, Universität-sklinikum Freiburg, Germany. The data have been collected with the approval of the ethics boards of the respective Institutions, and the informed consent of all the patients.

The EEG signals were obtained with scalp and intracranial electrodes. Recordings with a continuous duration of at least 96 hours (4 days) per patient, with a minimum of three clinically manifest seizures, are available. Depending on the patient, the dataset contains

from 22 to 37 scalp EEG channels; electrodes that were placed according to the 10-20 International System, with sampling rate of 256 Hz, 512 Hz, 400 Hz or 1024 Hz [148].

In the present thesis, Unipolar EEG derivations have been considered, where the EEG has been monitored with a common reference electrode for all the channels.

The database provides the information about the sleep macrostructure, and start and end timestamps. The macrostructure scoring, provided only for the seizures duration, was performed by sleep experts based on the R&K rules [39].

In this thesis, only the sleep state, that includes the NREM and REM stages, are considered. From the European Epilepsy Database, 22 patients (14 males and 8 females) affected by nocturnal frontal and/or temporal epilepsy with two or more epileptic seizures have been selected.

For each patient, the information in terms of localization of seizures, onset age of the epilepsy, the gender, the number of nocturnal seizures and the hospital where the data has been recorded, is reported in detail in Table 7-1. Five nocturnal epileptic seizures have been averagely occurred per patient.

Table 7-1 Selected patients from European Epilepsy Database.

Patient	Localization of seizures	Onset age	Gender	Number of nocturnal seizures	Hospital
Pat 1	Frontal	31	M	5	Coimbra
Pat 2	Frontal	4	F	7	Paris
Pat 3	Frontal	3	F	3	Paris
Pat 4	Frontal	23	M	9	Paris
Pat 5	Frontal	30	M	5	Coimbra
Pat 6	Frontal	22	F	8	Freiburg
Pat 7	Frontal	7	F	2	Paris
Pat 8	Frontal	10	F	2	Paris
Pat 9	Frontal	18	M	15	Freiburg
Pat 10	Frontal and Temporal	1	F	3	Freiburg
Pat 11	Frontal and Temporal	13	F	2	Freiburg
Pat 12	Frontal and Temporal	35	M	3	Coimbra
Pat 13	Frontal and Temporal	3	M	3	Coimbra
Pat 14	Frontal and Temporal	20	M	8	Freiburg
Pat 15	Frontal and Temporal	0	M	4	Freiburg
Pat 16	Frontal and Temporal	5	M	6	Freiburg
Pat 17	Frontal and Temporal	8	M	6	Freiburg
Pat 18	Frontal and Temporal	8	M	13	Freiburg
Pat 19	Frontal and Temporal	28	M	5	Freiburg
Pat 20	Frontal and Temporal	47	M	5	Freiburg
Pat 21	Frontal and Temporal	4	F	4	Freiburg
Pat 22	Temporal	33	M	5	Paris

7.2 NFLE Seizure Detection System

The aim of the first proposed detection system is to early detect sleep seizures of the specific patient affected by NFLE, in order to prevent accidents and minimize the probability of injuries [1]. The detection problem has been formalized as a two-class classification problem where the two classes correspond to seizure (SZ) or non-seizure (NS) states during the sleep.

Note that, in the available database, the scoring of the sleep macrostructure has been provided only for the seizure periods. Hence, in order to characterize the NFLE seizure state, only the recordings referring to seizures during the sleep have been considered. However, for non-seizure recordings, no information is available if the patient is sleeping or awake. In order to characterize the NS state, in this thesis, a hypothesis has been made that the patient is sleeping during the ten minutes preceding a seizure in his sleep. Therefore, only the recordings referring to ten minutes preceding the seizures in the sleep have been considered to characterize the non-seizure state.

Several patients affected by NFLE have been considered (the first nine in Table 7.1) and a detection system has been optimized for each of them to investigate the robustness of the proposed approach.

The recordings from eleven EEG channels have been selected for each patient: three central channels (C3, C4, Cz); two frontal channels (F3, F4); two fronto-polar channels (Fp1, Fp2); two parietal channels (P3, P4), and two occipital channels (O1, O2). As inputs to the detection system, two different sets of EEG channels have been considered: the first seven channels, which record the brain activity in the frontal region, and all eleven channels in order to take into account the possibility of the propagation of the seizures in the parietal and occipital regions.

Preprocessing and Feature Extraction

Firstly, the EEG signals have been preprocessed. The 50 Hz power line noise have been filtered and then EPILAB [149] has been used to extract features from the filtered EEG signals.

EPILAB is an extensive software application, efficient and user-friendly, in Matlab® environment, freely downloadable from http://epilepsiae.eu/DownloadEpilab, to support research in epilepsy, namely seizure prediction and detection, features extraction, etc., on long-term EEG/ECG recordings. It is a product of the European project *EPILEPSIAE*.

Sleep staging is usually scored by doctors using 30-second epochs, in [15] changing the window to 10 s is suggested to identify clinical interictal activities and paroxysmal changes

in EEG tracings. As NFLE seizures can last around 5s [150], in this study features have been extracted on a non-overlapping 5s-sliding window.

Note that, in this thesis, each 5-s sliding window is named as *Epoch*, as usually ascribed in clinical EEG analysis.

For each epoch, the relative band power (RP) in the following frequency sub-bands has been computed:

- delta [0.1-4 Hz];
- theta [4-8 Hz];
- alpha [8-15 Hz];
- beta [15-30 Hz];
- gamma [30-Nyquist frequency Hz].

The relative band power has been calculated using an autoregressive model to evaluate the power spectral density, as suggested by the EPILAB software.

Moreover, the energy of wavelet coefficients (WE) of six decomposition levels has been calculated, using the db4 as basis wavelet function.

As previously reported in chapter 6 (§6.1), the wavelet transform is widely used in seizure detection systems, given the non-stationary nature of the EEG signals.

In this thesis the WEs have been firstly used to construct 2 different training input matrices, depending on the number of channels (7 or 11), whose number of rows is the total number of selected SZ and NS epochs. Then, the RPs and their combination with WEs have been considered as features, with the aim to improve the results.

The training epochs have been selected from about half of the seizures and from the preceding 10 minutes. In order to train the classification model with correct NSs information and avoid possible misclassification due to the non-overlapping 5s sliding window, the five seconds before the seizure onset has been assumed to be part of the seizures. The number of columns is the product of the number of feature and the number of selected channels.

In Table 7-2, the numbers of training and test nocturnal epileptic seizures and the number of corresponding epochs have been reported for each patient.

Table 7-2 Training and Test set

Patient	Number training seizures	Number training Epochs	Number test seizures	Number of test Epochs
Pat 1	2	279	3	391
Pat 2	3	383	4	507
Pat 3	1	129	2	264
Pat 4	4	511	5	635
Pat 5	2	256	3	385
Pat 6	4	576	4	537
Pat 7	1	129	1	126
Pat 8	1	141	1	131
Pat 9	7	980	8	1086

Feature Reduction

In order to find the most important features, the Principal Component Analysis (PCA) [58] has been applied to the input matrix as feature reduction method. The choice of the components has been assessed considering a cut-off value $l^* < 1$, as suggested in [75]. In this study two values, $l^* = 0.8$ and $l^* = 0.5$ have been empirically considered.

Each feature vector has been normalized in the interval $[0 - 1]$.

Building Model

A SOM has been used as classification model. The SOM Toolbox 2.0 for Matlab [91] has been used to train the SOM classification models, and the Euclidean distance has been selected.

The neighborhood function determines how strongly the neurons are connected to each other, influencing the training result of SOM procedure. The considered four different neighborhood functions are: Bubble, Gaussian, Cut-Gaussian, and Epanechnikov functions, as described in the chapter 5 (§5.4.6).

The map dimensions (K) have been chosen according to the number of training epochs; the four map dimensions considered are:

- $K = 5 \cdot N^{0.54321}$, as suggested in the toolbox, where N is the number of training epochs.
- K = 80;
- K = 100;

- K = 225.

As previously described, each SOM has been trained with the epochs coming from the first half of the seizures and the associated 10 minutes of NS epochs before the seizure onset, as reported in Table 7-2. The remaining seizures and the related NS epochs have been used as test cases.

The seizure (SZ) or non-seizure (NS) label associated to each epoch can be used to identify four main categories of neurons in the SOM, depending on their composition:

- No seizure neurons: which contain only the NS epochs. These neurons have been colored in blue;
- Seizure neurons: which contain only seizure (SZ) epochs, which have been colored in red;
- Mixed neurons: which contain both NS and SZ epochs. They have been colored in grey;
- Empty neurons: which do not contain any epoch, have been colored in white.

Then, a post-processing approach has been applied, with the aim to count also for neurons that contained a significant number of epochs of the same class, recoloring the mixed neurons. In particular, the neurons that contain more than 80% of epochs of the same class (NS or SZ) have been reassigned to that class.

As an example, one of the SOM maps of the Patient 4, is reported in Fig. 7-1.

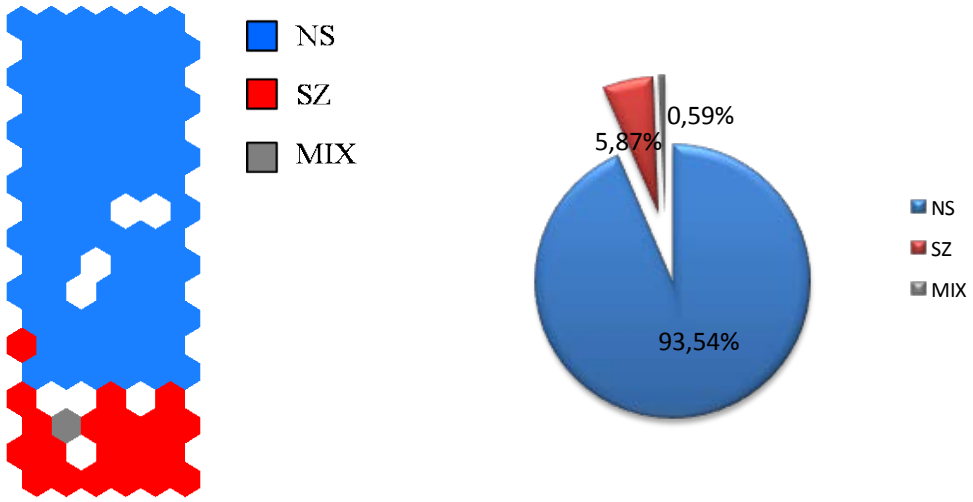


Fig. 7-1 SOM output of patient 4: a) the map colored on the basis of the clusters type (blue NS, red SZ epochs, grey mixed, white empty neurons); b) map composition in terms of epochs into the clusters, colored on the basis of the clusters type (blue NS, red SZ epochs, grey mixed).

The 2-D SOM in Fig. 7-1a clearly highlights the presence of a large blue region with an associated low risk of seizure, a red region, with a high risk of seizure, well separated from the blue region, a unique mixed cluster, and few empty clusters all over the map. The SOM composition is reported in Fig. 7-1b in terms of epochs into the neurons. As it can be seen, non-seizure blue neurons contain the 93.54% of the total epochs, the seizure region contains the 5.87% of the total epochs. Only 0.42% of NS epochs and 3.23% of SZ epochs belong to the mixed neuron.

Results and Performance

For each patient, 96 SOM networks have been initially evaluated, which differ for the number of EEG channels (7 or 11) from which the WE features have been extracted, the PCA pre-processing method, with the two cut-off values (it can be either not considered), the four neighborhood functions, and the four map dimensions.

The potentiality of the available toolbox for the SOM suggests the possibility to track the temporal sequence of the 5s sliding windows (i.e., epochs) on the map, depicting the movement of the operating point during the time. Following the trajectory in the SOM, it will be possible to eventually recognize the proximity to an operational region where the risk of an imminent seizure is high. In Fig. 7-2, the trajectories of an EEG test record on one the SOM map of Patient 1 is reported as a bold line. The neurons framed in black and yellow indicate where the starting and the ending operating points are projected, respectively. As it can be noted, the recording starts in a NS (blue) neuron, crosses a mixed neuron, and arrives in a SZ (red) neuron. Hence, using the SOM to display the operational states of the EEG records, it is possible to discriminate among different brain activities.

To better understand the temporal evolution of the trajectory onto the map, a bar graph is depicted at the bottom of the Fig. 7-3. For each epoch, the bar indicates the neuron where the operating point is projected. A blue bar corresponds to a NS neuron, a red bar to a SZ neuron, a white bar corresponds to an empty neuron. If the operating point is projected onto a grey mixed neuron, the corresponding bar is colored in proportion to the epochs (NS and SZ epochs) contained into such neuron.

Similarly, in the upper bar diagram in Fig. 7-3, the real evolution of the EEG is reported: each bar is colored on the basis of the label manually assigned by the neurologists: a blue bar represents a NS epoch, whereas a red bar characterizes a SZ epoch.

By comparing the bars corresponding to the same time window a very good correspondence is found. Moreover, few red bars before the real seizure onset are present suggesting the presence of seizure precursors before the neurologists reporting.

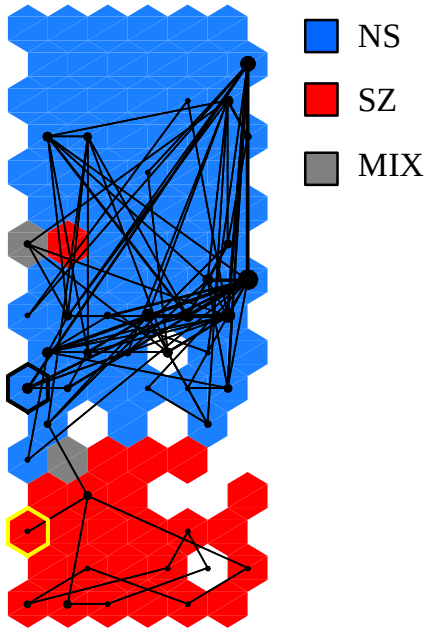


Fig. 7-2 SOM map of patient 1, coloured on the basis of the neurons type (blue NS, red SZ epochs, grey mixed, white empty neurons). The bold line is the trajectory of an EEG test record, the neurons framed in black and yellow indicate where the starting and the ending operating points are projected, respectively.

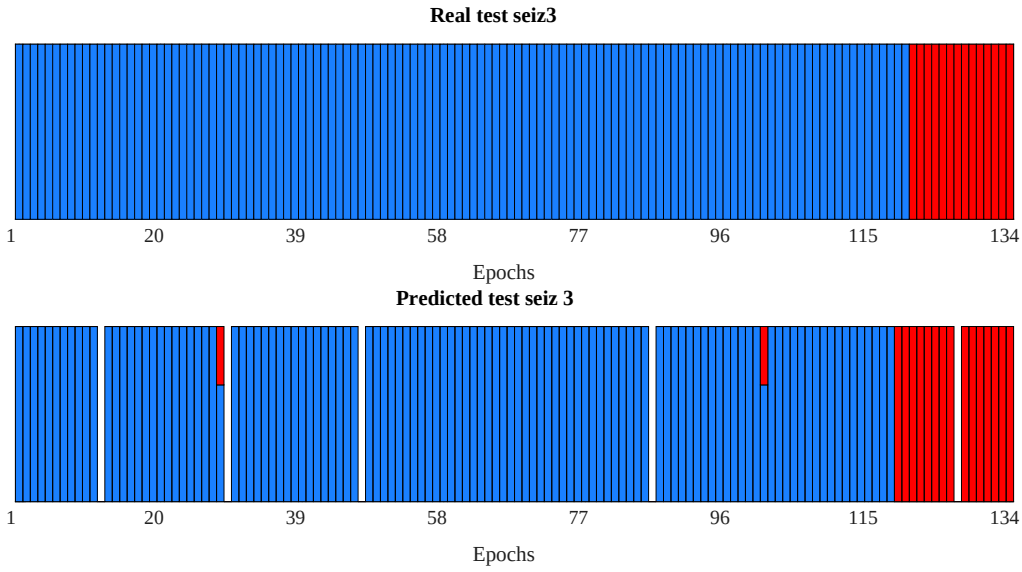


Fig. 7-3 Top of the figure - the real test containing the first 10 minutes of NS (blue) and then the SZ (red). Bottom of the figure: Bar graph (with the same colour code of the SOM in Fig. 7-2) corresponding to the clusters on which the operating point is progressively projected.

In order to evaluate the suitability of the method to detect seizures, for each patient, all the test records have been projected onto the 96 SOMs. During the signal evolution, an epoch is considered correctly classified if the color in the corresponding bars (real and predicted) is the same.

To select the better SOM for each patient and balancing the values of sensitivity (SS) and specificity (SP), a global performance index has been considered [151], which is the geometric mean of SS and SP ($Gmean$):

$$Gmean = \sqrt{SS \cdot SP}$$

Where SS is the ratio between the SZ epochs correctly classified by the SOM and the sum of all SZ epochs, whereas the SP is the ratio between the NS epochs correctly classified by the SOM and the sum of all NS epochs. The maximum value of $Gmean$ was used to select the best SOMs.

The best networks, selected for each patient, in terms of the associated value of the maximum $Gmean$, sensitivity and specificity values, using seven and eleven channels are reported in Table 7-3, respectively. The inclusion of four EEG channels localized in the posterior hemisphere increased the average value of $Gmean$, evaluated for all patients, from 77.39% to 78.40%.

It can be noted that, by comparing the performances obtained involving 7 and 11 EEG channels, Patient 1, Patient 3 and Patient 5 reach higher performance in terms of $Gmean$ and specificity values using 11 channels.

Additionally, the best networks of Patient 2, Patient 6 and Patient 9 show good performance of $Gmean$ and sensitivity but lower values of specificity compared to those obtained using 7 EEG channels.

Despite the good results are achieved with eleven EEG channels, Patient 4 reaches better performances of sensitivity and specificity using the seven EEG channels located in the anterior hemisphere.

Table 7-3 Best network per patient using the energy of wavelet coefficients

Patient	7 EEG channels			11 EEG channels		
	Best Gmean	SS	SP	Best Gmean	SS	SP
Pat 1	87,85%	93,55%	82,50%	89,75%	93,55%	86,11%
Pat 2	80,41%	70,37%	91,88%	81,37%	74,07%	89,38%
Pat 3	77,46%	66,67%	90,00%	79,41%	66,67%	94,58%
Pat 4	83,59%	71,43%	97,83%	79,91%	65,71%	97,17%
Pat 5	66,53%	48,00%	92,22%	72,19%	56,00%	93,06%
Pat 6	85,58%	75,44%	97,08%	90,57%	85,96%	95,42%
Pat 7	81,65%	66,67%	100,00%	81,31%	66,67%	99,17%
Pat 8	62,89%	54,55%	72,50%	60,30%	45,45%	80,00%
Pat 9	70,58%	52,38%	95,10%	70,81%	60,32%	83,13%

However, the performances obtained for some patients do not show a good balance between sensitivity and specificity values. As an example, the best network of Patient 5, using 11 EEG channels, achieves a sensitivity and specificity equal to 56.00% and 93.06%, respectively.

In order to increase and balance the performances, both the RP and WE features have been considered, together with the combination of these two.

Summarizing, eighteen different simulation scenarios have been considered, which differ for the extracted features (WE, RP, and WE & RP together), for the number of the input EEG channels (eleven or seven), and for the feature reduction method (no-one, and PCA with the two cut-off values). In Table 7-4, the eighteen considered scenarios are summarizes.

Hence, for each patient, 288 SOM networks have been built, i.e., for each of the 18 scenarios, four different neighborhood functions, and four different map dimensions were calculated.

Table 7-5 shows the best SOM selected for each patient, in terms of the associated value of $Gmean$, the scenario number according to Table 7-4, the sensitivity and the specificity values.

For the same best SOMs, the related neighborhood function and map dimension, are reported in Table 7-6. From these tables it can be noted that the best networks are obtained with a large variety of scenarios and neighborhood functions, whereas the map dimension equal to 100 is predominant among them.

Table 7-4 Investigated scenarios

	Features	Number of channels	Pre-processing
1.	WE	11 channels	
2.	WE	11 channels	PCA (l^*=0.8)
3.	WE	11 channels	PCA (l^*=0.5)
4.	WE	7 channels	
5.	WE	7 channels	PCA (l^*=0.8)
6.	WE	7 channels	PCA (l^*=0.5)
7.	RP	11 channels	
8.	RP	11 channels	PCA (l^*=0.8)
9.	RP	11 channels	PCA (l^*=0.5)
10.	RP	7 channels	
11.	RP	7 channels	PCA (l^*=0.8)
12.	RP	7 channels	PCA (l^*=0.5)
13.	WE & RP	11 channels	
14.	WE & RP	11 channels	PCA (l^*=0.8)
15.	WE & RP	11 channels	PCA (l^*=0.5)
16.	WE & RP	7 channels	
17.	WE & RP	7 channels	PCA (l^*=0.8)
18.	WE & RP	7 channels	PCA (l^*=0.5)

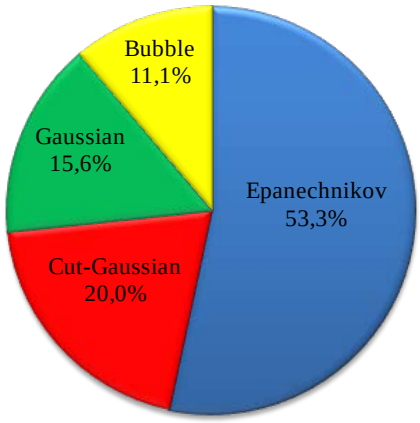
Table 7-5 Best network per patient; Scenario from Table 7-4

Patient	Best Gmean	Scenario	SS	SP
Pat 1	89,75%	1	93,55%	86,11%
Pat 2	81,37%	2	74,07%	89,38%
Pat 3	84,41%	16	79,17%	90,00%
Pat 4	85,24%	4	71,43%	97,83%
Pat 5	79,71%	7	76,00%	83,61%
Pat 6	90,57%	2	85,96%	95,42%
Pat 7	95,31%	15	100,00%	90,83%
Pat 8	67,76%	9	54,55%	84,17%
Pat 9	70,81%	1	60,32%	83,13%

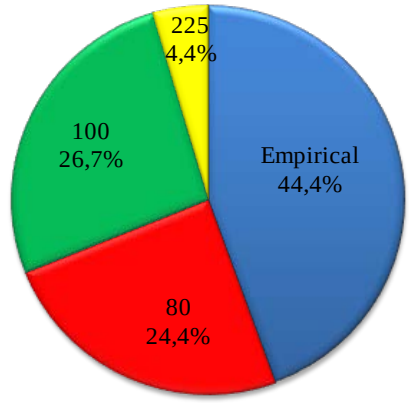
Table 7-6 Neighbourhood function and map size for each best network

Patient	Neighborhood Function	Map Size
Pat 1	Cut-Gaussian	100
Pat 2	Epanechnikov	100
Pat 3	Gaussian	100
Pat 4	Cut-Gaussian	108
Pat 5	Gaussian	100
Pat 6	Bubble	100
Pat 7	Cut-Gaussian	54
Pat 8	Epanechnikov	63
Pat 9	Bubble	225

However, by considering the first best five networks with the associated neighborhood functions and map dimensions, the Epanechnikov function and the empirical formula of the map dimension suggested by the toolbox prove to be the common parameters, with a percentage of 53.3% and 44.4%, respectively, as reported in Fig. 7-4.



a.



b.

Fig. 7-4 a) Percentage of the neighbourhood function used in the first five best SOMs; b) Percentage of the map dimensions used in the first five best SOMs.

As an example, the best SOM, reported in Table 7-5, of Patient 1 achieves a maximum $Gmean$ value equal to 89.75%, obtained training the map with the energy of wavelet coefficients of 11 channels, applying the Cut-Gaussian as neighborhood function, and a map dimension of 100 neurons. The testing performances of 93.55% and 86.11% for SS and SP respectively, were obtained training the map with two seizures and using three seizures as test.

An SS equal to 77.23% and an SP equal to 88.94% are obtained as average values for all patients. The average value of SS sharply increases of the 16.03% and 13.12%, compared to those achieved using WEs considering 7 and 11 channels, respectively; whereas, the average value of SP slightly decreases of 2.28% and 2.14%, respectively. Note that, for all the patients but two, the use of eleven channels gathered the best results proving the beneficial contribution of the signals coming from parietal and occipital channels in the detection of nocturnal frontal lobe epilepsy.

Test projections of Pat 4 on the SOM are displayed in Fig. 7-5 and Fig. 7-6, in the form of bar graphs. Both NS epochs and SZ epochs are identified with great success: in Fig. 7-5, NS epochs are detected without misclassification, whereas a very limited number of misclassified bars are depicted in Fig. 7-6. The red bars before the seizure onset, suggest a variation of the temporal evolution of the signal, showing an early detection, with a few seconds of prediction.

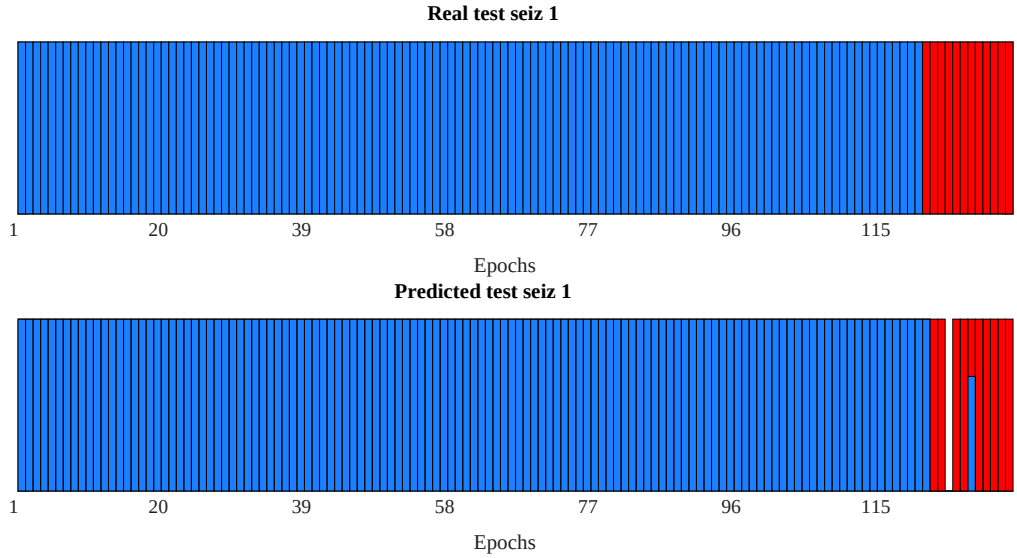


Fig. 7-5 Example of the test seizure 1 of patient 4: top of the figure - the real test containing the first 10 minutes of NS (blue) and then the SZ (red), bottom of the figure - prediction of the SOM map corresponding to neurons on which the operating point is progressively projected.

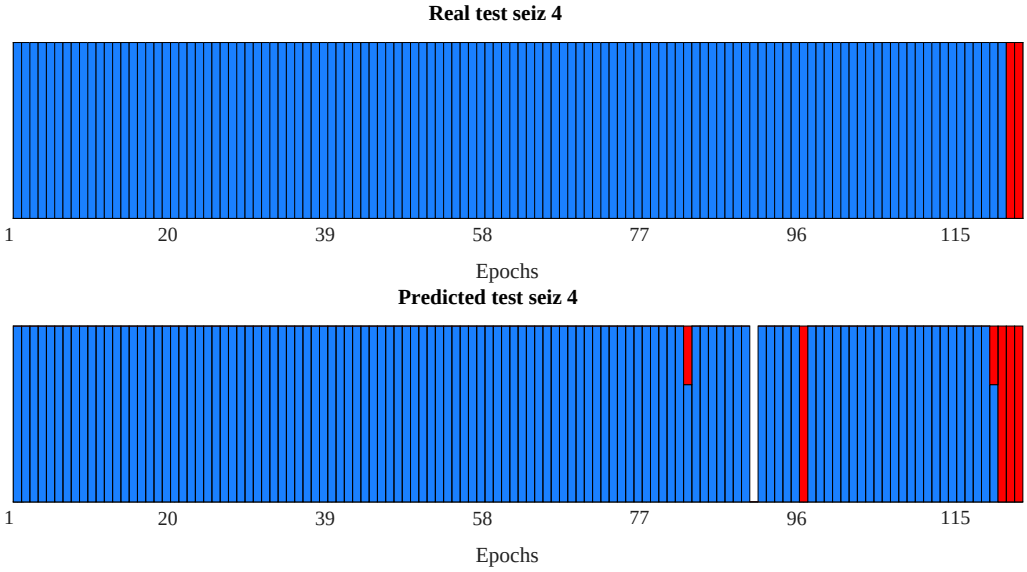


Fig. 7-6 Example of the test seizure 4 of patient 4: top of the figure - the real test containing the first 10 minutes of NS (blue) and then the SZ (red), bottom of the figure - prediction of the SOM map corresponding to neurons on which the operating point is progressively projected.

Conclusions

The results presented in this section demonstrated the suitability of the proposed approach to develop a patient customized NFLE Seizure Detection Systems. Several simulations were performed in order to find the best combination of features to be extracted from EEG signals and map parameters.

The investigation among 288 scenarios for each of the nine patients, determined a classification system capable to detect with good accuracy the nocturnal frontal lobe seizures, obtaining a mean value of 77.23% (between 54.55% and 100%) and 88.94% (between 83.13% and 97.83%) of sensitivity and specificity, respectively.

The best networks show a different scenario for each patient, except in the case of the number of channels, where 7 out of 9 patients achieve good performance training the network with 11 EEG channels, whereas the choice of features can improve the results. Hence, each patient presents a proper combination of features and network parameters to achieve the best metrics, confirming the choice to develop a patient-customized system.

Moreover, the temporal evolution of the projection of the test operating points onto the map demonstrated the possibility to early detect the seizures, underlining the possibility to use the combination of selected features as precursor of the seizures.

7.3 Nocturnal Epileptic Seizure Detection System

In order to further validate the machine learning approach proposed in the previous section, all the 22 patients affected by nocturnal epilepsy, recorded in the European Epilepsy Database, have been considered. As previously cited, they are affected by epilepsy with different focal position (as reported in Table 7-1). Additionally, the performances of the patient-customized SOM-based systems have been compared with those of other detection systems, which use a SVM as classification model. The features and model parameters involved in the systems have been optimized, thereby comparing the best performances for each patient.

As the patients are affected by frontal and/or temporal nocturnal epilepsy, for each patient nine EEG channels have been selected from those available in the database according to the information given by neurologists about the location of the seizures. In the previous section, good performances have been reached using 11 EEG channels. In this section, a reduced number of customized signals are investigated, with the aim to minimize discomfort to patients due to avoidable electrodes. In fact, for each patient, the seizures may be located in different positions; hence, a different set of channel could maximize the performance of the detection system with a lower number of EEG channels than the previous approach. In Table 7-7 the selected EEG channels for each patient are reported.

Table 7-7 Selected channels for each patient.

Patient	EEG channels								
Pat 1	C3	F3	F4	C4	T7	T8	P8	FT7	FT8
Pat 2	C4	F4	FZ	F8	FP2	FPZ	F7	CZ	FP1
Pat 3	FT10	F7	F8	C3	C4	P3	P4	FT9	F4
Pat 4	TP9	F7	T5	FT9	PZ	T9	TP10	F3	C3
Pat 5	C4	P4	F8	FT8	T8	P8	TP8	CZ	AF8
Pat 6	C3	C4	CP1	CP2	P4	F4	PZ	PO2	PO1
Pat 7	F8	F4	C4	TP10	CZ	FT10	P4	T10	FP2
Pat 8	F7	F3	FP2	F4	FT9	C3	FZ	TP10	T9
Pat 9	F4	F7	C3	P3	O1	F8	P4	F3	FP2
Pat 10	C3	F7	T5	FZ	F3	SP1	CZ	C4	T6
Pat 11	F7	F8	FP2	C4	F4	FP1	C3	F3	CZ
Pat 12	FP1	F7	T7	P7	F3	C3	P3	O1	FT7
Pat 13	T8	O2	FT8	F8	T10	AF8	P4	P8	C4
Pat 14	C4	F8	SP2	T6	FZ	FP2	O2	P4	F4
Pat 15	F7	CP1	CP5	F3	C3	FZ	PZ	TP7	T5
Pat 16	FP1	F3	F7	T5	T3	SP1	O1	C3	P3
Pat 17	F3	C3	FZ	CZ	FP1	FP2	PZ	O2	F7
Pat 18	C3	C4	T5	T6	PZ	FP1	FP2	F4	FZ
Pat 19	F4	F8	C4	FZ	CZ	T6	FP2	C3	O2
Pat 20	FZ	CZ	PZ	T5	F8	T6	P4	C4	F7
Pat 21	F7	T5	SP1	CP5	C3	FC5	CP1	FC1	FZ
Pat 22	F8	T6	FT10	T10	T4	C4	O2	FZ	CZ

Preprocessing and Feature Extraction

EEG signals have been preprocessed with a notch filter to remove the 50 Hz power line interference.

As in the previous case study, two types of features, the relative band power (RP) and the energy of the wavelet coefficients (WE), have been extracted from a 5s sliding window without overlapping, using the EPILAB software [149].

The RPs have been calculated for the delta [0.1-4 Hz], theta [4-8 Hz], alpha [8-15 Hz], beta [15-30 Hz] and gamma [30-Nyquist frequency [Hz] sub-bands; whereas, the WEs have been calculated using the db4 as wavelet function at the six level of decomposition.

Then, EEG recordings have been divided into four stages:

- inter-ictal: considered as the "normal" brain state, free from seizures;
- pre-ictal: the period before the occurring of the seizure;
- ictal: the seizure period;
- post-ictal: the interval after the seizure and before the inter-ictal.

In this study, the pre-ictal and post-ictal stages are both assumed to last 10 minutes, whereas the ictal stages is considered including the 5s window before the seizure onset, in order to avoid possible misclassifications due to the non-overlapping window.

In order to find significant differences between seizure regions, the cross information is evaluated. Each new feature is evaluated as the ratio of the features between two different frequency sub-bands belonging to the same or different EEG channel. This feature, also applied for predicting epileptic seizures [121], gives the cross information not just between two channels but also between two frequency sub-bands.

The new features have been evaluated for both the extracted features (RP and WE). The new approach leads to an increase of the number of features that depends on the combination of the considered EEG channels and frequency sub-bands. Assuming v as the product of the m EEG channels and n frequency sub-bands, the dimension of the new feature set is:

$$N = \frac{v!}{2!\,(v-2)!}$$

As an example, the RPs are evaluated in 5 frequency sub-bands from 9 EEG channels, the set is composed by 990 new features.

Feature Preprocessing and Feature Reduction

In order to limit the effects from noise and interfering sources [121], each feature has been smoothed with the moving average, using a window of 12 epochs, i.e., one minute length.

Successively, the inter-ictal and post-ictal epochs have been removed, retaining the pre-ictal (named as NS) and ictal (named as SZ) epochs for this study. Hence, each feature has been normalized with zero mean and unity standard deviation.

To reduce the high number of the new features, in this second approach, a different feature reduction method has been adopted, employing the minimum Redundancy maximum Relevance (mRMR) approach [80]. The mRMR is suitable to extract a subset of features that have the minimal redundancy among them and that are more representative for the classification. To evaluate the redundancy and relevance, for discrete variable, the mutual information has been applied, and the maximization of the ratio between relevance and redundancy has been chosen in this work, as suggested by [80] for the case of discrete variables.

The mRMR software packages (home.penglab.com/proj/mRMR), implemented in Matlab®, have been used to discretize and select the subset of features. Eight different subsets have been considered, changing the number of the selected features from 5 to 40, with a step of 5, with the goal to find the appropriate subset that achieves the best performance, reducing up to 99% the dataset dimension for both the extracted features (RP and WE).

Building Models

The SVM and SOM classifiers have been developed in order to distinguish between NS and SZ epochs. Each classifier has been trained with the first half of the SZ and the related NS epochs.

The SVM classifier has been used to distinguish between the two classes of NS and SZ epochs and the regularization parameter has been optimized to obtain good classification performances.

For each patient, 16 classification models have been calculated, i.e., for both the extracted features (RPs and WEs), 8 subsets have been evaluated. Hence, the best combination of number and type of features has been selected based on the maximum value of $Gmean$.

The SOM has been calculated with the SOM Toolbox 2.0 for Matlab®, using the Euclidean distance. In the previous approach, the Epanechnikov function, as neighborhood function, and the empirical formula of the map dimension suggested by the SOM toolbox, proved to

be suitable training network parameters, in this second approach, these network parameters have been fixed to train the SOMs

Then, the post-processing method has been performed, as well as the previous case study: the neurons containing more than 80% of epochs of the same class (NS or SZ), have been reassigned to that class.

In this work, the SVM has been trained using the LibSVM library package [149], implemented in Matlab®, that enables the selection of different SVM parameterizations such as the kernel type (linear, polynomial, radial basis function or sigmoid) and the value of the regularization parameter (cost). The Radial Basis Function (RBF) kernel has been used in this application.

The tuning of the free parameters of the SVM, i.e., C (the cost) and γ (the parameter of the RBF) has been performed using a grid search; C was varied in the range $[2^{-3} - 2^{8}]$ with a step of 2^{2}, γ in the range $[2^{-4} - 2^{3}]$ with a step of $2^{0.5}$.

To validate the model and to choose the best parameters, the 10-fold cross-validation [152] has been employed. The k-fold cross validation consists of separating the training dataset into k random partitions, equally distributed in size and class proportion. Then, among the k subsets, one is retained as the validation data for testing the model, and the remaining k-1 subsets are used in the training phase. This process is repeated k times and the resulting performance in terms of accuracy is the average of the k performances.

The best two parameters, which maximize the average accuracy, have been adopted to build the model with the entire training set

Results and Performance

For each classifier and for each patient, the test epochs have been performed on the sixteen networks. As well as the training set, the features have been extracted from a 5s non-overlapping sliding window and the cross-information has been calculated. Consequently, the selected features have been tested to the SOM and SVM classifiers.

Finally, the performances are evaluated in terms of sensitivity and specificity in order to evaluate the $Gmean$ value, and obtain the best combination of features that achieves good performances.

In Table 7-8 and Table 7-9 the best networks for each patient, selected on the base of the maximum value of $Gmean$, are reported for the SOM and the SVM models, respectively.

For each best network, the tables show the associated best features, number of features and training parameters of each model, as well as the sensitivity and specificity values obtained.

Table 7-8 Best SOM network per patient.

Patient	Best Gmean	Features	Number of features	Map dimension	SS	SP
Pat 1	87,55%	RP	10	80	93,55%	81,94%
Pat 2	81,04%	WE	5	98	70,37%	93,33%
Pat 3	78,35%	RP	40	56	70,83%	86,67%
Pat 4	80,78%	WE	25	108	68,57%	95,17%
Pat 5	86,02%	RP	10	80	80,00%	92,50%
Pat 6	82,03%	WE	15	120	71,93%	93,54%
Pat 7	**98,32%**	**WE**	**10**	**57**	**100,00%**	**96,67%**
Pat 8	67,53%	WE	40	66	63,64%	71,67%
Pat 9	82,92%	WE	20	160	85,71%	80,21%
Pat 10	79,28%	WE	30	56	85,71%	73,33%
Pat 11	94,47%	WE	15	56	90,00%	99,17%
Pat 12	81,98%	WE	15	63	72,00%	93,33%
Pat 13	88,83%	WE	15	64	89,74%	87,92%
Pat 14	77,65%	RP	35	119	64,62%	93,31%
Pat 15	93,72%	RP	30	84	92,86%	94,58%
Pat 16	87,72%	WE	15	102	77,59%	99,17%
Pat 17	84,10%	WE	35	84	76,92%	91,94%
Pat 18	90,98%	RP	30	144	88,35%	93,69%
Pat 19	95,85%	WE	40	80	96,15%	95,56%
Pat 20	93,39%	RP	10	80	94,29%	92,50%
Pat 21	82,56%	WE	15	80	81,40%	83,75%
Pat 22	90,56%	WE	10	80	95,24%	86,11%

Table 7-9 Best SVM network per patient.

Patient	Best Gmean	Features	Number of features	SS	SP
Pat 1	90,56%	WE	15	87,10%	94,17%
Pat 2	92,02%	WE	40	92,59%	91,46%
Pat 3	77,06%	RP	5	83,33%	71,25%
Pat 4	87,19%	RP	15	88,57%	85,83%
Pat 5	87,05%	WE	15	88,00%	86,11%
Pat 6	89,84%	WE	35	89,47%	90,21%
Pat 7	**97,47%**	**WE**	**5**	**100,00%**	**95,00%**
Pat 8	70,06%	WE	10	81,82%	60,00%
Pat 9	82,13%	WE	5	88,10%	76,56%
Pat 10	62,30%	WE	5	57,14%	67,92%
Pat 11	73,48%	RP	10	80,00%	67,50%
Pat 12	84,14%	RP	5	72,00%	98,33%
Pat 13	65,83%	WE	10	100,00%	43,33%
Pat 14	83,01%	WE	40	73,85%	93,31%
Pat 15	92,30%	WE	40	89,29%	95,42%
Pat 16	83,53%	RP	10	81,03%	86,11%
Pat 17	0,00%	WE	5	0,00%	99,17%
Pat 18	82,80%	WE	10	85,44%	80,24%
Pat 19	95,85%	WE	5	96,15%	95,56%
Pat 20	94,92%	WE	20	94,29%	95,56%
Pat 21	82,97%	RP	5	81,40%	84,58%
Pat 22	79,33%	WE	10	96,83%	65,00%

As mentioned before, the sensitivity is evaluated as the ratio of SZ epochs correctly classified and the sum of all SZ epochs, whereas, the specificity is the ratio of NS epochs correctly classified and the amount of the NS epochs considered.

It can be noted that, the energy of wavelet coefficients is the predominant feature to achieve the best network for both classifiers. Furthermore, for each classifier, two subsets of features prevail: 10 and 15, and 5 and 10 are the common number of features used to achieve the best SOM and SVM network, respectively.

As reported in the tables, for both SOM and SVM networks, the maximum value of $Gmean$ is reached from Patient 7, achieving values equal to 98.32% and 97.47%, respectively. The best networks have been trained with the energy of the wavelet coefficients, using 10 and 5 features for SOM and SVM models, respectively.

To better visualize the output of the models, for each patient, the best $Gmean$ obtained in the previous approach, whose line is colored in black, and the best $Gmean$ of SOM (colored in orange) and SVM (colored in cyan) networks, achieved in the present approach, are depicted in Fig. 7-7.

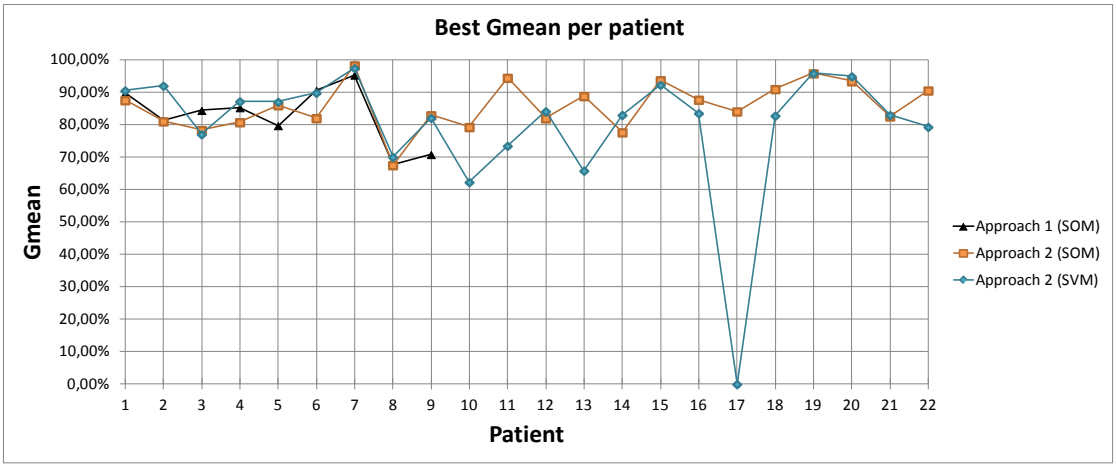


Fig. 7-7 Comparison between the performances, corresponding to the best Gmean per patient, of the first approach (black) and the second approach using SOM (orange) and SVM (cyan) models.

Comparing the results of the first nine patients, affected by NFLE, with the previous system, as shown in Fig. 7-7, the performances of the SOM networks are equal or higher than the first system except for three patients (Patient 3, Patient 4 and Patient 6), that can be due to the choice of the selected EEG channels combined with the parameters of the SOM models. On the other hand, the SVM networks present higher performance than the first system except for Patient 3.

It is noticeable that, the best SVM network for Patient 17 reaches the lowest value, not recognizing any seizure epochs and obtaining a SS equal to 0%, as reported in Table 7-9.

Conversely, a $Gmean$ value of 84.10% has been reached by the SOM model, with a SS of 76.92%, confirming its good discriminating capability.

Patient 17 presents a majority of nocturnal seizures during REM stage compared to the other patients. In particular, the first 4 (out of 6) nocturnal epileptic seizures occur during REM stages. The different frequency range of REM stage compared to NREM stage can lead to misclassification and a lack of generalization of the problem using the SVM model, preferring, in this case, the SOM model.

In the present approach, an SS equal to 82.25% and an SP equal to 89.92% have been reached as average values for all patients, using the SOM networks; whereas by applying the SVM model, the average of SS and SP, equal to 82.11% and 82.85%, have been obtained. For both the performance metrics, the SOM manages to accomplish higher values than those of the SVM model, revealing the good ability of the SOM to detect different brain states.

In Fig. 7-8 and in Fig. 7-9 the signal evolution of a test record of Pat 11 for the best SOM and SVM networks, are respectively shown. The top of the figures represents the expected output, whereas the bottom of the figures reports the predicted output of the models. As reported in Fig. 7-8, both NS (in blue) and SZ (in red) epochs are well detected by the SOM classifier; on the other hand, the predicted SVM output, (see the bottom of Fig. 7-9) presents a high number of misclassified epochs.

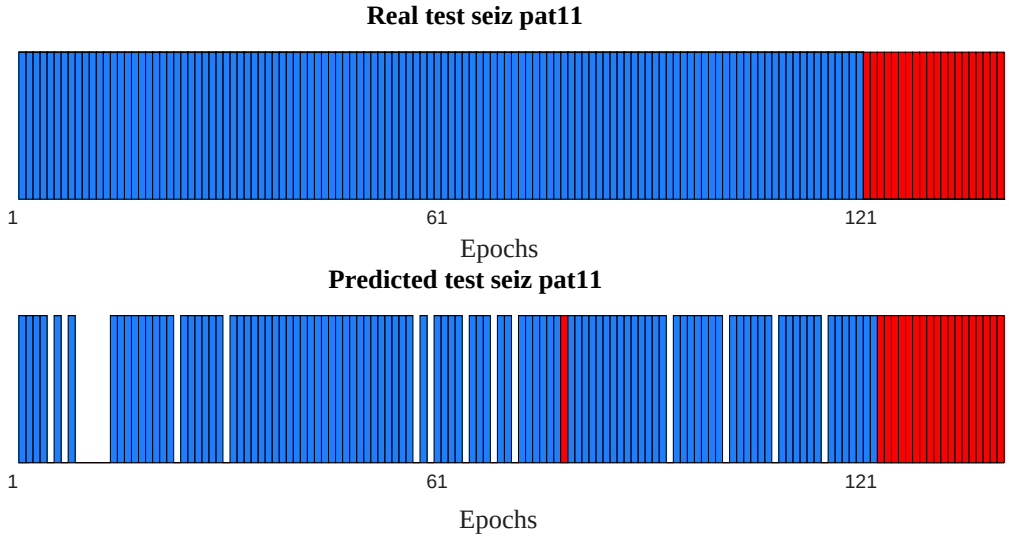


Fig. 7-8 Example of the test seizure of patient 11: top of the figure - the real test containing the first 10 minutes of NS (blue) and then the SZ (red), bottom of the figure - prediction of the SOM map corresponding to neurons on which the operating point is progressively projected.

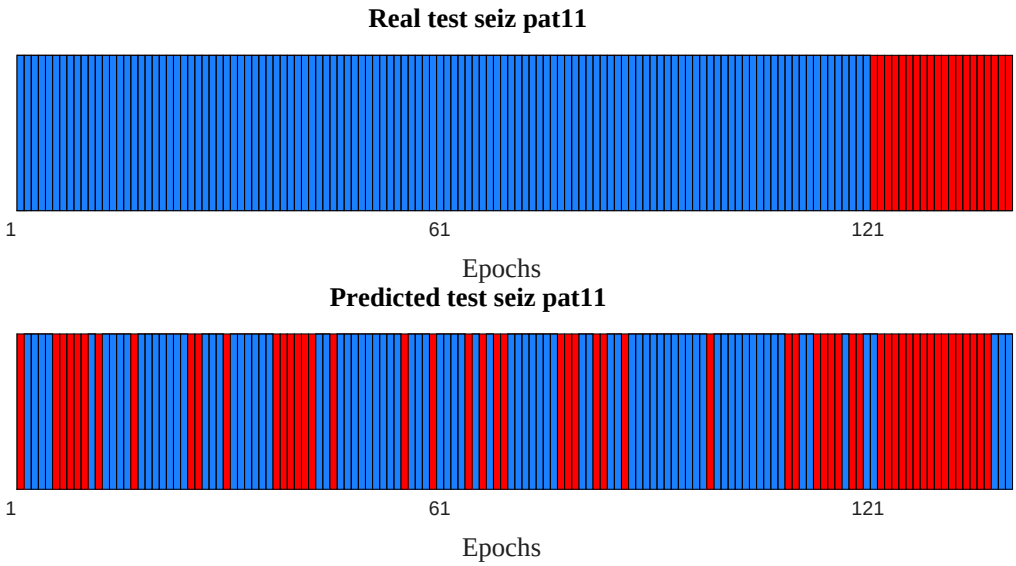


Fig. 7-9 SVM Example of the test seizure of patient 11: top of the figure - the real test containing the first 10 minutes of NS (blue) and then the SZ (red), bottom of the figure - prediction of the SVM classifier.

Conclusions

In this section, the proposed patient-specific approach highlights the good capability of the SOM model, compared to the SVM model, to detect nocturnal epileptic seizures. The SOM and the SVM have been trained and several features and EEG channels have been selected in order to find the best combination for each patient.

For each classifier, sixteen combinations of parameters and features have been examined for each of the 22 patients affected by frontal and/or temporal nocturnal epilepsy.

The balance performances, i.e., the geometric mean between the sensitivity and specificity, have been considered to select the best networks, achieving an average of sensitivity equal to 82.11% (between 0% and 100%) and 82.85% (between 63.64% and 100%) for SVM and SOM classifiers, respectively. Additionally the average of specificity of 82.85% (between 43.33% and 99.17%) and 89.92% (between 71.67% and 99.17%) have been reached for SVM and SOM models, respectively.

The SOM networks show comparable performance than the SVM networks, widely used in epileptic seizure detection. As can be shown in Fig. 7-7, the SOM-based systems demonstrate better performance than the SVM networks for some patients, such as for Patient 11, Patient 13 and Patient 17 whose performances increase by 21%, 23% and 84%, respectively.

Moreover, from the proposed patient-specific systems, the energy of wavelet coefficients is the predominant feature that achieves satisfying performances, confirming the wavelet transform as a suitable feature extraction technique for EEG signals. However, a good combination of features and selected channels is required to obtain good results for each patient. This confirms the proposed approach as a valid patient-customized system for detection of nocturnal epileptic seizures.

Moreover, these results highlight the good performance of the SOM to correctly map the seizure and no seizure epochs.

8 PATIENT-INDEPENDENT SYSTEM FOR NFLE SEIZURE DETECTION

The previous chapter highlighted the good capability of the SOM to detect both frontal and temporal nocturnal epileptic seizures. These very high performances have been obtained by customizing the detection system for each patient.

In this chapter, an investigation is done in order to evaluate the suitability of the approach to develop a patient-independent system.

Since in literature [46]–[48] studies on the sleep microstructure reported that NFLE\ADNFLE typically happens during the phases A3 of the Cyclic Alternating Pattern (CAP), only the A3 phases of the scalp EEG signals provided by the Sleep Centers of Cagliari and Parma, have been considered.

Three sections are presented in this chapter: the first describes the database provided by the Sleep Centers of Cagliari and Parma, the second presents a patient independent classification system to automatically discriminate the nocturnal seizures of patients affected by NFLE, whereas in the third section an improvement of the model has been performed.

8.1 Database - Cagliari & Parma

Data for developing ADNFLE and NFLE seizure detection models and for testing their performances has been provided by the Sleep Centers of Cagliari and Parma.

The database consists of 15 polysomnographic recordings, that include scalp EEG (19 channels, electrodes placed according to the 10–20 International System), electrooculogram (EOG), electromyogram (EMG) of the submentalis muscle, EMG of the right and left tibialis anterior muscle and electrocardiogram (ECG). All subjects underwent one overnight polysomnographic recording, after one adaptation night.

Only the EEG signals were extracted and considered in the present study. All EEG signals were sampled at 256 Hz, a notch filter to remove the 50 Hz power line interference and a low-pass filter at 0-120 Hz to retain the significant frequency components, were applied.

The investigated dataset is composed by the EEG signals of 7 healthy subjects from the Sleep Center of Parma and 8 patients affected by NFLE and ADNFLE from the Sleep Center of Cagliari. It has been divided in two subsets: the control group, represented by the healthy subjects, which comprises 4 females and 3 males ranging in an age from 24 to 52 years, and the patient group, composed by 5 females and 3 males ranging in age from 25 to 34 years.

Three patients are affected by ADNFLE, which is a genetically heterogeneous disorder in which family members can manifest the epileptic events; these patients are indeed relatives: two of them are sisters and the third one is daughter of one of the two.

The information on sleep macrostructure and sleep microstructure were provided by neurologists, according to the sleep macrostructure and microstructure rules defined by R&K [39] and Terzano et al [44], respectively. Additionally, doctors provided information about NFLE/ADNFLE seizures that distinguished in:

- **Major Attacks** (MA), that represent stereotyped movements such as asymmetric tonic or dystonic postures, or other bizarre behaviors lasting about 20-30 s;
- **Paroxysmal Arousals** (PA) whose short attacks lasting about 5 10 s; they are characterized by sudden awakenings, often with dystonic-dyskinetic features;
- **Minor Motor Events** (MME), that include short and stereotyped movements (2 – 4 s), involving the limbs, axial muscle and/or the head.

As mentioned before, some studies have observed an increased amount of the Cyclic Alternating Pattern (CAP) sequences, associated to an epileptic disease [46]–[48]. Hence, the A phases of the CAP, have been extracted and investigated to detect and discriminate epileptic seizures. The A phases were distinguished in the three different subtypes A1, A2, A3. For each subject the composition of the three phases for the control and patient groups are reported in Table 8-1 and Table 8-2, respectively.

In Table 8-3 the amount of all phases, evaluated for the control and the patient groups, is shown. It can be noted that, the A3 phases of the patient group are more than twice the A3 phases of the control group confirming the studies in literature [46]–[48], that ADNFLE/NFLE typically happens in phases A3 of CAP, due to the fact that contain the predominance of EEG desynchronization. The amount of phases A1 and A2 do not exhibit considerably differences from the control and the patient groups; conversely, a focalization on phases A3 and their relation to the crisis, is necessary.

Table 8-1 Phases A of CAP of the control group.

Control group			
Subject	**A1**	**A2**	**A3**
Subj 1	200	101	42
Subj 2	275	76	71
Subj 3	186	55	84
Subj 4	254	117	60
Subj 5	323	122	78
Subj 6	186	111	79
Subj 7	217	69	20

Table 8-2 Phases A of CAP of the patient group.

Patient group			
Patient	**A1**	**A2**	**A3**
Pat 1	207	96	210
Pat 2	252	83	135
Pat 3	423	105	115
Pat 4	206	9	111
Pat 5	283	65	37
Pat 6	59	125	128
Pat 7	269	160	149
Pat 8	136	80	87

Table 8-3 Comparison of the three subtypes A from control and patient groups.

	A1	**A2**	**A3**	
Control group	1641	651	434	**2726**
Patient group	1835	723	972	**3530**

As the neurologists marked only the start time of major attacks, paroxysmal arousals and minor motor events, and do not provide the duration of the seizures, the A3 phases of the patient group have been classified in three classes, labelled as MPA, which contains Major Attacks (seizures) and\or Paroxysmal Arousal; MME, which contains Minor Motor Events, and NS (No Seizures), which does not contain any of the above events. In Fig. 8-1 an example of the A3 phase of a patient, containing both a MA and MME seizures, is reported.

In Table 8-4 the classification of the A3 phases of the patient group, subdivided in NS, MME and MPA labels, are reported. Pat 5 and Pat 7 do not present crisis in phases A3, whereas Pat 6 and Pat 8 present only MME seizures.

Table 8-4 Phases A3 of the group of patients, subdivided in NS, MME and MPA classes.

Patient	NS	MME	MPA
Pat 1	193	14	3
Pat 2	121	12	2
Pat 3	102	4	5
Pat 4	108	1	2
Pat 5	37	0	0
Pat 6	116	12	0
Pat 7	149	0	0
Pat 8	86	1	0

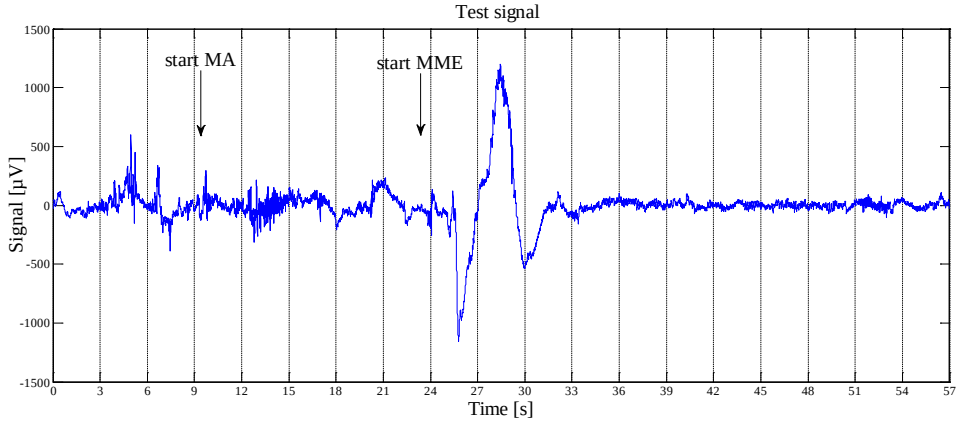


Fig. 8-1 Signal from C4/A1 EEG channel for a patient. A major attack is present, which starts at 9,566 s. Moreover, in the same A3 epoch also a Minor Motor Event appears which starts at 23,316 s.

8.2 NFLE Seizure Detection System

An automatic patient-independent classification system based on the application of the SOM has been developed.

Three patients affected by Autosomal Dominant Nocturnal Frontal Lobe Epilepsy (ADNFLE) from the database *Cagliari&Parma* have been chosen and analyzed with the aim to detect nocturnal seizures in A3 phases.

Preprocessing

In order to verify the suitability of selecting only the A3 phases to characterize ADNFLE, the spectral analyses of the subtypes A1, A2 and A3 are carried out and compared with those of the control group (H).

Only the A phases lasting more or equal than 4s have been considered in order to have a reliable comparison between the A phases, as suggested in [153]. This establishes a reduction of the data set of about 4%, not involving any removal of MME phases, whose durations are greater than 4s.

In order to verify the suitability of selecting only the A3 phases to characterize ADNFLE, the spectral analyses of the subtypes A1, A2 and A3 are carried out.

For each phase, the C4-A1 EEG channel, which is one of the monopolar derivations currently used for the conventional sleep staging and arousal scoring [44], was chosen to evaluate the average Power Spectral Density estimate (PSD). The PSD has been evaluated in the typical EEG frequency range [0÷32] Hz, as suggested in [153], using the Welch's non-overlapped windows of 0.5 s.

In Fig. 8-2 and Fig. 8-3 the PSDs of the A1 and A2 phases are shown, respectively. For both of them, the blue curves, representing the A phases without seizures (NS) of the three patients, are within the standard deviation of the PSD evaluated for the control group (H), depicted in black dotted line. The spectral analysis shows a prominent peak in the low delta range between 0 and 2.5 Hz for A1 subtype for both healthy subjects and patients, not highlighting substantial differences, whereas, the average amplitudes of the spectral components in delta and theta bands in A2 subtypes for the patients are slightly greater than those of healthy subjects.

Conversely, as can be shown in Fig. 8-4, spectra of A3 subtypes show a visible pick in the alpha band, approximately between 7 and 12 Hz, both for control group and patients. The behavior of the PSDs evaluated for NS (blue), MME (green) and MPA (red) phases are outside the range given by the standard deviation of H. This reveals an increase of the energy content compared to the control group, endorsing the choice to retain only the subtype A3 to characterize the ADNFLE EEG behavior. At the same time, in the highest

frequencies [14 – 25 Hz], the PSDs of NS and MME phases exhibit a behavior similar to the PSD of H, explaining the complexity of this form of epilepsy and the problem to automatically deduce a detection model from sample data.

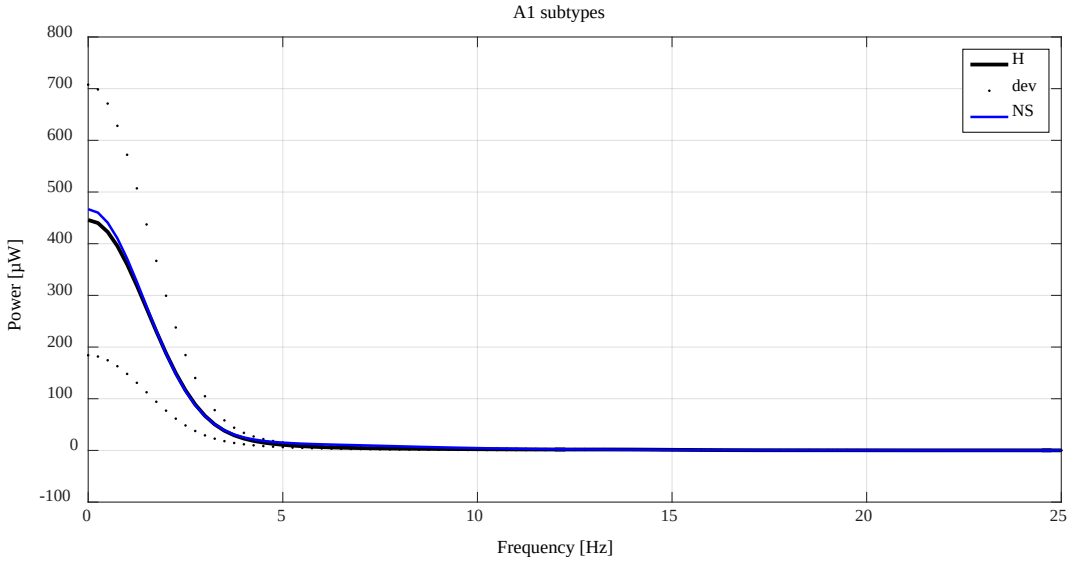


Fig. 8-2.The average of PSDs calculated for each A1 subtype group: Healthy group H (black bold line) and standard deviation of H dev (black dot line), No Seizures group NS (blue line).

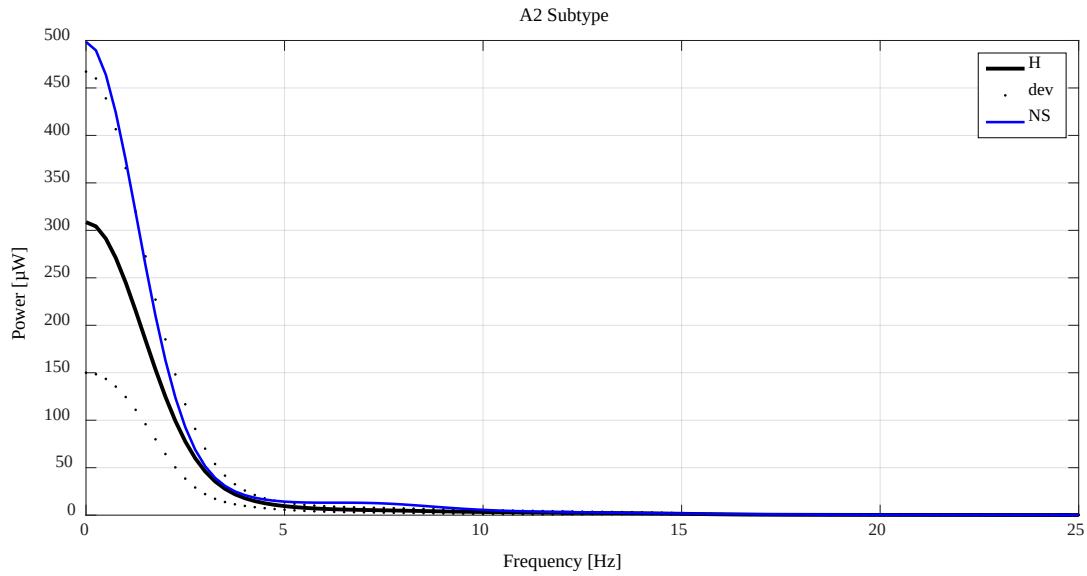


Fig. 8-3.The average of PSDs calculated for each A2 subtype: Healthy group H (black bold line) and standard deviation of H dev (black dot line), No Seizures group NS (blue line).

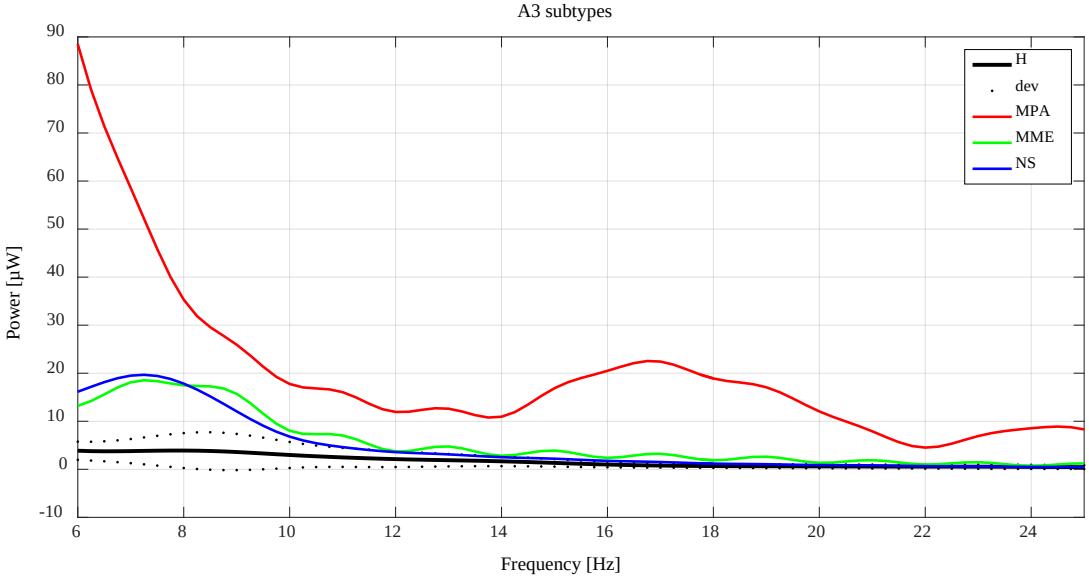


Fig. 8-4 The average of PSDs calculated for each A3 subtype group: Healthy group H (black bold line) and standard deviation of H dev (black dot line), No Seizures group NS (blue line), Major Attacks and\or Paroxysmal Arousal MPA (red line), and Minor Motor Events MME (green line).

Hence, in Table 8-5, the number of A3 phases for the three examined patients, subdivided in NS, MME and MPA phases is reported.

Table 8-5 A3 phases, lasting more than 4s, of the three patients

Patient	NS	MME	MPA
Pat 1	187	14	3
Pat 2	114	12	2
Pat 3	102	4	5
	403	**30**	**9**

Feature Extraction

In order to extract suitable features from the A3 phases to be used as input to the subsequent classification stage, the Wavelet Packet Transform has been carried out to decompose the EEG signals.

The 4th order Daubechies wavelet (db4), considered the best suitable wavelet basis function for the analysis of epileptic EEG data [69], was applied in the present study, at the third level of decomposition, thereby considering a minimum number of samples equal to 128 (0.5s) for each of the eight packets.

Subsequently, for each packet, the standard deviation of the coefficients has been calculated to gather the information on the spectral distribution of the coefficients at different frequency sub-bands. In this way, a unique value for each packet is retained, and an input matrix of dimension $[N \times 8]$ is obtained, where N is the number of A3 phases.

The matrix is the training input applied to build the classification model. To each row of the matrix a label has been associated as the output of the classification, according to the information given by neurologists (NS, MME, MPA and H).

Data Reduction

Being the number of the NS phases much larger than the MME and MPA phases, in this study, a data reduction method has been applied with the aim to balance these numbers.

Data reduction has been performed as proposed in [154]. In particular, for each patient, a SOM has been trained using, as features, the normalized spectral power of the following seven frequency bands: low Delta(<0.5 Hz), high Delta (0.5-4 Hz), Theta (4-8 Hz), Alpha (8-12 Hz), Sigma (12-16 Hz), Beta (16-30 Hz), and Gamma (30-32 Hz). The Sigma frequency band has been considered in order to take into account the contribution of particular sleep patterns as the sleep spindles, which have frequency range from 12 to16 Hz.

Data reduction procedure consists in selecting, for every neuron, the single epoch closest to the SOM centroid, allowing to automatically select a limited and representative number of epochs in the NS set for the subsequent classification. In this case, only 37% of the NS phases of the three patients have been retained, reducing the number from 403 to 148. The number of the NS reduced phases for each patient is reported in detail in Table 8-6.

Table 8-6 Data reduction of NS phases for each patient.

	NS (before data reduction)	NS (after data reduction)
Pat 1	187	**65**
Pat 2	114	**47**
Pat 3	102	**36**
	403	**148**

Building Model

As in the previous applications, the SOM has been used to identify some aggregation or properties of the input patterns. As described in the previous chapter 5 (§5.4.6) the SOM defines a map that preserves the topological properties of the input. This means that points close to each other in the input space are mapped on the same or neighbouring neurons in the output space.

The SOM Toolbox 2.0 for Matlab [91] has been used in this thesis to train the SOM. During the SOM training, the map dimension, i.e., the number of neurons K, has been fixed as the heuristic formula $\mathrm{K} = 5 \cdot \mathrm{N}^{0.54321}$ suggested by the toolbox, where N is the number of the A3 phases.

The resulting map has eight inputs and 12×6 neurons, as shown in Fig. 8-5. Each color in the map is representative of a particular composition of the neuron in terms of MME, MPA and NS phases.

In particular:

- NS neurons (blue) contain only A3 phases with no seizures,
- MME neurons (green) contain only A3 phases with minor motor events,
- MPA neurons (red) contain only A3 phases with major attacks and\or paroxysmal arousal,
- MME+MPA neurons (magenta) contain both MME and MPA phases,
- MIX neurons (grey), contain both NS and MME phases, or NS and MPA phases, or all the three (NS, MME, and MPA phases).

- EMPTY neurons (white), do not contain any phase.

The lower the number of the MIX neurons, the greater is the discrimination capability of the SOM. In Fig. 8-5, two different representations of SOM are reported: in Fig. 8-5a, the SOM is colored on the basis of the neurons type whereas in Fig. 8-5 b, it is colored on the base of the density of the different classes in the neuron.

The map clearly highlights some macro-regions: a large cluster of blue neurons in the top of the map and smaller clusters of red, green and magenta neurons. Moreover, a mixed grey region, where NS and the other classes coexist, appears between the latter two regions. The SOM exhibits a good generalization capability.

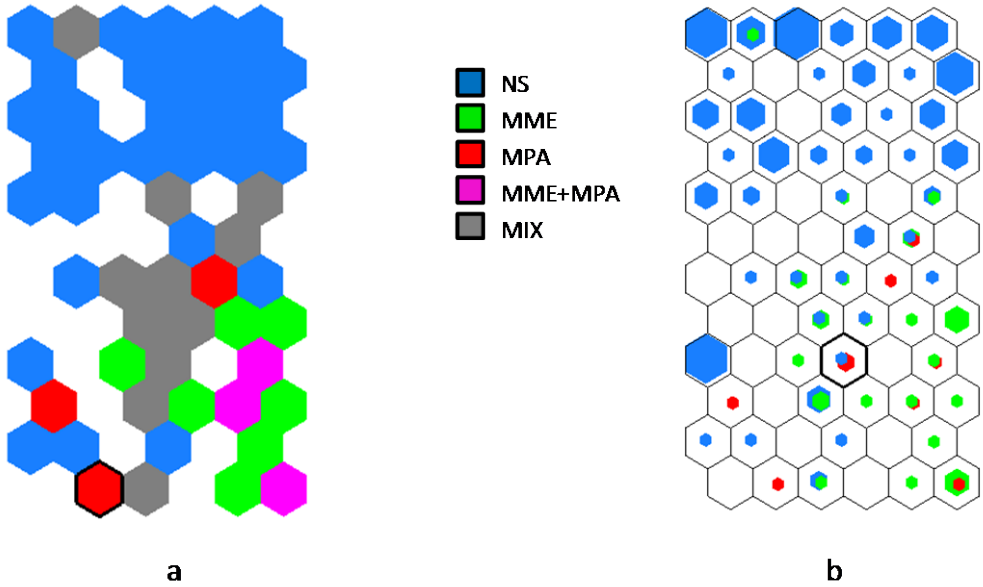


Fig. 8-5. 2-D SOM map colored on the basis of: a) the neurons type (blue NS, green MME, red MPA, magenta MME+MPA, grey mixed, white empty neurons); b) the classes type (blue NS, red MPA, green MME) and their density.

Analyzing the map composition, it results that, the 85.8% of NS classes belongs to NS neurons, 53.3% of MME classes belongs to MME or MME+MPA neurons, and 66.7% of MPA classes belongs to MPA or MPA+MME neurons. The composition of NS, MME and MPA classes in the different clusters of the map are represented in percentage in Fig. 8-6, Fig. 8-7 and Fig. 8-8, respectively.

As reported in Fig. 8-6, the 0.7% of NS classes is contained in neurons that include the MPA classes, the 12.8% of NS classes are included in neurons with MME classes, and the 0.7% of NS classes is enclosed in neurons where both MME and MPA classes are present. Considering the percentage of misclassified MME classes, the 40% of them is included in neurons with NS classes, and the 6.7% is contained in MIX neurons, as shown in Fig. 8-7.

Additionally, as reported in Fig. 8-8, one-third of MPA classes, i.e., 3 MPA classes, belong to mixed clusters, one is contained in MIX neuron and two in the neuron with NS classes, which is marked in black in Fig. 8-5b.

The framed mixed neuron contains one NS class of the Pat 1 and two MPA classes of the Pat 3; in Fig. 8-9 the signals recorded by C4-A1 EEG channel for the NS class (blue) and the two A3 phases manually labelled as MPA, colored in red and black respectively, are depicted. For the latest ones, the starting time of major attacks and paroxysmal arousals have been pointed out. As it can be noted, the NS class presents a slip signature close to those of the crisis. Then, the "errors" of the SOM map can be justified.

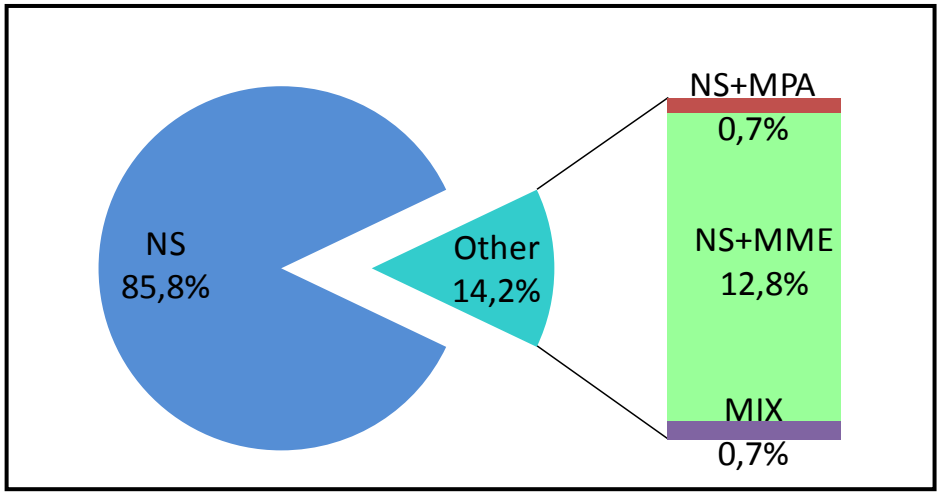


Fig. 8-6 Distribution of NS phases in SOM.

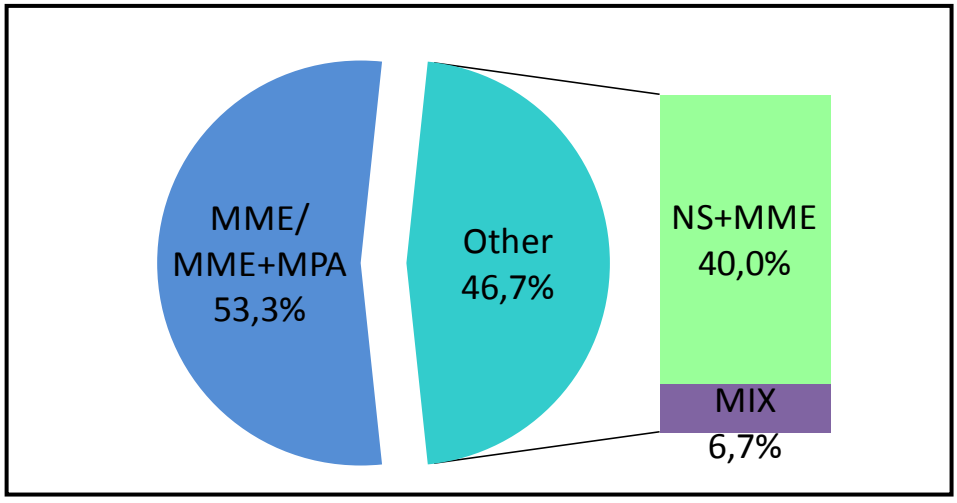


Fig. 8-7 Distribution of MME phases in SOM.

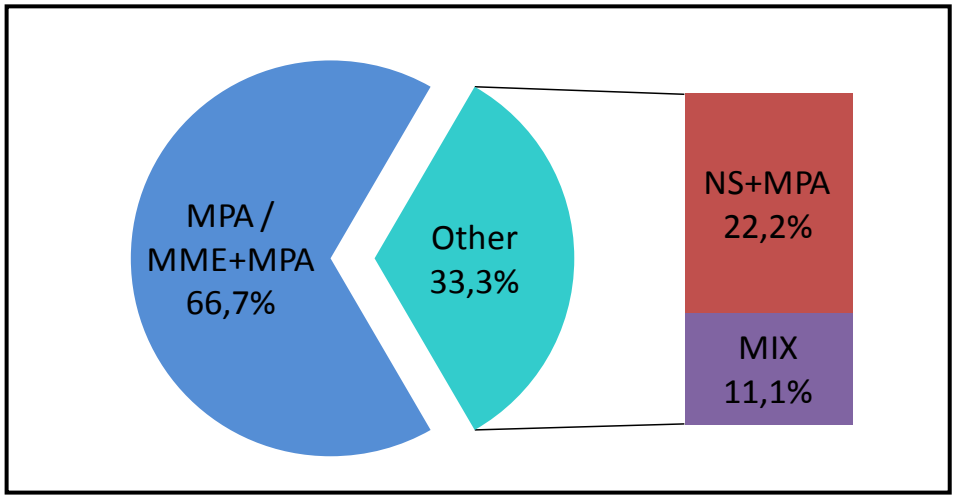


Fig. 8-8 Distribution of MPA phases in SOM.

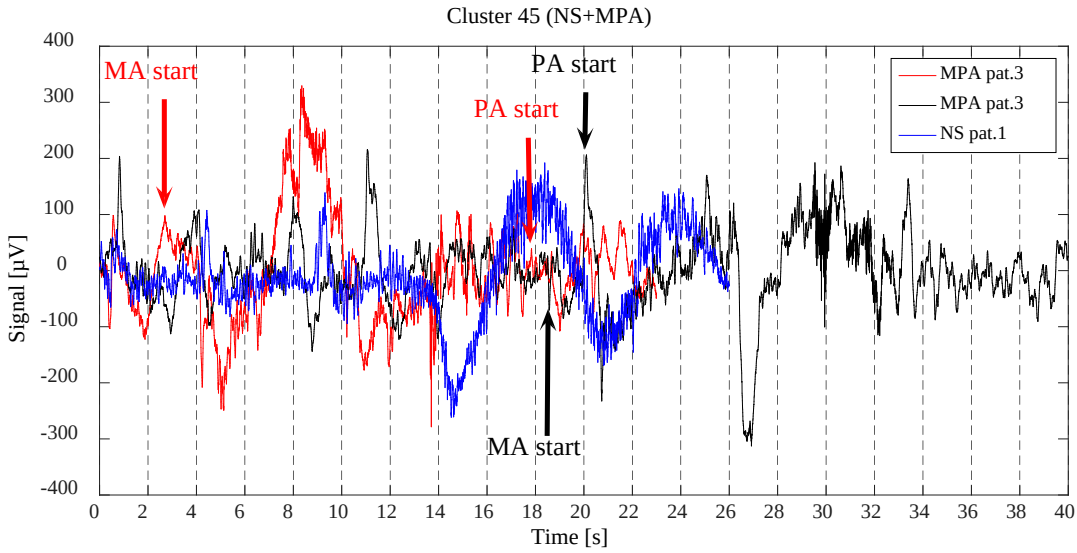


Fig. 8-9. Signals from the C4-A1 EEG channel of a NS class of Pat.1, and two phases of Pat 3 containing both major attacks and paroxysmal arousals. The starting time of MAs and PAs are pointed out.

Results and Performance

The SOM has been used to analyze the temporal evolution of the EEG signals in order to discriminate among different brain activities. To this purpose, 4s sliding windows with an overlapping of 3s have been considered. The length of the moving windows has been chosen by taking into account the minimum length of the events in the A3 signals. For each window, wavelet packet decomposition with the 4th order Daubechies wavelet and the standard deviation of the coefficients have been computed. In this way each sliding window in the signal has been associated to an 8-D point that can be projected onto the SOM, and the time evolution of the EEG signal can be displayed as a trajectory on the map.

The A3 phase, represented in Fig. 8-1, has been retained in order to test the SOM. At the top of the Fig. 8-10, the temporal evolution of the C4-A1 EEG signal is reported. The start times of the two critical events (a major attack and a minor motor event) are indicated. The bottom of the figure shows the SOM output represented by a bar graph (with the same color code of the SOM in Fig. 8-5a), corresponding to neurons on which the operating point is progressively projected. Every bar corresponds to a 4s sliding window. The first bar (labelled with 0) refers to the time window [0-4]s whereas the last bar (labelled with 53) refers to the time window [53-57]s. Note that, as the signal lasts 57s, the subsequent windows would contain a number of samples less than 4 and have not been considered.

Following the trajectory on the SOM, it is possible to clearly recognize the major attack, which begins to show its signatures starting from second 9. Also the MME is clearly identified by the SOM. As can be seen, the start time of that event is 23.316s and the signal trajectory on the map enters a MPA red neuron at the 24^{th} s, then it passes through MME green neurons. Finally, it remains in NS neurons (blue) almost until the end.

The observed behaviour suggests that an MPA starts at 9s and lasts for about 6s. Then, the SOM recognizes some crisis precursors and clearly identifies a MME, which starts at 24s and lasts for about other 6s. Finally, the EEG record no longer contains any crisis signatures.

In Fig. 8-11, the output of the SOM corresponding to a NS phase of one of the three patients, is shown. As can be seen, the trajectory on the SOM crosses only NS (blue) neurons.

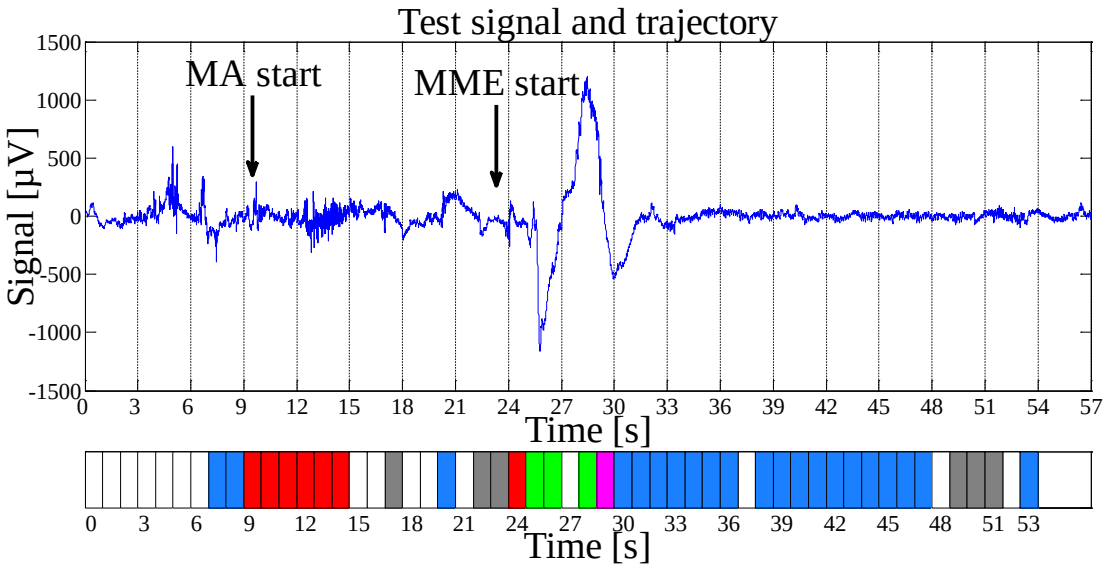


Fig. 8-10. SOM predictor output corresponding to the test signal: top of the figure - the test signal containing both major attack and minor motor event, for which the corresponding starting times are pointed out; bottom of the figure - bars (with the same colour code of the SOM in Fig. 8-5a) corresponding to neurons on which the operating point is progressively projected.

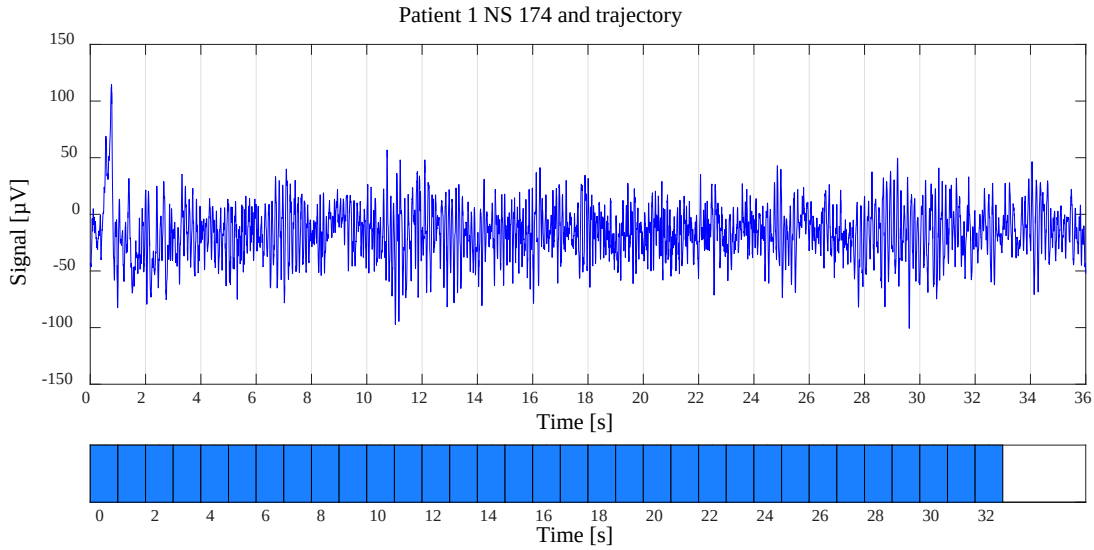


Fig. 8-11 SOM predictor output corresponding to an NS epoch of patient n.1: top of the figure - the test signal; bottom of the figure - bars (with the same colour code of the SOM in Fig. 8-5a) corresponding to neurons on which the operating point is progressively projected.

To further test the detection capability of the SOM, the specificity (SP), i.e., the percentage ratio between true negatives and the sum of true negatives and false positives, has been calculated. The evaluation of the metrics has been assessed applying the majority method: each A3 phase whose number of 4s sliding windows (i.e., bars) projected in NS neurons is greater than the half of the sum of all bars, is defined as true negative, otherwise, it is considered as a false positive.

The specificity, calculated on the 255 NS phases not contained in the training set, reached an average value of 90.20% where for the three patients was 93.44%, 89.55% e 84.85%, respectively. Furthermore, the SOM has been tested using also the 5 patients of the database, the test mapping achieved a performance of specificity equal to 63.40%, whereas, the evaluation of sensitivity has not assessed due to the limited number of seizures.

With the aim to test whether the brain activity of healthy subjects during A3 sleep phases is different from that of patients during NS phases, the SOM has been tested using the control group. In the majority of the tested epochs the trajectory evolves within regions containing NS or mixed neurons reaching a value of specificity of 84.19%. No MPA, MME or MPA+MME neurons were crossed.

Conclusions

A patient-independent system to automatically classify and label epileptic seizures from patients affected by ADNFLE is proposed.

The features extracted by the decomposition of the C4-A1 EEG channel by means of the wavelet packet transform have been used to train the SOM. The SOM has been subsequently tested using the remaining data of the same three patients affected by ADNFLE and other five patients affected by NFLE.

For the case of the three patients affected by ADNFLE, the obtained results show the suitability of the SOM mapping to detect different brain electrical activities, achieving an average specificity equal to 90.20%. Conversely, the performance decreases considerably when the model is tested on the other five patients, reaching a specificity of 63.40%. Furthermore, the SOM has been tested on the control group reaching an interesting value of specificity corresponding to 84.19%.

These results, despite the acceptable performances, suggest that the development of a patient-specific NFLE Seizure Detection System can ensure an improvement of the metrics.

8.3 NFLE Seizures Characterization System

In order to improve the potential of the SOM to discriminate MPA and MME classes from NS classes, a second approach has been developed using data from the database *Cagliari&Parma*, including different features.

The A3 phases from two EEG channels, the C4-A1 and the F4-A1 have been investigated. The choice of the F4-A1 EEG channel is given by the fact that it records the electrical activity of the frontal region of the brain. Only the A3 phases lasting more than 4s have been considered.

Since the information, given by neurologists, is only related to the starting time of the seizure, the A3 phases have been preprocessed. The samples before the starting time of the seizures have been removed, for both MME and MPA phases.

Feature Extraction and Data Visualization

The db4 has been used as wavelet function to decompose the EEG signals in the most relevant frequency sub-bands:

- Low delta: 0 – 2 Hz.
- High delta: 2 – 4 Hz.
- Theta: 4 – 8 Hz.
- Alpha: 8 – 16 Hz.
- Beta: 16 – 32 Hz.

In this second approach, the mean and the standard deviation of the wavelet coefficients have been calculated for each frequency sub-bands.

In order to know the discriminant capability of these features, for each frequency sub-band and for each EEG channel, the standard deviations of the NS, MME and MPA phases used in the training set have been represented in function of their means. In Fig. 8-12 and in Fig. 8-13, the scatter plots for F4-A1 and C4-A1 EEG channels are depicted, respectively. It can be noted that the NS phases, symbolized with blue circles, are separated from the MME and MPA phases, described with green stars and red crosses, respectively. More precisely, a good separation can be shown in the high delta frequency band of both EEG channels; in addition, NS phases are well distinguished from MME and MPA phases in the beta and alpha frequency bands of F4-A1 and C4-A1 EEG channels, respectively. This suggests a good aptitude of these features to discriminate NS from MME and MPA phases and their suitability to classify seizures. Thus, mean and standard deviation of the wavelet coefficients have been used in this study as features to train the model.

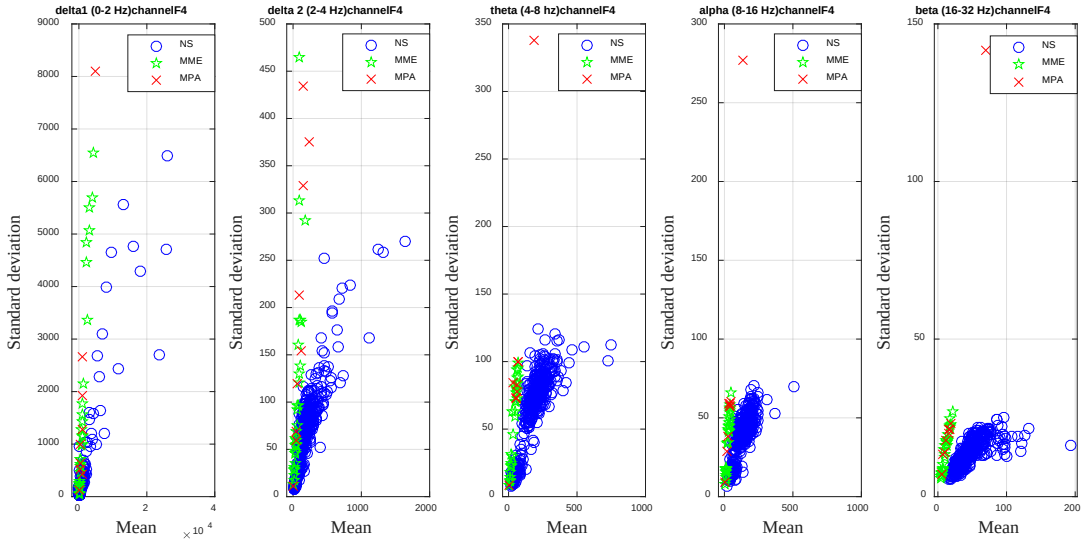


Fig. 8-12 Representation of the standard deviation in function of the mean for all frequency bands considered in F4 channel.

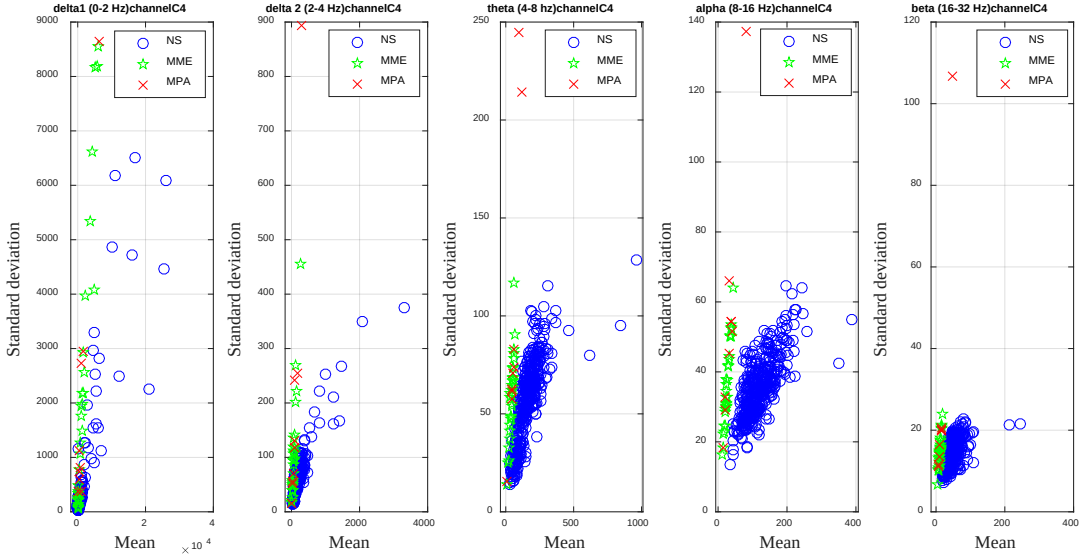


Fig. 8-13 Representation of the standard deviation in function of the mean for all frequency bands considered in C4 channel.

Building model

The SOM has been trained with the data of three patients affected by ADNFLE (442 A3 phases), whereas, the remaining 487 A3 phases of the five patients were used as tests. The heuristic formula of the map dimension, suggested by the SOM toolbox, has been considered. The resulting map, shown in Fig. 8-14a, has twenty inputs and 5×22 neurons. Each color in the map is representative of a particular composition of the neuron in terms of MME (green), MPA (red) and NS (blue) classes, applying the same color code used in the previous presented system (§8.2).

The map clearly highlights a large cluster of blue neurons and smaller clusters of red, green and magenta neurons, showing a good discriminant propensity.

The map is composed by 86.9% of NS classes that belong to NS neurons, by 60% of MME classes belonging to MME or MME+MPA neurons, and by 44.4% of MPA classes included in MPA or MPA+MME neurons, the density of the classes for each neuron is represented in Fig. 8-14b.

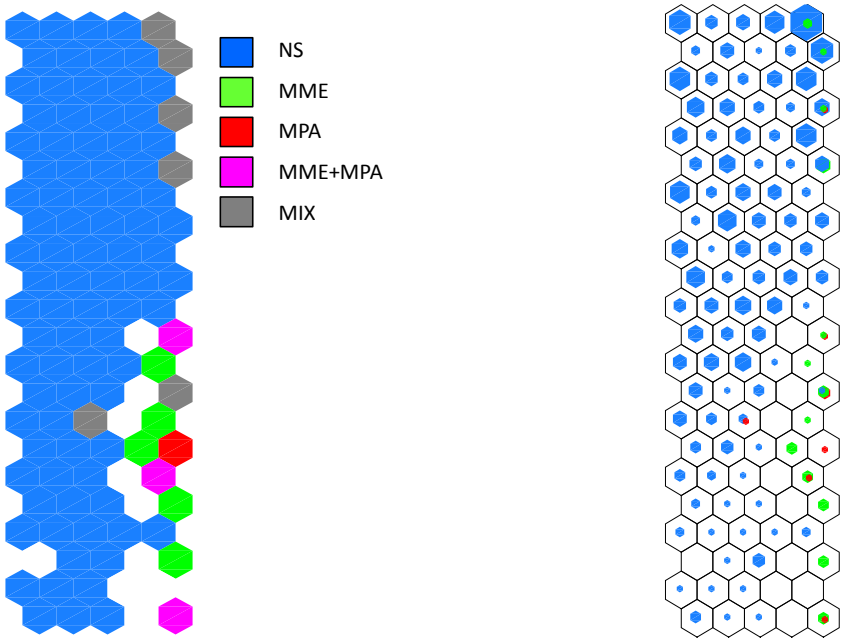


Fig. 8-14 SOM map colored on the basis of: a) the neurons type (blue NS, green MME, red MPA, magenta MME+MPA, grey mixed, white empty neurons); b) the classes type (blue NS, red MPA, green MME) and their density.

Results and Performance

To evaluate the output of the SOM model, the A3 phases of the control group and of the five patients (not included in the SOM training) have been tested. The features extracted from a 4s sliding window (with an overlapping of 3s) have been projected on the map in order to analyze the temporal evolution of the EEG signals.

An example of the SOM output for A3 phase that include a MPA seizure is reported in Fig. 8-15. At the top of the figure the temporal evolution of the F4-A1 EEG channel is shown. The vertical red line identifies the starting time of the seizure, manually identified by neurologists. At the bottom of the figure the trajectory on the map is described by the bar graph. The first 4s window is projected on a MME neuron and then, at the beginning of the seizure, the operating points move to the MME+MPA neurons, highlighting a change of the signal towards a major attack.

The outputs of the system for two examples of NS and H phases are shown at the bottom of the Fig. 8-16 and Fig. 8-17, respectively. It can be noted that, both the examples are correctly identified in NS neurons.

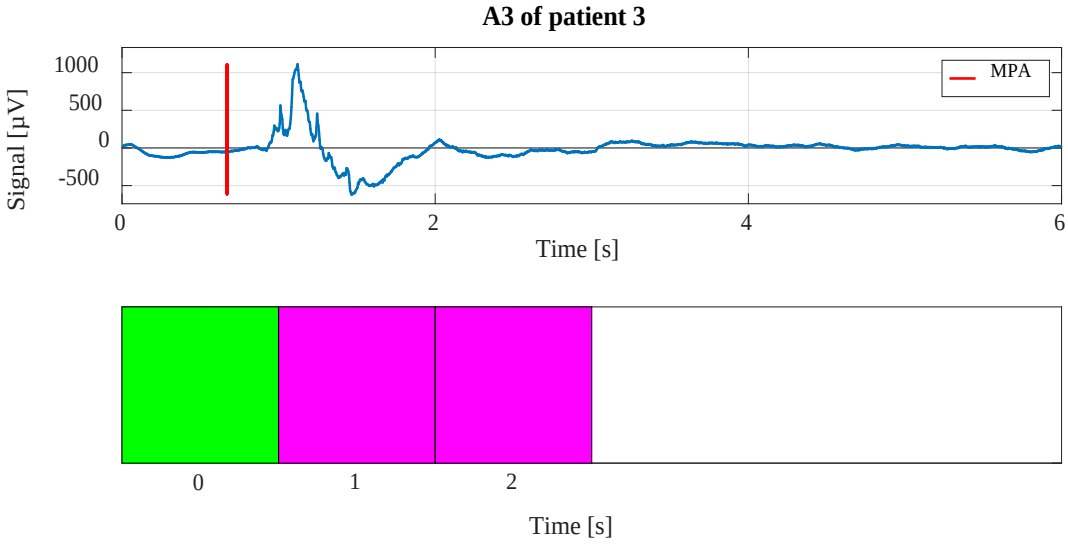


Fig. 8-15 Representation of SOM output: top of the figure – Signal of F4–A1 EEG channel of MPA phase; bottom - bars (with the same colour code of the SOM), corresponding to neurons on which the operating point is progressively projected.

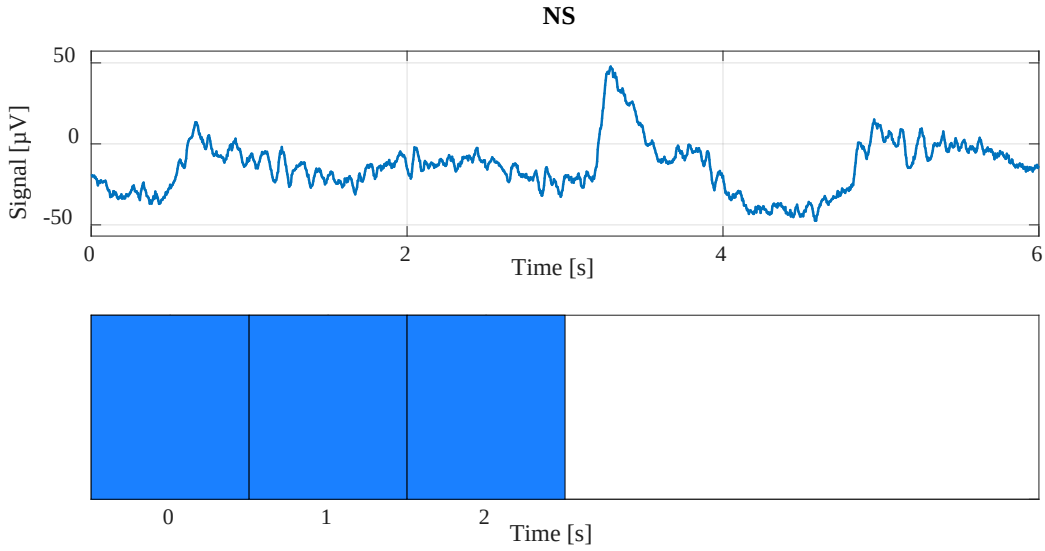


Fig. 8-16 Representation of SOM output: top of the figure – Signal of F4–A1 EEG channel of a NS phase; bottom of the figure- bars (with the same colour code of the SOM), corresponding to neurons on which the operating point is progressively projected.

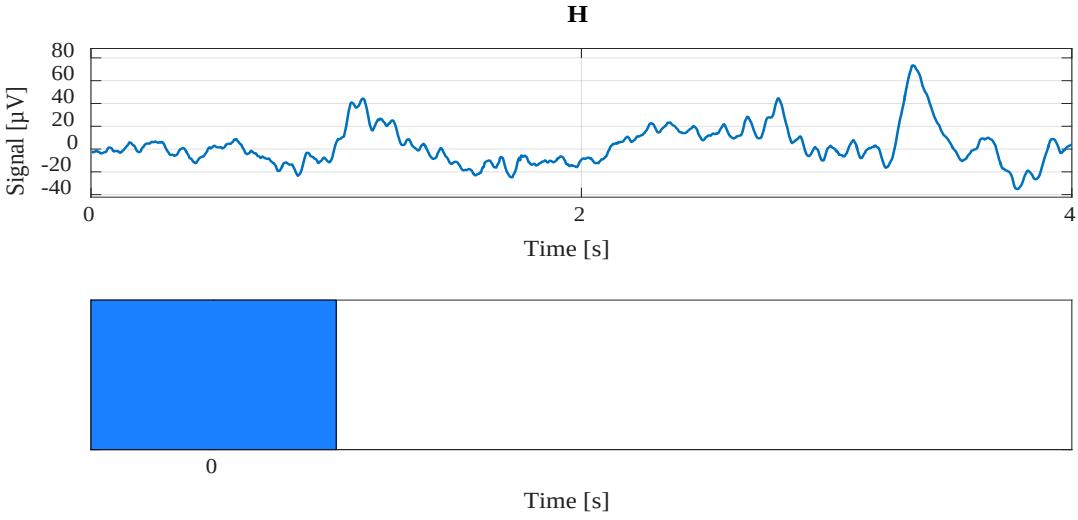


Fig. 8-17 Representation of SOM output: top of the figure – Signal of F4–A1 EEG channel of an H phase from the control group; bottom of the figure - bars (with the same colour code of the SOM), corresponding to neurons on which the operating point is progressively projected.

Because of the poor number of the MME and MPA phases, only the specificity has been evaluated. The evaluation of the metrics has been assessed evaluating the majority classes for each bar and, consequently, for the entire phases: each A3 phase whose number of NS bars, defined as the 4s-sliding windows projected in neurons with the majority of NS classes, is greater than the half of the sum of all bars, is defined as true negative; otherwise, it is considered as a false positive.

The specificity, evaluated for the 5 patients affected by NFLE not used in the training phase, reached an average value of 70%, whereas the test using the control group achieved a value of 81.5%.

Conclusions

In this study a patient-independent system to detect Nocturnal Frontal Lobe Epilepsy has been presented. In order to improve the performance of a patient-independent system that uses only one channel, two EEG channels have been used to train a SOM.

The A3 phases of three patients affected by ADNFLE, previously cleaned from the samples before the beginning of the seizures, were used to train the SOM. The proposed method, tested on five patients and the control group, reached the values of specificity equal to 70% and 81.5%, respectively. The model is capable to recognize the changing of the brain state during the seizures, identifying the correct cluster.

Comparing the specificity evaluated on the 5 NFLE patients with the previous approach, this proposed system achieves a slightly higher value; however, the specificity calculated testing the control group decreases slightly.

Both the proposed patient-independent models showed good performances in the identification of phases without seizures. However, the patient-specific models demonstrated higher performances due to the customization of the parameters, thus confirming the choice of using a personalized approach for each patient.

9 CONCLUSION

This thesis presents a machine learning multi-feature approach for detection of nocturnal frontal lobe epileptic seizures using EEG signals.

Despite the intense study carried out during the last decades in order to find a best epileptic seizure detection system, few contributions are reported to detect nocturnal epilepsy.

Among the available methods, a system based on SOM mapping has been developed. The capability of the SOM to convert complex nonlinear relationships between high dimensional data into a low-dimensional space preserving the topological properties of the original space has been chosen to identify the nocturnal frontal lobe seizures. Additionally, the SOM allow us to visualize and follow the projection of the temporal sequence of the EEG recordings on the map and analyze the dynamics of the brain state.

Firstly, a patient-subjected system has been developed to detect nocturnal seizures in patients affected by NFLE. For each of the nine patients of the European Epilepsy Database, different simulation scenarios have been performed in order to find the best combination of features, number of EEG channels and training parameters.

As the focus of this type of epilepsy is in the frontal regions, the study has been carried out using seven EEG channels from the anterior cerebral hemisphere. Furthermore, the increase of four recordings from the posterior hemisphere has been considered, taking into account the possibility of seizure propagation to the occipital and parietal regions.

Data have been labelled, according to the information provided by neurologists, as no-seizure (NS) and seizure (SZ) epochs and the SOM has been trained with the first half of SZ epochs and the associated previously 10 minutes of NS epochs.

The remaining data have been subsequently mapped onto the 2-D SOM and the performances of sensitivity and specificity have been evaluated.

The system reached interesting mean values of 77.23% (between 54.55% and 100%) and 88.94% (between 83.13% and 97.83%) of sensitivity and specificity, respectively, with networks predominantly trained using eleven EEG channels. Additionally, by means of the SOM visualization capabilities, it was possible to identify SZ and NS regions and follow the temporal evolution of the projection of the test operating points onto the map.

In order to improve the performance and validate the capability of the SOM mapping in nocturnal epileptic seizures, multiple scenarios have been trained using data of patients affected by frontal and/or temporal nocturnal epileptic seizures of the European Epilepsy Database.

For each of the 22 patients, nine EEG channels have been considered according to the information provided by neurologists about the location of the focal seizures. The reduced

number of EEG channels has been chosen with a view to minimizing patient discomfort in the amount of electrodes placed on the head. Several features have been extracted and sixteen combinations of parameters and features have been examined in order to achieve the best SOM for each patient.

The results highlighted an increase of the performance, the system reached an average sensitivity of 82.85% (between 63.64% and 100%) and specificity of 89.92% (between 71.67% and 99.17%).

The proposed method has been compared training an SVM classifier, widely used in epileptic seizure detection systems. For each patient, the best SVM network has been chosen and the performances have been analyzed. The SVM networks achieved an average sensitivity equal to 82.11% (between 0% and 100%) and a mean specificity of 82.85% (between 43.33% and 99.17%).

It can be argued that the lowest value of sensitivity using SVM model can be attributed to the sleep stage where the nocturnal seizures occur. In fact, in this case, the majority of the epileptic seizures arise during the REM stage which contains high frequency that may be misclassified with normal REM sleep by the SVM model. However, SOM model manages to detect both SZ and NS epochs.

It can be noted that, the SOM achieved comparable and, in some patients, higher performance than the SVM, indicating the SOM mapping method as a good alternative to detect nocturnal epileptic seizures.

A sleep stage detector can be incorporate to the nocturnal epileptic seizure detection system to improve the performance.

In order to generalize the approach, a patient-independent system has been developed using the scalp EEG signals from the Sleep Centers of Cagliari and Parma. The information on sleep microstructure, provided in the database, has been used to extract only the significant phases in nocturnal frontal lobe epilepsy according to the studies in literature [47] [48]. Thus, the features extracted from the A3 phases of the Cyclic Alternating Pattern have been used to train the SOM using selected data from the C4-A1 EEG channel of three patients affected by ADNFLE. The remaining data have been used to test the model, as well as the five patients not included in the training phase.

The performance, evaluated in term of specificity, achieved an average value of 90.20% for the remaining data of the three ADNFLE patients, whereas, the test using the five patients revealed a specificity of 63.40%. Due to lack information about the final signature of the seizures and the poor number of seizures, the sensitivity cannot be assessed. However, following the trajectory into the SOM map, it has been observed that the model recognizes the variation of the signal at the imminence of the epileptic seizures.

Furthermore, the SOM has been tested on healthy subjects in order to establish if the A3 phases were different from those of patients. The mean specificity reached a value of 84.19%; proving a good resemblance between the A3 phases without seizures of the patients and those of the healthy subjects.

It can be noted that, the model shows good performance for the same patients used in the training phase, whereas a low performance is reached when other NFLE patients are tested. With the aim to improve this performance, a system, including a frontal EEG channel, has been developed. The A3 phases of three patients affected by ADNFLE have been previously cleaned from the samples before the starting of the seizures, and then the mean and the standard deviation of the wavelet coefficients have been used as features to train the SOM.

The testing of the model on five NFLE patients, reached a mean specificity of 70%, slightly greater than the previous model. Additionally, testing the control group, the specificity decreases slightly, reaching values of 81.5%.

Different methodologies have been investigated both for patient-dependent and patient-independent systems, the performances of the proposed methods, highlight the potential of the patient-specific system to detect nocturnal epileptic seizures, with a customized combination of number of EEG channels and features per patient.

Furthermore, the capability of SOM mapping to distinguish seizure from non-seizure events has proved to be a suitable alternative to other detection models.

According to the existing literature, the use of accelerometers reaches good performance in nocturnal epileptic seizure detection, but, it is subjected to a high number of artefacts and false positive outputs.

The use of scalp EEG signals and wavelet transform, as feature extraction technique, proposed in this thesis, have shown interesting results of sensitivity and specificity values, despite the difficulty in nocturnal epileptic seizure detection; where, in almost 50% of cases, scalp EEGs do not show abnormalities.

However, from the results presented in this thesis, further investigations on prediction time are needed to evaluate the possibility of a clinical use, and to allow the development of an epileptic seizure detection system that can be integrated with implantable stimulation devices.

10 BIBLIOGRAPHY

[1] S. Ramgopal *et al.*, "Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy," *Epilepsy & Behavior*, vol. 37, pp. 291–307, Aug. 2014.

[2] R. S. Fisher *et al.*, "Epileptic Seizures and Epilepsy: Definitions Proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)," *Epilepsia*, vol. 46, no. 4, pp. 470–472, Apr. 2005.

[3] R. S. Fisher *et al.*, "ILAE Official Report: A practical clinical definition of epilepsy," *Epilepsia*, vol. 55, no. 4, pp. 475–482, Apr. 2014.

[4] C. N. X. OpenStax, *Biology*. 2016.

[5] F. A. C. Azevedo *et al.*, "Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain," *J. Comp. Neurol.*, vol. 513, no. 5, pp. 532–541, Apr. 2009.

[6] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, "Neurotransmitters, Synapses, and Impulse Transmission," 2000.

[7] D. Purves *et al.*, "Excitatory and Inhibitory Postsynaptic Potentials," 2001.

[8] G. L. Holmes and Y. Ben-Ari, "The Neurobiology and Consequences of Epilepsy in the Developing Brain," *Pediatr. Res.*, vol. 49, no. 3, pp. 320–325, Mar. 2001.

[9] R. S. Fisher *et al.*, "Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology," *Epilepsia*, vol. 58, no. 4, pp. 522–530, Apr. 2017.

[10] P. Ryvlin, S. Rheims, and G. Risse, "Nocturnal Frontal Lobe Epilepsy," *Epilepsia*, vol. 47, pp. 83–86, Nov. 2006.

[11] F. Provini, G. Plazzi, P. Tinuper, S. Vandi, E. Lugaresi, and P. Montagna, "Nocturnal frontal lobe epilepsy. A clinical and polygraphic overview of 100 consecutive cases," *Brain J. Neurol.*, vol. 122 (Pt 6), pp. 1017–1031, Jun. 1999.

[12] I. E. Scheffer, "Autosomal Dominant Nocturnal Frontal Lobe Epilepsy," *Epilepsia*, vol. 41, no. 8, pp. 1059–1060, Aug. 2000.

[13] O. K. Steinlein *et al.*, "A missense mutation in the neuronal nicotinic acetylcholine receptor alpha 4 subunit is associated with autosomal dominant nocturnal frontal lobe epilepsy," *Nat. Genet.*, vol. 11, no. 2, pp. 201–203, Oct. 1995.

[14] M. Puligheddu *et al.*, "Rationale for an adjunctive therapy with fenofibrate in pharmacoresistant nocturnal frontal lobe epilepsy," *Epilepsia*, vol. 58, no. 10, pp. 1762–1770, Oct. 2017.

[15] R. B. Berry, *Fundamentals of Sleep Medicine,* Elsevier Health Sciences, 2011.

[16] M. T. Salam, M. Sawan, and D. K. Nguyen, "Low-Power Implantable Device for Onset Detection and Subsequent Treatment of Epileptic Seizures: A Review," *Journal of Healthcare Engineering*, 2010.

[17] V. De Herdt *et al.*, "Vagus nerve stimulation for refractory epilepsy: A Belgian multicenter study," *Eur. J. Paediatr. Neurol.*, vol. 11, no. 5, pp. 261–269, Sep. 2007.

[18] S. C. Schachter, J. Guttag, S. J. Schiff, D. L. Schomer, and Summit Contributors, “Advances in the application of technology to epilepsy: the CIMIT/NIO Epilepsy Innovation Summit,” *Epilepsy Behav. EB*, vol. 16, no. 1, pp. 3–46, Sep. 2009.
[19] I. Osorio *et al.*, “Automated seizure abatement in humans using electrical stimulation,” *Ann. Neurol.*, vol. 57, no. 2, pp. 258–268, Feb. 2005.
[20] F. T. Sun and M. J. Morrell, “Closed-loop Neurostimulation: The Clinical Experience,” *Neurotherapeutics*, vol. 11, no. 3, pp. 553–563, Jul. 2014.
[21] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, 1995.
[22] J. O. Ollikainen, M. Vauhkonen, P. A. Karjalainen, and J. P. Kaipio, “Effects of local skull inhomogeneities on EEG source estimation,” *Med. Eng. Phys.*, vol. 21, no. 3, pp. 143–154, Apr. 1999.
[23] M. Fuchs, M. Wagner, and J. Kastner, “Development of volume conductor and source models to localize epileptic foci,” *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.*, vol. 24, no. 2, pp. 101–119, Apr. 2007.
[24] M. X. Huang, J. C. Mosher, and R. M. Leahy, “A sensor-weighted overlapping-sphere head model and exhaustive head model comparison for MEG,” *Phys. Med. Biol.*, vol. 44, no. 2, p. 423, 1999.
[25] M. Fuchs, R. Drenckhahn, H. A. Wischmann, and M. Wagner, “An improved boundary element method for realistic volume-conductor modeling,” *IEEE Trans. Biomed. Eng.*, vol. 45, no. 8, pp. 980–997, Aug. 1998.
[26] S. Rush and D. A. Driscoll, “EEG Electrode Sensitivity-An Application of Reciprocity,” *IEEE Trans. Biomed. Eng.*, vol. BME-16, no. 1, pp. 15–22, Jan. 1969.
[27] P. L. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, 2006.
[28] M. Akhtari *et al.*, “Conductivities of three-layer human skull,” *Brain Topogr.*, vol. 13, no. 1, pp. 29–42, 2000.
[29] S. Tong and N. V. Thakor, *Quantitative EEG Analysis Methods and Clinical Applications*. Artech House, 2009.
[30] R. Flink *et al.*, “Guidelines for the use of EEG methodology in the diagnosis of epilepsy,” *Acta Neurol. Scand.*, vol. 106, no. 1, pp. 1–7, Jul. 2002.
[31] G. H. Klem, H. O. Lüders, H. H. Jasper, and C. Elger, “The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology,” *Electroencephalogr. Clin. Neurophysiol. Suppl.*, vol. 52, pp. 3–6, 1999.
[32] R. Flink *et al.*, “Guidelines for the use of EEG methodology in the diagnosis of epilepsy. International League Against Epilepsy: commission report. Commission on European Affairs: Subcommission on European Guidelines,” *Acta Neurol. Scand.*, vol. 106, no. 1, pp. 1–7, Jul. 2002.
[33] J. D. Bronzino, *Medical Devices and Systems*. CRC Press, 2006.
[34] H. Türker and H. Sözen, “Surface Electromyography in Sports and Exercise,” 2013.
[35] W. Heide, E. Koenig, P. Trillenberg, D. Kömpf, and D. S. Zee, “Electrooculography: technical standards and applications. The International Federation of Clinical

Neurophysiology," *Electroencephalogr. Clin. Neurophysiol. Suppl.*, vol. 52, pp. 223–240, 1999.
[36] S. P. Singh, "Magnetoencephalography: Basic principles," *Ann. Indian Acad. Neurol.*, vol. 17, no. Suppl 1, pp. S107–S112, Mar. 2014.
[37] J. S. González-Salazar *et al.*, "Characterization of the autonomic system during the cyclic alternating pattern of sleep," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 3805–3808.
[38] F. Leutmezer, C. Schernthaner, S. Lurger, K. Pötzelberger, and C. Baumgartner, "Electrocardiographic changes at the onset of epileptic seizures," *Epilepsia*, vol. 44, no. 3, pp. 348–354, Mar. 2003.
[39] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.," *Wash. DC Public Health Serv. US Gov. Print. Off. 1968*.
[40] T. Hori *et al.*, "Proposed supplements and amendments to 'A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects', the Rechtschaffen & Kales (1968) standard," *Psychiatry Clin. Neurosci.*, vol. 55, no. 3, pp. 305–310, Jun. 2001.
[33] http://eegatlas-online.com/index.php/en/.
[42] RazerM, *English: Hypnogram of sleep between midnight and 6.30 am*. 2011.
[43] M. G. Terzano, D. Mancia, M. R. Salati, G. Costani, A. Decembrino, and L. Parrino, "The cyclic alternating pattern as a physiologic component of normal NREM sleep," *Sleep*, vol. 8, no. 2, pp. 137–145, 1985.
[44] M. G. Terzano *et al.*, "Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep," *Sleep Med.*, vol. 2, no. 6, pp. 537–553, Nov. 2001.
[45] B. Farina *et al.*, "Microstructure of sleep in depressed patients according to the cyclic alternating pattern," *J. Affect. Disord.*, vol. 77, no. 3, pp. 227–235, Dec. 2003.
[46] L. Parrino, P. Halasz, C. A. Tassinari, and M. G. Terzano, "CAP, epilepsy and motor events during sleep: the unifying role of arousal," *Sleep Med. Rev.*, vol. 10, no. 4, pp. 267–285, Aug. 2006.
[47] L. Parrino *et al.*, "Distinctive polysomnographic traits in nocturnal frontal lobe epilepsy," *Epilepsia*, vol. 53, no. 7, pp. 1178–1184, Jul. 2012.
[48] F. De Paolis *et al.*, "Effects of antiepileptic treatment on sleep and seizures in nocturnal frontal lobe epilepsy," *Sleep Med.*, vol. 14, no. 7, pp. 597–604, Jul. 2013.
[49] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM J Res Dev*, vol. 3, no. 3, pp. 210–229, Jul. 1959.
[50] T. M. Mitchell, *Machine Learning*. McGraw-Hill, 1997.
[51] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn. 20 273-297 1995*.
[52] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst. Man Cybern.*, vol. 21, no. 3, pp. 660–674, May 1991.
[53] M. H. Hassoun, *Fundamentals of Artificial Neural Networks*. MIT Press, 1995.
[54] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, Feb. 2009.

[55] P. Melin and O. Castillo, *Hybrid Intelligent Systems for Pattern Recognition Using Soft Computing: An Evolutionary Approach for Neural Networks and Fuzzy Systems*. Springer Science & Business Media, 2005.
[56] S. Pavalakodi, *A New Method for Dimensionality Reduction using K- Means Clustering Algorithm for High Dimensional Data Set*. .
[57] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.
[58] I. Jolliffe, *Principal component analysis*. Springer-Verlag New York, 2002.
[59] H. Wang, X. Lu, Z. Hu, and W. Zheng, "Fisher Discriminant Analysis With L1-Norm," *IEEE Trans. Cybern.*, vol. 44, no. 6, pp. 828–842, Jun. 2014.
[60] G. McLachlan and D. Peel, *Finite Mixture Models*. John Wiley & Sons, 2004.
[61] U. R. Acharya, S. Vinitha Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: A review," *Knowl.-Based Syst.*, vol. 45, pp. 147–165, Jun. 2013.
[62] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, and F. E. A. El-Samie, "EEG seizure detection and prediction algorithms: a survey," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, p. 183, 2014.
[63] A. V. Oppenheim and R. W. Schafer, *Discrete-time Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1989.
[64] S. Smith, *Digital Signal Processing: A Practical Guide for Engineers and Scientists*. Newnes, 2013.
[65] D. Chen, S. Wan, J. Xiang, and F. S. Bao, "A high-performance seizure detection algorithm based on Discrete Wavelet Transform (DWT) and EEG," *PLOS ONE*, vol. 12, no. 3, p. e0173138, Mar. 2017.
[66] S. B. Wilson, "Algorithm architectures for patient dependent seizure detection," *Clin. Neurophysiol.*, vol. 117, no. 6, pp. 1204–1216, Jun. 2006.
[67] A. S. Zandi, M. Javidan, G. A. Dumont, and R. Tafreshi, "Automated Real-Time Epileptic Seizure Detection in Scalp EEG Recordings Using an Algorithm Based on Wavelet Packet Transform," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1639–1651, Jul. 2010.
[68] L. Wang *et al.*, "Automatic Epileptic Seizure Detection in EEG Signals Using Multi-Domain Feature Extraction and Nonlinear Analysis," *Entropy*, vol. 19, no. 6, p. 222, May 2017.
[69] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform," *J. Neurosci. Methods*, vol. 123, no. 1, pp. 69–87, Feb. 2003.
[70] C. K. Chui, *An Introduction to Wavelets*. San Diego, CA, USA: Academic Press Professional, Inc., 1992.
[71] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.
[72] S. M. Pincus, "Approximate entropy as a measure of system complexity.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 88, no. 6, pp. 2297–2301, Mar. 1991.
[73] N. A. Campbell and W. R. Atchley, "The Geometry of Canonical Variate Analysis," *Syst. Zool.*, vol. 30, no. 3, pp. 268–280, 1981.
[74] Kaiser H.F., "The application of electronic computers to factor analysis," no. Educ. Psychol. Meas., 20, 141–151., 1960.

[75] I. T. Jolliffe, “Discarding Variables in a Principal Component Analysis. I: Artificial Data,” *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 21, no. 2, pp. 160–173, 1972.
[69] Vallsvg, *Scree plot for the ALR transformed dataset*. 2008.
[77] J. L. Semmlow, *Biosignal and Medical Image Processing*. CRC Press, 2004.
[78] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2004.
[79] R. A. Fisher, “The Use of Multiple Measurements in Taxonomic Problems,” *Ann. Eugen.*, vol. 7, no. 2, pp. 179–188, Sep. 1936.
[80] C. Ding and H. Peng, “Minimum Redundancy Feature Selection from Microarray Gene Expression Data,” *J. Bioinform. Comput. Biol.*, vol. 3, no. 2, pp. 185–205, 2005.
[81] A. B. Carlson, P. B. Crilly, and J. C. Rutledge, *Communication systems : an introduction to signals and noise in electrical communication*, 4th ed. Boston : McGraw-Hill, 2002.
[82] M. Caudill, “Neural Networks Primer, Part I,” *AI Expert*, vol. 2, no. 12, pp. 46–52, Dec. 1987.
[83] S. Agatonovic-Kustrin and R. Beresford, “Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research,” *J. Pharm. Biomed. Anal.*, vol. 22, no. 5, pp. 717–727, Jun. 2000.
[84] F. Rosenblatt, *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Spartan Books, 1962.
[85] J. L. Elman, “Finding Structure in Time,” *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, Mar. 1990.
[86] B. M. Wilamowski and H. Yu, “Improved Computation for Levenberg-Marquardt Training,” *Trans Neur Netw*, vol. 21, no. 6, pp. 930–937, Jun. 2010.
[87] D. F. Specht, “Probabilistic Neural Networks,” *Neural Netw*, vol. 3, no. 1, pp. 109–118, Jan. 1990.
[88] S. Dua and X. Du, *Data Mining and Machine Learning in Cybersecurity*. CRC Press, 2016.
[89] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. Wiley, 2001.
[90] C. J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition,” *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, Jun. 1998.
[91] J. Vesanto, J. Himberg, E. Alhoniemi, J. Parhankangas, “SOM Toolbox for Matlab 5,” 2000.
[92] K. Polat and S. Güneş, “Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform,” *Appl. Math. Comput.*, vol. 187, no. 2, pp. 1017–1026, Apr. 2007.
[93] A. Watkins, J. Timmis, and L. Boggess, “Artificial Immune Recognition System (AIRS): An Immune-Inspired Supervised Learning Algorithm,” *Genet. Program. Evolvable Mach.*, vol. 5, no. 3, pp. 291–317, Sep. 2004.
[94] K. Polat and S. Güneş, “Artificial immune recognition system with fuzzy resource allocation mechanism classifier, principal component analysis and FFT method based new hybrid automated identification system for classification of EEG signals,” *Expert Syst. Appl.*, vol. 34, no. 3, pp. 2039–2048, Apr. 2008.

[95] H. Ocak, “Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy,” *Expert Syst. Appl.*, vol. 36, no. 2, Part 1, pp. 2027–2036, Mar. 2009.
[96] A. Subasi, “EEG signal classification using wavelet feature extraction and a mixture of expert model,” *Expert Syst. Appl.*, vol. 32, no. 4, pp. 1084–1093, May 2007.
[97] C. A. M. Lima, A. L. V. Coelho, and M. Eisencraft, “Tackling EEG signal classification with least squares support vector machines: A sensitivity analysis study,” *Comput. Biol. Med.*, vol. 40, no. 8, pp. 705–714, Aug. 2010.
[98] O. Faust, U. R. Acharya, H. Adeli, and A. Adeli, “Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis,” *Seizure*, vol. 26, pp. 56–64, Mar. 2015.
[99] A. Berdakh and S. H. Don, “Epileptic Seizures Detection Using Continuous Time Wavelet Based Artificial Neural Networks,” in *2009 Sixth International Conference on Information Technology: New Generations*, 2009, pp. 1456–1461.
[100] A. J. Gabor, “Seizure detection using a self-organizing neural network: validation and comparison with other detection strategies,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 107, no. 1, pp. 27–32, Jul. 1998.
[101] C. J. James, R. D. Jones, P. J. Bones, and G. J. Carroll, “Detection of epileptiform discharges in the EEG by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages,” *Clin. Neurophysiol.*, vol. 110, no. 12, pp. 2049–2063, Dec. 1999.
[102] U. Orhan, M. Hekim, and M. Ozer, “EEG signals classification using the K-means clustering and a multilayer perceptron neural network model,” *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13475–13481, Sep. 2011.
[103] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, “Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state,” *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, vol. 64, no. 6 Pt 1, p. 061907, Dec. 2001.
[104] Y. Kocyigit, A. Alkan, and H. Erol, “Classification of EEG Recordings by Using Fast Independent Component Analysis and Artificial Neural Network,” *J. Med. Syst.*, vol. 32, no. 1, pp. 17–20, Feb. 2008.
[105] K. Mahajan, M. R. Vargantwar, and S. M. Rajput, “Article: Classification of EEG using PCA, ICA and Neural Network,” *IJCA Proc. Int. Conf. Comput. Intell. ICCIA 2012*, vol. ICCIA, no. 6, pp. 1–4, Mar. 2012.
[106] A. Subasi and M. Ismail Gursoy, “EEG signal classification using PCA, ICA, LDA and support vector machines,” *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8659–8666, Dec. 2010.
[107] U. R. Acharya, S. V. Sree, and J. S. Suri, “Automatic detection of epileptic EEG signals using higher order cumulant features,” *Int. J. Neural Syst.*, vol. 21, no. 5, pp. 403–414, Oct. 2011.
[108] K. C. Chua, V. Chandran, U. R. Acharya, and C. M. Lim, “Application of Higher Order Spectra to Identify Epileptic EEG,” *J. Med. Syst.*, vol. 35, no. 6, pp. 1563–1571, Dec. 2011.

[109] V. Srinivasan, C. Eswaran, and N. Sriraam, “Approximate Entropy-Based Epileptic EEG Detection Using Artificial Neural Networks,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 3, pp. 288–295, May 2007.

[110] U. R. Acharya, F. Molinari, S. V. Sree, S. Chattopadhyay, K.-H. Ng, and J. S. Suri, “Automated diagnosis of epileptic EEG using entropies,” *Biomed. Signal Process. Control*, vol. 7, no. 4, pp. 401–408, Jul. 2012.

[111] T. J. Ross, *Fuzzy Logic with Engineering Applications*. John Wiley & Sons, 2009.

[112] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Elsevier, 2011.

[113] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan, “EEG-based neonatal seizure detection with Support Vector Machines,” *Clin. Neurophysiol.*, vol. 122, no. 3, pp. 464–473, Mar. 2011.

[114] K. AlSharabi, S. Ibrahim, R. Djemal, and A. Alsuwailem, “A DWT-entropy-ANN based architecture for epilepsy diagnosis using EEG signals,” in *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 2016, pp. 288–291.

[115] P. Fergus, D. Hignett, A. Hussain, D. Al-Jumeily, and K. Abdel-Aziz, “Automatic Epileptic Seizure Detection Using Scalp EEG and Advanced Artificial Intelligence Techniques,” *BioMed Res. Int.*, vol. 2015, 2015.

[116] R. R. Borhade and M. S. Nagmode, “EEG Signal Classification for Detection of Epileptic Seizures by extracting various Features –A Review,” 2016.

[117] A. Temko, N. Stevenson, W. Marnane, G. Boylan, and G. Lightbody, “Inclusion of temporal priors for automated neonatal EEG classification,” *J. Neural Eng.*, vol. 9, no. 4, p. 046002, 2012.

[118] N. Moghim and D. W. Corne, “Predicting Epileptic Seizures in Advance,” *PLoS ONE*, vol. 9, no. 6, Jun. 2014.

[119] C. Alexandre Teixeira *et al.*, “Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients,” *Comput. Methods Programs Biomed.*, vol. 114, no. 3, pp. 324–336, May 2014.

[120] B. Direito, C. Teixeira, M. Bandarabadi, F. Sales, and A. Dourado, “Automatic warning of epileptic seizures by SVM: the long road ahead to success,” *IFAC Proc. Vol.*, vol. 47, no. 3, pp. 1158–1163, Jan. 2014.

[121] M. Bandarabadi, C. A. Teixeira, J. Rasekhi, and A. Dourado, “Epileptic seizure prediction using relative spectral power features,” *Clin. Neurophysiol.*, vol. 126, no. 2, pp. 237–248, Feb. 2015.

[122] J. Rasekhi, M. R. K. Mollaei, M. Bandarabadi, C. A. Teixeira, and A. Dourado, “Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods,” *J. Neurosci. Methods*, vol. 217, no. 1, pp. 9–16, 2013.

[123] Y. Park, L. Luo, K. K. Parhi, and T. Netoff, “Seizure prediction with spectral power of EEG using cost-sensitive support vector machines,” *Epilepsia*, vol. 52, no. 10, pp. 1761–1770, Oct. 2011.

[124] Y. U. Khan and J. Gotman, “Wavelet based automatic seizure detection in intracerebral electroencephalogram,” *Clin. Neurophysiol.*, vol. 114, no. 5, pp. 898–908, May 2003.

[125] N. Kannathal, M. L. Choo, U. R. Acharya, and P. K. Sadasivan, “Entropies for detection of epilepsy in EEG,” *Comput. Methods Programs Biomed.*, vol. 80, no. 3, pp. 187–194, Dec. 2005.

[126] A. Subasi and E. Erçelebi, “Classification of EEG signals using neural network and logistic regression,” *Comput. Methods Programs Biomed.*, vol. 78, no. 2, pp. 87–99, May 2005.

[127] S. B. Wilson, “A neural network method for automatic and incremental learning applied to patient-dependent seizure detection,” *Clin. Neurophysiol.*, vol. 116, no. 8, pp. 1785–1795, Aug. 2005.

[128] R. Hopfengärtner, F. Kerling, V. Bauer, and H. Stefan, “An efficient, robust and fast method for the offline detection of epileptic seizures in long-term scalp EEG recordings,” *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 118, no. 11, pp. 2332–2343, Nov. 2007.

[129] S. Ghosh-Dastidar, H. Adeli, and N. Dadmehr, “Principal Component Analysis-Enhanced Cosine Radial Basis Function Neural Network for Robust Epilepsy and Seizure Detection,” *IEEE Trans. Biomed. Eng.*, vol. 55, no. 2, pp. 512–518, Feb. 2008.

[130] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 1st ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1994.

[131] L. Guo, D. Rivero, J. A. Seoane, and A. Pazos, “Classification of EEG Signals Using Relative Wavelet Energy and Artificial Neural Networks,” in *Proceedings of the First ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, New York, NY, USA, 2009, pp. 177–184.

[132] L. Guo, D. Rivero, and A. Pazos, “Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks,” *J. Neurosci. Methods*, vol. 193, no. 1, pp. 156–163, Oct. 2010.

[133] G. R. Minasyan, J. B. Chatten, M. J. Chatten, and R. N. Harner, “Patient-Specific Early Seizure Detection from Scalp EEG,” *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.*, vol. 27, no. 3, pp. 163–178, Jun. 2010.

[134] C. P. Shen *et al.*, “Epileptic Seizure Detection for Multichannel EEG Signals with Support Vector Machines,” in *2011 IEEE 11th International Conference on Bioinformatics and Bioengineering*, 2011, pp. 39–43.

[135] D. Wang, D. Miao, and C. Xie, “Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection,” *Expert Syst. Appl.*, vol. 38, no. 11, pp. 14314–14320, Oct. 2011.

[136] Y. Uzzaman Khan, N. Rafiuddin, and O. Farooq, “Automated seizure detection in scalp EEG using multiple wavelet scales,” in *2012 IEEE International Conference on Signal Processing, Computing and Control (ISPCC)*, 2012, pp. 1–5.

[137] M. Z. Parvez and M. Paul, “Epileptic seizure detection by analyzing EEG signals using different transformation techniques,” *Neurocomputing*, vol. 145, pp. 190–200, Dec. 2014.

[138] M. Shoaib, K. H. Lee, N. K. Jha, and N. Verma, “A 0.6-107 µW Energy-Scalable Processor for Directly Analyzing Compressively-Sensed EEG,” *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 61, no. 4, pp. 1105–1118, Apr. 2014.

[139] A. Kavitha and V. Krishnaveni, “A novel automatic stepwise signal processing based computer aided diagnosis system for epilepsy-seizure detection and classification for EEG,” *Biomed. Res.*, vol. 0, no. 0, 2016.
[140] X. Liu, A. Jiang, and N. Xu, “Automated epileptic seizure detection in EEGs using increment entropy,” in *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2017, pp. 1–4.
[141] E. D. Übeyli, “Support vector machines for detection of electrocardiographic changes in partial epileptic patients,” *Eng. Appl. Artif. Intell.*, vol. 21, no. 8, pp. 1196–1203, Dec. 2008.
[142] M. B. Malarvili* and M. Mesbah, “Newborn Seizure Detection Based on Heart Rate Variability,” *IEEE Trans. Biomed. Eng.*, vol. 56, no. 11, pp. 2594–2603, Nov. 2009.
[143] T. M. E. Nijsen, J. B. A. M. Arends, P. A. M. Griep, and P. J. M. Cluitmans, “The potential value of three-dimensional accelerometry for detection of motor seizures in severe epilepsy,” *Epilepsy Behav. EB*, vol. 7, no. 1, pp. 74–84, Aug. 2005.
[144] A. Van de Vel *et al.*, “Long-term home monitoring of hypermotor seizures by patient-worn accelerometers,” *Epilepsy Behav.*, vol. 26, no. 1, pp. 118–125, Jan. 2013.
[145] B. Pisano *et al.*, “Autosomal dominant nocturnal frontal lobe epilepsy seizure characterization through wavelet transform of eeg records and self organizing maps,” in *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2016, pp. 1–6.
[146] B. Pisano, A. Fanni, C. A. Teixeira, and A. Dourado, “Application of self organizing map to identify nocturnal epileptic seizures,” in *2017 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM)*, 2017, pp. 1–7.
[147] J. Klatt *et al.*, “The EPILEPSIAE database: an extensive electroencephalography database of epilepsy patients,” *Epilepsia*, vol. 53, no. 9, pp. 1669–1676, Sep. 2012.
[148] H. Jasper, “The 10-20 electrode system of the International Federation.,” *Electroencephalogr Clin Neurophysiol 1958 10 370-375*.
[149] C. A. Teixeira *et al.*, “EPILAB: A software package for studies on the prediction of epileptic seizures,” *J. Neurosci. Methods*, vol. 200, no. 2, pp. 257–271, Sep. 2011.
[150] M. H. Kryger, *Atlas of Clinical Sleep Medicine.* Elsevier Health Sciences, 2013.
[151] M. Kubat and S. Matwin, “Addressing the Curse of Imbalanced Training Sets: One-Sided Selection,” in *Proceedings of the Fourteenth International Conference on Machine Learning*, 1997, pp. 179–186.
[152] P. Refaeilzadeh, L. Tang, and H. Liu, “Cross-Validation,” in *Encyclopedia of Database Systems*, Springer, Boston, MA, 2009, pp. 532–538.
[153] R. Ferri, O. Bruni, S. Miano, and M. G. Terzano, “Topographic mapping of the spectral components of the cyclic alternating pattern (CAP),” vol. 6, no. 1, pp. 29–36, Jan. 2005.
[154] B. Cannas, A. Fanni, G. Pautasso, G. Sias, and P. Sonato, “An adaptive real-time disruption predictor for ASDEX Upgrade,” *Nucl. Fusion*, vol. 50, no. 7, p. 075004, 2010.