



**Politecnico  
di Torino**



**UNIVERSITÀ  
DEGLI STUDI  
DI TORINO**

Doctoral Dissertation  
Doctoral Program in Bioengineering and Medical-Surgical Sciences (33<sup>rd</sup> Cycle)

# **Analysis of Signal Decomposition and Stain Separation methods for biomedical applications**

**Nicola Michielli**

\*\*\*\*\*

**Supervisor**

Full Professor Filippo Molinari

**Doctoral Examination Committee:**

Prof. V. Bevilacqua, Referee, Department of Electrical and Information Engineering, Politecnico di Bari

Prof. V. Corino, Referee, Department of Electronics, Information and Bioengineering, Politecnico di Milano

Prof. M. Casadio, Referee, Department of Informatics, Bioengineering, Robotics and Systems Engineering, University of Genoa

Prof. A. Cereatti, Department of Electronics and Telecommunications, Politecnico di Torino

Politecnico di Torino - University of Turin  
April 8<sup>th</sup>, 2021

This thesis is licensed under a Creative Commons License, Attribution - Noncommercial - NoDerivative Works 4.0 International: see [www.creativecommons.org](http://www.creativecommons.org). The text may be reproduced for non-commercial purposes, provided that credit is given to the original author.

I hereby declare that the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

*Nicola Michielli*

.....  
Nicola Michielli  
Turin, April 8<sup>th</sup>, 2021

*“If people do not believe  
that mathematics is simple,  
it is only because they do not  
realize how complicated life is”*

*John von Neumann*

# Abstract

Nowadays, the biomedical signal processing and classification and medical image interpretation play an essential role in the detection and diagnosis of several human diseases. The problem of high variability and heterogeneity of information, which is extracted from digital data, can be addressed with the separation of single components from non-stationary signals and multi-stained images. In this context, signal decomposition and stain separation techniques can be useful approaches to highlight hidden patterns or rhythms in biological signals and specific cellular structures in histological color images, respectively. The decomposition strategy is often a pre-processing step that enables the effective extraction of quantitative data from biomedical signals and medical images in the development of computer-assisted diagnosis (CAD) systems to overcome the limitation of time-consuming manual processes which are also subjected to intra- and inter-operator variability.

This thesis work can be divided into two macro-sections: firstly, a novel automatic sleep stage classification (ASSC) system based on single-channel electroencephalographic (EEG) signals and recurrent neural network (RNN) architecture is presented and several signal decomposition techniques are employed to extract significant features from sub-signals with different frequency content. Secondly, a novel multi-tissue and multiscale automated solution for the separation and normalization of histological stains in the field of digital histopathology image analysis is proposed and validated.

In the first part (Part I), a novel cascaded RNN model based on long short-term memory (LSTM) blocks is presented with the aim to classify sleep stages automatically. Sleep scoring is a time-consuming and difficult task which is manually performed by sleep experts. A general workflow based on single-channel EEG signals is developed to enhance the low performance in staging N1 sleep without reducing the performances in the other sleep stages (i.e. Wake, N2, N3 and REM). In the same context, several signal decomposition techniques and time-frequency representations are deployed for the analysis of EEG signals. All extracted features are analyzed by using a novel correlation-based timestep feature selection and finally the selected features are fed to a bidirectional RNN model. The results of network architectures are tested by performing a robust stratified cross-validation strategy.

In the second part (Part II), a fully automated method named SCAN (Stain Color Adaptive Normalization) is proposed for the separation and normalization of staining in digital pathology. This normalization system allows to standardize digitally, automatically and in a few seconds, the color intensity of a tissue slide with respect to that of a target image considered as a gold standard. The stain normalization strategy is useful to increase the speed and accuracy of the pathologist's diagnosis, to reduce delays due to re-staining of histological preparations and can be considered as a pre-processing method for subsequent CAD systems developed for accurate cellular structure segmentation, classification and quantification of histological prognostic parameters, due to the reduction of enormous variability of non-optimal tissue color intensity. Multiscale evaluation and multi-tissue comparison are performed for assessing the robustness of the proposed method. In addition, a stain normalization based on a novel mathematical technique, named ICD (Inverse Color Deconvolution) is developed for immunohistochemical (IHC) staining in histopathological images.

In conclusion, the proposed techniques achieve satisfactory results compared to state-of-the-art methods in the same research field. The workflow proposed in this thesis work and the developed algorithms can be employed for the analysis and interpretation of other biomedical signals and for digital medical image analysis.

# Contents

<b>1. Introduction - Part I</b> .....	<b>1</b>
1.1 Sleep Analysis .....	1
1.2 Automatic Sleep Stage Classification system .....	6
1.2.1 Data acquisition and signal pre-processing .....	7
1.2.2 Feature extraction .....	8
1.2.3 Feature selection or Dimensionality reduction .....	9
1.2.4 Classification .....	9
1.3 Aim of the Thesis - Part I .....	10
References .....	12
<b>2. Automated Sleep Stage Classification of single-channel EEG signals using Recurrent Neural Networks</b> .....	<b>16</b>
2.1 Introduction .....	17
2.2 Materials and Methods .....	20
2.2.1 Data acquisition .....	21
2.2.2 Signal pre-processing.....	22
2.2.3 Feature extraction .....	22
2.2.4 Feature selection .....	25
2.2.5 Dimensionality reduction.....	26
2.2.6 Recurrent Neural Networks .....	27
2.3 Results .....	31
2.4 Discussion and Conclusion.....	34
References .....	38
<b>3. Analysis of Signal Decomposition methods and timestep Feature Selection</b> .....	<b>42</b>
3.1 Introduction .....	42
3.2 Materials and Methods .....	44
3.2.1 Data acquisition and pre-processing .....	45
3.2.2 Signal decomposition.....	46
3.2.3 Time-Frequency distributions.....	51
3.2.4 Correlation-based timestep feature selection.....	54
3.2.5 Bidirectional Recurrent Neural Networks .....	56
3.3 Results .....	57
3.4 Discussion and Conclusion.....	59
References .....	63

<b>4. Introduction - Part II</b> .....	<b>67</b>
4.1 Digital Pathology.....	67
4.2 Color Deconvolution .....	70
4.3 Aim of the Thesis - Part II.....	74
References .....	76
<b>5. Stain Separation and Normalization of Histological Images in Digital Pathology</b> .....	<b>78</b>
5.1 Introduction .....	79
5.1.1 Color deconvolution-based methods .....	80
5.1.2 Histogram transformation-based methods .....	81
5.1.3 Iterative optimization-based methods.....	82
5.2 Materials and Methods .....	83
5.2.1 Preliminary stain separation.....	85
5.2.2 Cellular structures segmentation.....	86
5.2.3 Final stain separation .....	87
5.2.4 Stain normalization.....	88
5.3 Experiments and Results .....	89
5.3.1 Stain separation performance.....	90
5.3.2 Stain normalization performance.....	94
5.4 Discussion and Conclusion.....	99
References .....	103
<b>6. Inverse Color Deconvolution for Stain Normalization in Immunohistochemistry</b> .....	<b>106</b>
6.1 Introduction .....	107
6.2 Materials and Methods .....	110
6.2.1 Stain estimation and separation .....	111
6.2.2 Inverse Color Deconvolution.....	114
6.3 Experiments and Results .....	115
6.4 Discussion and Conclusion.....	118
References .....	121
<b>7. Conclusions and Final Remarks</b> .....	<b>124</b>
<b>8. List of Contributions</b> .....	<b>125</b>



# Chapter 1

## Introduction - Part I

### 1.1 Sleep Analysis

Humans spend about one third of their existence sleeping and when deprived of it, they feel an urgent need, just like drinking or eating. It was shown that deprivation of sleep, also known as sleeplessness, leads to death faster than food deprivation, although the mechanisms are still unknown [1]. Sleep can therefore be considered a fully-fledged vital function, like nutrition, reproduction and so on, even if its functional role has not been clarified yet. The sleep quality can be compromised by sleep-related disorders such as sleep apnea, depression, insomnia, narcolepsy, breathing-related and circadian rhythm disorders [2], [3]. In addition, the chronic restriction of sleep has in fact serious negative consequences both on the health of the individual (e.g. at the level of metabolic, immune, and psychiatric balance) and on public security [4]. There are several accidents at work and when driving due to slow reflexes caused by drowsiness [5]. The National Highway Traffic Safety Administration had proved that during driving the main consequence of drowsiness was the reduction of reaction times and this caused between 56,000 and 100,000 car accidents, with more than 71,000 injuries and 1500 deaths each year in USA [6]. In recent decades, the development of new technologies in the field of neuroscience has allowed a more detailed study on the mechanisms that regulate human sleep. The relationship between alertness level and sleepiness, which is well known at the experience level by anyone, has been proven and clinically studied using several techniques and indicators. There are objective and validated assessments of vigilance state, which can be measured by numerous psychological tests. A widely used alertness index for its simple execution is the measurement of reaction times. There are several devices and tests dedicated to the measurement of reaction times, including the psychomotor vigilance task (PVT), developed in 1985, which measures over a period of time the rapidity of the subject's response to a visual stimulus [7]. On the other hand, a difference exists

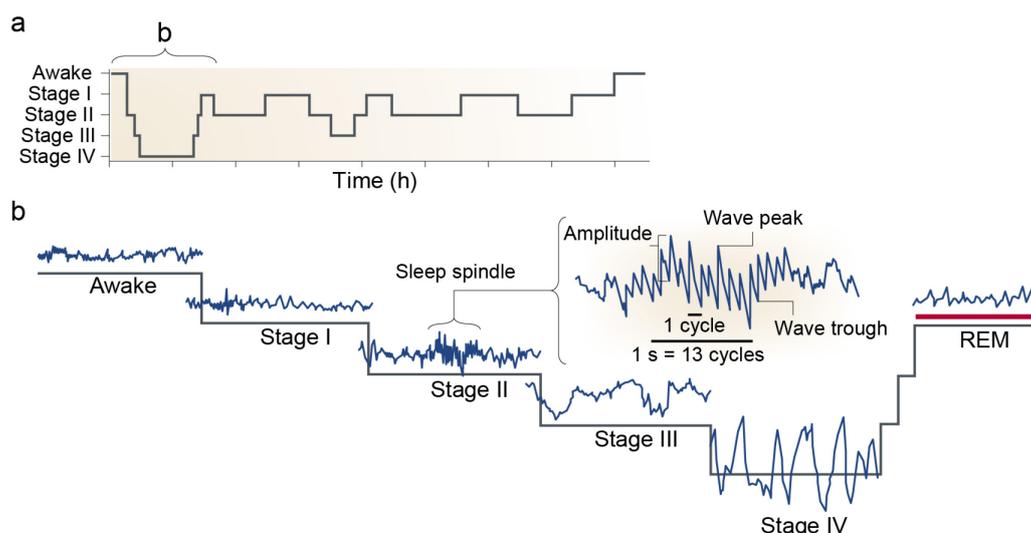
between subjectively perceived sleepiness, measurable with various validated and internationally diffused questionnaires, and the drowsiness objectively measurable in the sleep laboratory, thanks to the detection of parameters such as heart rate, respiratory rate, blood pressure, eye movements, muscle tone and others, but above all with the study of the encephalographic (EEG) signals. In fact, specific markers of sleepiness have been identified in the EEG signal, including the presence of slow frequencies, specific waveforms and variations in the power spectrum of the various frequencies both globally and focused on individual brain areas [8].

The EEG signal is the electrical biopotential recorded on the scalp due to the variation of the electric field generated by cortical neurons. The functioning of neurons is poorly synchronized in a normal condition, and for this reason the EEG is considered a highly nonlinear and non-stationary random process and the recognition of specific patterns in time domain is very complex [9]. The EEG signal can be employed for the monitoring of brain activity since it can be accessible in a non-invasive way and due to its high temporal resolution. The information extracted in time domain from the EEG signals is not as significant as that contained in frequency domain since EEG has a limited frequency bandwidth in the range of 0.3-80 Hz. From a clinical point of view and for the sleep analysis, the interesting band is between 0.5 and 45 Hz and is split into six psychological sub-bands: delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma or the sleep spindle sub-band (12-16 Hz), beta (16-30 Hz) and gamma sub-band ( $> 30$  Hz). The lower the frequency range of a sub-band, the higher the amplitude of the EEG signal, and the more mental activity associated to that sub-band is reduced. Therefore delta and theta sub-bands are related to sleep, drowsiness or pathological conditions, instead alpha sub-band occurs when the subject is relaxed or starts to fall asleep and finally beta and gamma sub-bands are recorded during concentration or attention related processes [10]. There is also a high inter-subject variability for the non-stationary characteristics of this signal.

The process of falling asleep cannot be considered as an abrupt transition from wakefulness to sleep but as a process that involves a series of steps and gradual changes between the relaxed waking state and deep sleep. The macrostructure of sleep refers to the various sleep stages that occur and repeat in relatively regular cycles during sleep [11]. These stages, and the characteristics that allow them to be detected and scored are by convention accepted and used worldwide, to ensure diagnostic and therapeutic uniformity. These sleep scoring guidelines were proposed in 1968 by Rechtschaffen and Kales (R&K) [12] and were updated in 2007 by the American Academy of Sleep Medicine (AASM) [13]. The process of sleep scoring is a time-consuming and difficult task which is manually performed by sleep experts. A sleep expert is a professional figure in the sleep laboratory which is trained in sleep technology and relevant aspects of sleep medicine. The classification of sleep stages is performed through the analysis and interpretation of polysomnographic (PSG) signals: the EEG signals for the study of brain activity, the electrooculographic (EOG) signals used to detect the amplitude and rapidity of eye movements and the electromyographic (EMG) signals to quantify mylohyoid

muscle tone. In addition, other biomedical signals can be recorded to study respiration, heart rate, oxygen saturation, body temperature and limb movements since they can provide useful information on the diagnosis of sleep-related diseases.

The main difference between the guidelines proposed by R&K and the AASM manual is the number of scored sleep phases: the AASM manual classifies sleep into 5 stages, one less than R&K's guidelines, since in the AASM manual the two R&K stages S3 and S4 are merged into a single sleep class, named slow wave sleep (SWS) stage [14]. The summary diagram for the sleep cycle according to R&K and the EEG signals for each sleep stage, is reported in Figure 1.1.



**Fig. 1.1** Summary diagram of the sleep cycle (a) and EEG signals for each of the 6 sleep stages scored according to the R&K's standard (b).

The more recent sleep scoring system proposed by the AASM classified sleep into three main types of phases: wakefulness (stage W), non-rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep. NREM sleep, which represents almost 75-80% of the whole sleep cycle duration of a healthy individual, is in turn split into three stages (N1, N2 and N3) each of which is denoted by specific PSG patterns, although the distinction between one stage and another is not always well defined. For a healthy human subject, the N1 sleep is the shortest phase in sleep cycle (about 2-5%), while stage N2 constitutes the most lasting one (about 45-55% of the total sleep episode) [15]. By convention, a single PSG epoch scored by the sleep expert according to the AASM manual [16] lasts 30 seconds and the staging refers to the following indications:

- **Stage W:** more than 50% of the EEG activity in a single epoch at the occipital level must consist of alpha rhythm (i.e. with a frequency between 8 and 12 Hz, typical of wakefulness with closed eyes). In the EOG recording, movements typical of wakefulness, i.e. rapid and irregular, are noted and muscle tone is high. However, it should be noted that this phase includes both active wakefulness related to concentration and attention (characterized by high-

frequency beta waves) and early stages of falling asleep, so the EEG morphology is found to be extremely variable depending on the condition of the subject.

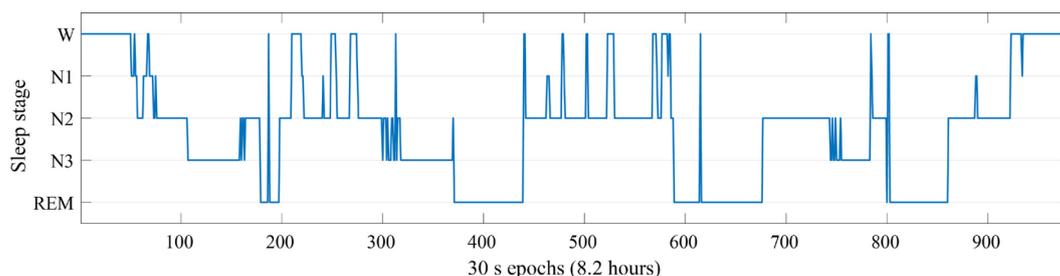
- **Stage N1:** represents the transition between awake state and sleep, where the alpha activity is replaced by mixed frequency waves (mostly in theta rhythm, between 4 and 8 Hz) with low amplitude. This must be present for more than 50% of the whole 30 s epoch. Eye movements are slower and more rhythmic than in stage W and muscle tone decreases. In EEG recording may be identified some sharply contoured and negative-going bursts called vertex sharp waves, lasting less than 0.5 s, particularly pronounced on the central brain regions and distinguishable from the underlying activity.
- **Stage N2:** at the EEG level a rhythm continues to be found mainly of theta type, with low voltage waves and mixed frequencies but predominantly between 4 and 8 Hz. The muscle tone continues to decrease, and the eyes present slow and wide movements. The presence of specific EEG patterns is what distinguishes this stage from the previous one. These waves are the sleep spindles (i.e. a train of visible waves especially at central level with frequency between 12 and 16 Hz and duration greater than 0.5 s) and the K-complexes (visible at the frontal level, are characterized by a positive deflection followed immediately by a negative component of duration greater than 0.5 s and clearly distinguishable from normal background activity).
- **Stage N3:** also called SWS stage, represents the deep sleep and EEG activity must be dominated by low-frequency delta rhythms and high amplitude (i.e. greater than 75  $\mu$ V) waves. The delta rhythm must represent more than 20% of the scored epoch. Nevertheless, sleep spindles may be still found. Typically, there are no eye movements and muscle activity is very low.
- **Stage REM:** low voltage and mixed frequency EEG activity must be present; muscle tone must be almost abolished and the presence of rapid and irregular eye movements typical of this phase must be identified. Sawtooth waves (i.e. sharp wave trains with a rhythm between 2 and 6 Hz) may be found in the EEG signal, especially at central level. In addition, very short myoclonic jerks (less than 0.25 s) may be superimposed on existing background muscle atony. Other physiological variables such as blood pressure, heart and respiratory rate are highly irregular during REM sleep.

The typical patterns for each sleep stage according to the AASM rules are listed in Table 1.1. As evident in Table 1.1, sleep scoring is a very difficult and complex procedure, because a robust discrimination between the five different sleep stages is not provided. In addition, the EEG patterns may be present or not in the scored epoch (is not a necessary or sufficient condition) and they are not mutually exclusive for each stage; hence this represents a crucial issue in the context. AASM manual can be used as a methodology to standardize the sleep scoring but the expertise of the neurologists and sleep specialists makes the difference in most of critical aspects of diagnosis.

**Table. 1.1** Characteristic waves for each sleep stage.

Stage	Scoring criteria (EEG)
<b>W (Wakefulness)</b>	Alpha waves (8-12 Hz), Beta waves (16-30 Hz)
<b>N1 (NREM 1)</b>	Theta waves (4-8 Hz), Vertex sharp waves may be present
<b>N2 (NREM 2)</b>	Theta waves, high voltage biphasic waves (K-complexes), Sleep spindles (12-16 Hz)
<b>N3 (NREM 3)</b>	High amplitude ( $> 75 \mu\text{V}$ ) Delta waves (0.5-4 Hz)
<b>REM</b>	Theta waves, Alpha and Sawtooth waves (2-6 Hz) may be present

From the study of the sequence and duration of the various sleep stages throughout the night, the sleep expert obtains the hypnogram, which is a graphical representation of the sleep stages as function of time and allows to obtain quantitative data related to subject's sleep. Among these, the parameters that are most clinically evaluated are the total time of sleep, time of falling asleep, duration of nocturnal wakefulness, the latency of the first REM phase, the total number of REM phases and the percentage of sleep spent in the various stages [16]. An example of the hypnogram is shown in Figure 1.2.

**Fig. 1.2** Manually labelled hypnogram of a whole night.

Typically, during the night, 75-80% is represented by NREM sleep and the remaining part (i.e. 20-25%) constitutes REM sleep. In generally, an adult subject has between 4 and 6 successive sleep cycles every night. Every cycle lasts between 70 and 120 minutes, but its relative composition varies during the time. The first cycles are dominated by the stages of deep sleep, while REM lasts a few minutes at the beginning and gradually increases over the subsequent cycles until the final part of the night, where the sleep cycles are composed predominantly of stage N2 and stage REM [15]. As evident from Figure 1.2, there are several epochs scored as wakefulness in addition to the initial and final part of the hypnogram; thus, the subject wakes up during the night and the number of nocturnal W epochs is very important for the diagnosis of sleep-related disorders. In addition, stage N1 rarely occurs during the night.

A novel and interesting topic in neurology and for research applications is the assessment of a subject's neurocognitive performance (NCP). The NCP can be defined as the process or mental action through which a subject acquires knowledge, perception, intuition, and reasoning [17]. An example of NCP linked to sleep deprivation is the case of a human subject who, while driving a vehicle,

decreases his alertness and becomes a danger to himself and others. The change in NCP is related to changes in brain activity in the phases of falling asleep, which are evaluated through the analysis of the EEG signal. For this reason, the study of the variation of NCP starts with the analysis of the variations in the EEG signal during the sleep stages (according to the previously described AASM standard) and then moves on to a more detailed analysis of the hypnagogic state of the subject, i.e. the transition between wakefulness and the first stage of sleep. In this context, Hori et al. proposed in 1994 a novel classification system for the hypnagogic state and defined an indicator called sleep onset period (SOP) which quantifies the time in the transition from alertness to drowsiness [18]. In their work, they introduced nine EEG hypnagogic stages, where the first two stages belonged to the awake state, the subsequent six stages were referred to stage N1 and the last one was the stage N2 according to the AASM gold standard. Hori and Tanaka focused their attention on the first stage of sleep (i.e. stage N1 in the AASM manual) since it represents the first moments when a subject changes his/her NCP by decreasing attention levels and reaction times consequently [18]. The problem of this new scoring is that the classification rules are not very robust and the EEG patterns are not discriminatory in each phase; in fact in the following years this new standard for the hypnagogic state has not been considered by sleep experts a sleep standard as the guidelines proposed by R&K and updated by the AASM.

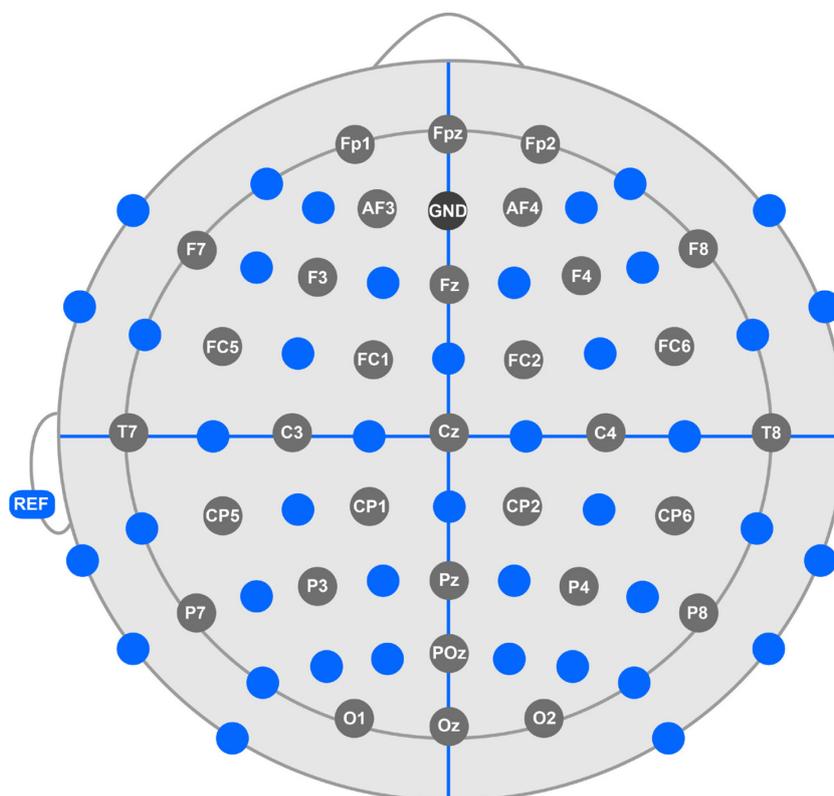
## 1.2 Automatic Sleep Stage Classification system

Sleep scoring is a complex process performed by sleep experts or sleep technologists who, according to the AASM criteria, assign a score (between W, N1, N2, N3 or REM) to a 30 s epoch, after having visually examined the PSG signals of a human subject. This process is repeated for all the 30 s epochs within the whole night recording. The time employed by the expert to score a full night is about 1-3 hours. In addition, the expert may sometimes request a second opinion for scoring agreement or to diagnose a sleep-related disease. Hence, sleep scoring is time-consuming, often tedious and suffers from intra- and inter-operator variability [19]. For this reason, in literature, several authors have proposed automatic sleep stage classification (ASSC) systems with low computational times to support sleep experts in this difficult and complex task. The main structure of these automated systems consists of four modules for the analysis and classification of PSG signals: *i*) data acquisition and signal pre-processing, *ii*) feature extraction, *iii*) feature selection or dimensionality reduction and *iv*) classification [20]. All these steps are deepened in the following sections. In addition, the ASSC systems can be based on multi-channel or single-channel approach. In the multi-channel approach, several EEG channels corresponding to the electrodes placed on the subject's scalp during the acquisition are employed for the subsequent classification process and in addition the EOG or EMG recording may be added to the processing. This kind of solution has the shortcoming of limiting the movement of the subjects during the signal acquisition. On the other hand, in the single-channel approach, it is very

important the choice of the channel to use, in order to be able to extract the most significant information from the data, while the experimental setup can be simplified and the crosstalk effect due to multiple electrodes can be reduced.

### 1.2.1 Data acquisition and signal pre-processing

The international standard 10-20 system is used for the electrode placement for the EEG signal acquisition in the sleep laboratory. The 10-20 system is based on the following anatomical landmarks: nasion, inion and right/left preauricular points used as reference points. All the electrodes are positioned at different distances equal to 10 or 20% of the length of the curve that connects nasion and inion. The electrodes placed in the left and right hemispheres are denoted with odd and even numbers, respectively. High numbers are related to lateral parts of the scalp, while low numbers are closer to midline. Mid-sagittal positions are denoted with the subscript z (e.g. Fz, Cz, Pz, Oz electrode locations) [14]. Figure 1.3 shows the electrode placement according to the standard 10-20 system.



**Fig. 1.3** Electrode placements in a 32-channel layout according to the 10–20 system.

Several electrodes are applied on the scalp, at pre-frontal, frontal, temporal, central parietal, and occipital level bilaterally, and referential electrodes are typically placed on mastoid, ear or chin for the EEG recording. The EEG signal acquisition can be bipolar if voltage difference is recorded between adjacent electrodes or referential if the recording is performed with respect to a common reference. For the EMG signal, mental or submental placements are employed, while two

electrodes for the EOG recording are placed on the outer canthus of the eye, offset 1 cm below and 1 cm above the horizon [14].

The analog signals are converted in a digital form and stored in the European Data Format (EDF), which is the standard file format intended for storage of multichannel physical and biological signals. Finally, all EEG time-series data are band-pass filtered in the range of 0.5 - 45 Hz to retain only physiological bandwidth relevant for sleep analysis and to attenuate the power line interference. The whole night recording is time-segmented into 30 s epochs for sleep scoring and the epochs which contain artifacts (e.g. eye blinking) are rejected.

## 1.2.2 Feature extraction

Features are employed to summarize raw data with the aim to minimize the loss of relevant information. The feature extraction process consists in computing mathematical parameters which contain quantitative information about the signal in different domains. The feature vector is a set of characteristics used to describe signal patterns with the aim of reducing the dimensional space necessary to explain EEG data. These characteristics are called handcrafted features and can be extracted from the whole 30 s epoch or can be computed over a smaller time period and then averaged or concatenated. The time duration of the EEG signals from which to extract information is defined by searching a window where the EEG can be at least considered a wide-sense stationary process and, at the same time, satisfying the frequency resolution according to Heisenberg's uncertainty principle. The extracted features are subsequently fed to an automated classifier which tries to find a relation between these features in order to correctly classify the five sleep stages. The main crucial point is the link between the extracted handcrafted features and the clinical problem. Independently on the used classifier, there could be a set of features that are more significant. Such features contain the “signature” of the physiological system and, if present, of the pathology.

Several domains can be used to extract quantitative information from the pre-processed signals. The most used in literature are four: time, frequency, time-frequency and nonlinear/complexity domain [20]. Features extracted in time domain (e.g. statistical moments, signal energy, etc.) are parameters extracted from the pre-processed EEG signals without any transformation in a new space but simply as function of time. Frequency-domain features (e.g. relative spectral power, mean frequency, etc.) are parameters extracted after estimating the power spectral density (PSD) which associates the relative power to each frequency component included in the original signal [21]. The frequency domain can be very useful for sleep analysis since each frequency sub-band of the EEG signal corresponds to a particular physiological condition of the subject, i.e. awake, drowsiness or deep sleep. Time-frequency distributions (e.g. Choi-Williams distribution, continuous wavelet transform, etc.) are mathematical tools of spectral analysis which represent the signal spectrum instantaneously, i.e. as a function of time and are intended for non-stationary signals. Features are extracted from a bidimensional plane where the

horizontal axis is the time and the vertical axis denotes the frequency [22]. Finally, nonlinear/complexity domain features (e.g. approximate entropy, Lyapunov exponent, correlation dimension, etc.) can be extracted from the EEG signals in the phase space. These features are complexity measures which quantify the nonlinear dynamics and the chaotic behavior of a non-stationary signal [23].

Typically, all these features are extracted from the EEG signals, especially in ASSC systems based on single-channel architecture. The EMG or EOG signals are excluded from the analysis, since the aim is to extract information only from the EEG which is regarded as the principal indicator of NCP.

### **1.2.3 Feature selection or Dimensionality reduction**

The objective of this step is to come up with the smallest set of features that best captures the characteristics of the problem being addressed. However, not all ASSC systems contain this step in their workflow but when the number of extracted features is very high, this step is necessary to reduce the computational time of the classification algorithm. In addition, the removal of not significant features could increase the classification accuracy since redundant features can be seen as an additive noise on the data and consequently are prone to misclassifications [20]. Features are selected both on the basis of how they represent a given class (e.g. a specific sleep stage) and on the basis of how they can distinguish between two or multiple classes (e.g. the five sleep stages scored with AASM manual). There must be a balance between expressiveness and compactness of the feature set.

There are two main types of techniques to consider in selecting features: either by removing irrelevant and redundant information (feature selection) or by finding a combination of the previously extracted feature set (dimensionality reduction). Sequential forward selection (SFS) and its counterpart sequential backward selection (SBS) methods [24], ReliefF and RReliefF algorithms [25] and minimum redundancy maximum relevance (mRMR) technique [26] are examples of feature selection algorithms. On the other hand, linear discriminant analysis (LDA) [27] and principal component analysis (PCA) [28] are two most common methods with the goal of finding a new lower dimensional subset that can effectively summarize the data.

The feature selection or dimensionality reduction step could be very useful for the following classification task since as the dimensionality increases, the searching space grows exponentially, and the input data fed to classifier become increasingly sparse.

### **1.2.4 Classification**

Various classification algorithms have been proposed to automatically score the 30 s EEG epoch as stage W, N1, N2, N3 or REM. These classifiers are able to build boundaries, which can be linear or nonlinear, around feature vectors in order to separate the features which belong to different classes. In addition, some of them try to minimize a cost function which quantifies the difference between the

predicted output and the true label. Some of the popular machine learning algorithms are the following: gaussian mixture model (GMM) [29], support vector machine (SVM) [30]–[33], K-means clustering [34], J-means clustering [35], random forest classifier [22], bootstrap aggregating (Bagging) [36], [37] and artificial neural networks (ANNs) [38], [39]. In recent years, with the advent of deep learning, which is a subset of artificial intelligence (AI), new algorithms have been proposed with the aim of simulating human like decision making. The main applications of deep learning are speech recognition, natural language processing (NLP) and image recognition [40]. Recurrent neural networks (RNNs) are used to process and classify sequential data such as text, genomes, handwriting and also time series [21], [41], [42]. Special architectures of RNNs have been designed to learn long-term dependencies: RNNs with gated recurrent units (GRUs) [43], with long short-term memory (LSTM) blocks [21], [44] and bidirectional RNNs (BRNNs) [45], [46]. On the other hand, convolutional neural networks (CNNs) are employed for image segmentation and classification and in the context of sleep staging, time-frequency transforms can be treated as bidimensional images for CNN-based applications [47], [48].

### **1.3 Aim of the Thesis - Part I**

The aim of the work presented here is to develop and validate an automatic sleep stage classification (ASSC) system based on single-channel electroencephalographic (EEG) signals and recurrent neural network (RNN) architecture. The workflow proposed in this part (Part I) consists of four main modules (i.e. data acquisition and pre-processing, feature extraction, feature selection and classification) and can also be applied to other biomedical signals with the aim of extracting relevant patterns for classification purposes.

The ASSC systems with low computational times are intended to support sleep experts in scoring sleep, which is a complex process, time-consuming (e.g. the time employed by the expert to score a full night is about 1-3 hours), often tedious and suffers from intra- and inter-operator variability. The single-channel approach alleviates the number of wire connections during the recording process with the consequence of reducing sleep disturbances and avoiding any limits for the movement of the subject during signal acquisition. Consequently, the experimental setup can be simplified especially when multi-channel recording is not suitable for particular patients or situations and the crosstalk effect due to multiple electrodes can be reduced.

In Chapter 2, a novel cascaded RNN with long short-term memory (LSTM) units is presented with the aim of automatically classifying sleep stages. To the best of our knowledge, this was the first study published in literature which employed a cascaded RNN architecture based on LSTM units for the automatic scoring of sleep. The aim is to define a general workflow based on single-channel EEG signals to enhance the low performance in staging N1 sleep without reducing the performances in the other four sleep stages (i.e. W, N2, N3 and REM). The question

related to N1 stage detection emerges in the context of neurocognitive performance (NCP) evaluation; in fact, the first step in testing the capabilities to track subtle NCP changes should be the analysis of the hypnagogic state of the subject, i.e. the transition between wakefulness and the first stage of sleep (i.e. stage N1) [18]. In fact, from a clinical point of view, the sleep onset period (SOP) which quantifies the time in the transition from alertness to drowsiness represents a potential physiological indicator of sleep quality which can be compromised by sleep-related disorders such as sleep apnea, depression, insomnia, narcolepsy, breathing-related and circadian rhythm disorders [11].

In Chapter 3, in the same context of EEG sleep scoring, several signal decomposition techniques were analyzed and employed in order to extract significant features from sub-signals with different frequency content. In addition, two time-frequency representations were deployed for the analysis of EEG signals. The segmentation of time-frequency transforms has been performed based on the frequency bands of the rhythms of EEG signals (i.e. delta, theta, alpha, sigma, beta and gamma sub-band) and for each sub-band were extracted several features in time-frequency domain. The entire set of features was analyzed by using a novel correlation-based timestep feature selection strategy and finally the selected features were fed to bidirectional RNN model.

The conclusions and final remarks of this work are reported in the last section of the thesis.

## References

- [1] J. M. Siegel, "Clues to the functions of mammalian sleep," *Nature*, vol. 437, no. 7063, pp. 1264–1271, 2005, doi: 10.1038/nature04285.
- [2] C. F. Reynolds and R. O'Hara, "DSM-5 Sleep-Wake Disorders Classification: Overview for Use in Clinical Practice," *Am. J. Psychiatry*, vol. 170, no. 10, pp. 1099–1101, 2013, doi: 10.1176/appi.ajp.2013.13010058.
- [3] M. M. Ohayon, "Epidemiology of insomnia: what we know and what we still need to learn," *Sleep Med. Rev.*, vol. 6, no. 2, pp. 97–111, 2002, doi: <https://doi.org/10.1053/smr.2002.0186>.
- [4] J. S. Higgins *et al.*, "Asleep at the Wheel-The Road to Addressing Drowsy Driving," *Sleep*, vol. 40, no. 2, Feb. 2017, doi: 10.1093/sleep/zsx001.
- [5] P. Philip *et al.*, "Fatigue, sleep restriction and driving performance," *Accid. Anal. Prev.*, vol. 37, no. 3, pp. 473–478, 2005, doi: <https://doi.org/10.1016/j.aap.2004.07.007>.
- [6] A. Garcés Correa, L. Orosco, and E. Laciari, "Automatic detection of drowsiness in EEG records based on multimodal analysis," *Med. Eng. Phys.*, vol. 36, no. 2, pp. 244–249, 2014, doi: <https://doi.org/10.1016/j.medengphy.2013.07.011>.
- [7] D. F. Dinges and J. W. Powell, "Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations," *Behav. Res. Methods, Instruments, Comput.*, vol. 17, no. 6, pp. 652–655, 1985, doi: 10.3758/BF03200977.
- [8] A. A. Putilov and O. G. Donskaya, "Association of an individual's ability to overcome desire to fall asleep with a higher anterior-posterior gradient in electroencephalographic indexes of sleep pressure," *Int. J. Psychophysiol.*, vol. 113, pp. 23–28, 2017, doi: <https://doi.org/10.1016/j.ijpsycho.2017.01.002>.
- [9] M.-T. Lo, P. H. Tsai, P.-F. Lin, C. Lin, and Y.-L. Hsin, "The Nonlinear and nonstationary Properties in EEG Signals: Probing the Complex Fluctuations by Hilbert-Huang Transform.," *Adv. Adapt. Data Anal.*, vol. 1, pp. 461–482, Jul. 2009, doi: 10.1142/S1793536909000199.
- [10] M. Roohi-Azizi, L. Azimi, S. Heysieattalab, and M. Aamidfar, "Changes of the brain's bioelectrical activity in cognition, consciousness, and some mental disorders," *Med. J. Islam. Repub. Iran*, vol. 31, p. 53, Sep. 2017, doi: 10.14196/mjiri.31.53.
- [11] R. D. Ogilvie, "The process of falling asleep.," *Sleep Med. Rev.*, vol. 5, no. 3, pp. 247–270, Jun. 2001, doi: 10.1053/smr.2001.0145.
- [12] 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, doi: 10.1046/j.1440-1819.2001.00810.x.
- [13] R. S. Rosenberg and S. Van Hout, "The American Academy of Sleep Medicine Inter-scoring Reliability Program: Sleep Stage Scoring," *J. Clin. Sleep Med.*, vol. 9, no. 1, pp. 81–87, Jan. 2013, doi: 10.5664/jcsm.2350.
- [14] R. K. Malhotra and A. Y. Avidan, "Chapter 3 - Sleep Stages and Scoring Technique," S. Chokroverty and R. J. B. T.-A. of S. M. (Second E. Thomas, Eds. St. Louis: W.B. Saunders, 2014, pp. 77–99.

- [15] H. R. Colten and B. M. Altevogt, *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. National Academies Press (US), 2006.
- [16] C. Iber, S. Ancoli-Israel, A. L. Chesson, and S. Quan, “The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications,” *Westchester, Am. Acad. Sleep Med.*, Jan. 2007.
- [17] A. Gevins, “Non-invasive human neurocognitive performance capability testing method and system,” *U.S. Patent*, no. 5,295,491, 1994.
- [18] H. Tanaka, M. Hayashi, and T. Hori, “Topographical characteristics and principal component structure of the hypnagogic EEG,” *Sleep*, vol. 20, no. 7, pp. 523–534, Jul. 1997, doi: 10.1093/sleep/20.7.523.
- [19] K. M. Gunnarsdottir *et al.*, “A novel sleep stage scoring system: Combining expert-based features with the generalized linear model,” *J. Sleep Res.*, vol. 29, no. 5, p. e12991, Oct. 2020, doi: <https://doi.org/10.1111/jsr.12991>.
- [20] R. Boostani, F. Karimzadeh, and M. Nami, “A comparative review on sleep stage classification methods in patients and healthy individuals,” *Comput. Methods Programs Biomed.*, vol. 140, pp. 77–91, 2017, doi: <https://doi.org/10.1016/j.cmpb.2016.12.004>.
- [21] N. Michielli, U. R. Acharya, and F. Molinari, “Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals,” *Comput. Biol. Med.*, vol. 106, pp. 71–81, 2019, doi: 10.1016/j.combiomed.2019.01.013.
- [22] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, “Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier,” *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 10–19, 2012, doi: <https://doi.org/10.1016/j.cmpb.2011.11.005>.
- [23] U. R. Acharya, O. Faust, N. Kannathal, T. Chua, and S. Laxminarayan, “Non-linear analysis of EEG signals at various sleep stages,” *Comput. Methods Programs Biomed.*, vol. 80, no. 1, pp. 37–45, 2005, doi: <https://doi.org/10.1016/j.cmpb.2005.06.011>.
- [24] P. Pudil, J. Novovičová, and J. Kittler, “Floating search methods in feature selection,” *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994, doi: [https://doi.org/10.1016/0167-8655\(94\)90127-9](https://doi.org/10.1016/0167-8655(94)90127-9).
- [25] M. Robnik-Šikonja and I. Kononenko, “Theoretical and Empirical Analysis of ReliefF and RReliefF,” *Mach. Learn.*, vol. 53, no. 1, pp. 23–69, 2003, doi: 10.1023/A:1025667309714.
- [26] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, 2005, doi: 10.1109/TPAMI.2005.159.
- [27] Q. Du and N. H. Younan, “Dimensionality Reduction and Linear Discriminant Analysis for Hyperspectral Image Classification BT - Knowledge-Based Intelligent Information and Engineering Systems,” 2008, pp. 392–399.
- [28] I. Jolliffe, “Principal Component Analysis BT - International Encyclopedia of Statistical Science,” M. Lovric, Ed. Springer Berlin Heidelberg, 2011, pp. 1094–1096.
- [29] U. R. ACHARYA, E. C.-P. CHUA, K. C. CHUA, L. I. M. C. MIN, and T. TAMURA, “ANALYSIS AND AUTOMATIC IDENTIFICATION OF SLEEP STAGES USING HIGHER ORDER SPECTRA,” *Int. J. Neural*

- Syst.*, vol. 20, no. 06, pp. 509–521, Dec. 2010, doi: 10.1142/S0129065710002589.
- [30] M. Sharma, D. Goyal, P. V Achuth, and U. R. Acharya, “An accurate sleep stages classification system using a new class of optimally time-frequency localized three-band wavelet filter bank,” *Comput. Biol. Med.*, vol. 98, pp. 58–75, 2018, doi: <https://doi.org/10.1016/j.compbiomed.2018.04.025>.
- [31] C. S. Huang, C. L. Lin, W. Y. Yang, L. W. Ko, S. Y. Liu, and C. T. Lin, “Applying the fuzzy c-means based dimension reduction to improve the sleep classification system,” in *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2013, pp. 1–5, doi: 10.1109/FUZZ-IEEE.2013.6622495.
- [32] T. Lajnef *et al.*, “Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines,” *J. Neurosci. Methods*, vol. 250, pp. 94–105, 2015, doi: <https://doi.org/10.1016/j.jneumeth.2015.01.022>.
- [33] S. Seifpour, H. Niknazar, M. Mikaeili, and A. M. Nasrabadi, “A new automatic sleep staging system based on statistical behavior of local extrema using single channel EEG signal,” *Expert Syst. Appl.*, vol. 104, pp. 277–293, 2018, doi: <https://doi.org/10.1016/j.eswa.2018.03.020>.
- [34] X. Shuyuan, W. Bei, Z. Jian, Z. Qunfeng, Z. Junzhong, and M. Nakamura, “An improved K-means clustering algorithm for sleep stages classification,” in *2015 54th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, 2015, pp. 1222–1227, doi: 10.1109/SICE.2015.7285326.
- [35] J. L. Rodríguez-Sotelo, A. Osorio-Forero, A. Jiménez-Rodríguez, D. Cuesta-Frau, E. Cirugeda-Roldán, and D. Peluffo, “Automatic Sleep Stages Classification Using EEG Entropy Features and Unsupervised Pattern Analysis Techniques,” *Entropy*, vol. 16, no. 12, pp. 6573–6589, 2014, doi: 10.3390/e16126573.
- [36] A. R. Hassan, S. K. Bashar, and M. I. H. Bhuiyan, “On the classification of sleep states by means of statistical and spectral features from single channel Electroencephalogram,” in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2015, pp. 2238–2243, doi: 10.1109/ICACCI.2015.7275950.
- [37] A. R. Hassan and M. I. H. Bhuiyan, “Computer-aided sleep staging using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and bootstrap aggregating,” *Biomed. Signal Process. Control*, vol. 24, pp. 1–10, 2016, doi: <https://doi.org/10.1016/j.bspc.2015.09.002>.
- [38] S. Charbonnier, L. Zoubek, S. Lesecq, and F. Chapotot, “Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging,” *Comput. Biol. Med.*, vol. 41, no. 6, pp. 380–389, 2011, doi: <https://doi.org/10.1016/j.compbiomed.2011.04.001>.
- [39] M. E. Tagluk, N. Sezgin, and M. Akin, “Estimation of sleep stages by an artificial neural network employing EEG, EMG and EOG,” *J. Med. Syst.*, vol. 34, no. 4, pp. 717–725, Aug. 2010, doi: 10.1007/s10916-009-9286-5.
- [40] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
- [41] Y.-L. Hsu, Y.-T. Yang, J.-S. Wang, and C.-Y. Hsu, “Automatic sleep stage recurrent neural classifier using energy features of EEG signals,” *Neurocomputing*, vol. 104, pp. 105–114, 2013, doi:

- <https://doi.org/10.1016/j.neucom.2012.11.003>.
- [42] N. F. Güler, E. D. Übeyli, and İ. Güler, “Recurrent neural networks employing Lyapunov exponents for EEG signals classification,” *Expert Syst. Appl.*, vol. 29, no. 3, pp. 506–514, 2005, doi: <https://doi.org/10.1016/j.eswa.2005.04.011>.
- [43] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” *arXiv Prepr. arXiv/1412.3555*, 2014, Accessed: Jul. 18, 2018. [Online]. Available: <http://arxiv.org/abs/1412.3555>.
- [44] S. Hochreiter and J. Schmidhuber, “Long short-term memory.,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–80, Nov. 1997, Accessed: Jul. 18, 2018. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9377276>.
- [45] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997, doi: 10.1109/78.650093.
- [46] Ö. Yildirim, “A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification,” *Comput. Biol. Med.*, vol. 96, pp. 189–202, 2018, doi: <https://doi.org/10.1016/j.combiomed.2018.03.016>.
- [47] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J.-F. Payen, “A convolutional neural network for sleep stage scoring from raw single-channel EEG,” *Biomed. Signal Process. Control*, vol. 42, pp. 107–114, 2018, doi: <https://doi.org/10.1016/j.bspc.2017.12.001>.
- [48] L. Wei, Y. Lin, J. Wang, and Y. Ma, “Time-Frequency Convolutional Neural Network for Automatic Sleep Stage Classification Based on Single-Channel EEG,” in *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2017, pp. 88–95, doi: 10.1109/ICTAI.2017.00025.

## Chapter 2

# Automated Sleep Stage Classification of single-channel EEG signals using Recurrent Neural Networks

*Part of this chapter has been published as:*

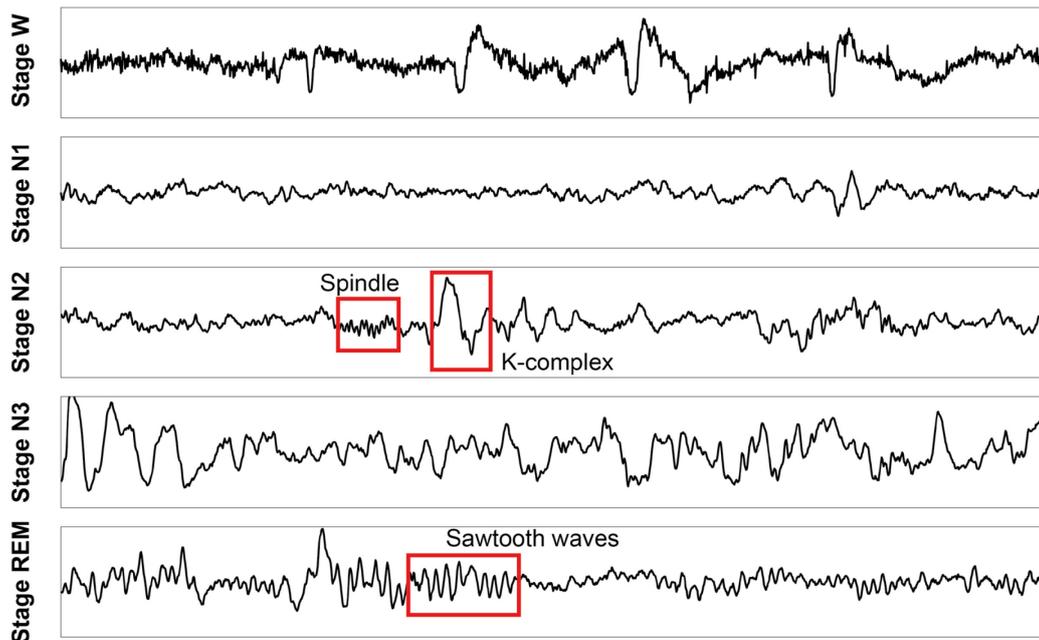
N. Michielli, U. R. Acharya and F. Molinari. "**Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals.**" *Computers in Biology and Medicine* 106 (2019): 71-81.

## 2.1 Introduction

As discussed in Chapter 1, an interesting topic in clinical and neurophysiological studies is the automated assessment of the neurocognitive performance (NCP) of a human subject. The NCP can be defined as the process or mental action through which a subject acquires knowledge, perception, intuition, and reasoning [1]. The critical issue is the development of study or research protocols in which the change of subject's NCP is clearly predictable and repeatable. The different phases of falling asleep may be considered as an example of temporal changes of NCP; hence sleep represents a suitable test case for developing NCP techniques. Humans spend about one third of their existence sleeping and when deprived of it, they feel an urgent need, just like drinking or eating. Sleep can therefore be considered a fully-fledged vital function, like nutrition, reproduction and so on, even if its functional role has not been clarified yet. The sleep quality can be compromised by sleep-related disorders such as sleep apnea, depression, insomnia, narcolepsy, breathing-related and circadian rhythm disorders [2]. The chronic restriction of sleep has serious negative consequences both on the health of the individual (e.g. at the level of metabolic, immune, and psychiatric balance) and on public security [3]. There are several accidents at work and when driving due to slow reflexes caused by drowsiness [4], [5]. In this context, the concept of sleep onset period (SOP) which is the time that elapses between the awake state and the drowsiness, becomes relevant for NCP studies [6].

Sleep scoring, considered a tedious, complex and time-consuming task performed by sleep technologists, consists in classifying the 30 s epochs contained in a subject's full night recording into five classes according to the recent rules introduced by the American Academy of Sleep Medicine (AASM) in 2007 [7] which update the manual written in 1968 by Rechtschaffen and Kales (R&K) [8]. In this manual, sleep is split into three main phases: wakefulness (stage W), non-rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep. NREM sleep is in turn split into three sub-stages (N1, N2 and N3). The five sleep stages are described in detail in Chapter 1. For the classification of sleep, a set of polysomnographic (PSG) signals are recorded, i.e. eye movements (EOG signals), muscle tone (EMG signals) and others, but above all, the encephalographic (EEG) signals. In fact, specific markers of sleepiness have been identified in the EEG signal, including the presence of slow frequencies, specific waveforms, and variations in the power spectrum of the various frequencies both globally and focused on individual brain areas. In addition, the information extracted from EEG signals and its classification and interpretation reflects primary NCP changes [9].

The AASM rules define some typical EEG patterns related to each of five sleep stages as listed in Chapter 1 and reported in Figure 2.1. The problem is that the distinction between one stage and another is not always well defined. For this reason, in literature, several authors have proposed automatic sleep stage classification (ASSC) systems with low computational times to support sleep experts in this difficult and complex task.



**Fig. 2.1** EEG patterns (Sleep Spindle, K-complex and Sawtooth waves).

The main structure of ASSC systems consists of four modules for the analysis and classification of EEG signals: *i*) data acquisition and signal pre-processing, *ii*) feature extraction, *iii*) feature selection or dimensionality reduction and *iv*) classification [10]. The first phase *i*), is the electrode placement on the human scalp according to the international 10-20 system and multiple-channel EEG signals are recorded and converted into digital form [11]. The standard format designed for storing multichannel physical and biological signals is the European Data Format (EDF). A pre-processing step is typically applied to reduce frequency components which are beyond the range of interest from a clinical point of view. In the second phase *ii*), a set of mathematical parameters are extracted from EEG signals and in the third phase *iii*) only the most significant and relevant for sleep analysis information is retained. In the last phase *iv*) a robust classification model is implemented to automatically score the 30 s EEG epoch as stage W, N1, N2, N3 or REM (i.e. it mimics the manual process performed by sleep experts). The robustness of the classification model must be assessed in order to face the high inter-subject variability typical of sleep acquisitions. The first step consists in splitting the original dataset into training and test sets: the test set is employed to estimate how well the model trained on the training data would generalize to new unseen data. The test set represents the unseen data but belongs to the same underlying distribution as the training set. Three main types of validation methods can be applied to test the performances of a classification model: “hold-out” validation, leave-one-out cross-validation (LOOCV) and k-fold cross-validation. The simplest form of validation, called “hold-out” validation consists in a random separation of the available data into a single training and test set. This approach is not a good choice because an unlucky subdivision of the dataset could cause an ineffective network. In the LOOCV, only one subject is retained in the test set and

the remaining subjects are employed to predict the model; the process is repeated until every single sample in the dataset has been used in the test set. This approach has very high computational times when the number of subjects increases. Finally, in k-fold cross-validation, data are randomly split into k mutually exclusive partitions of equal size, and the classification process is repeated k times using in turn different data partitions. This last method gives more robust and reliable estimates by running multiple different training test splits, instead of relying entirely on a single particular training set [12].

The ASSC systems can be multi-channel or single-channel signal based. In the multi-channel approach, several EEG channels corresponding to the electrodes placed on the subject's scalp during the acquisition are employed for the classification process and in addition the EOG or EMG recording could be added to the processing. This solution could be a problem for a portable device for the automatic NCP assessment, for example during driving a car. On the other hand, the single-channel approach alleviates the number of wire connections during the recording process with the consequence of reducing sleep disturbances and avoiding any limits for the movement of the subject during signal acquisition. Consequently, the experimental setup can be simplified especially when multi-channel recording is not suitable for particular patients or situations and the crosstalk effect due to multiple electrodes can be reduced. Therefore, with the use of single-channel EEG signals remains very difficult to improve the performances of sleep stages without relevant information, which is hidden in EOG and EMG signals, but with this approach, the acquisition setup is simpler with less noise interference and it may be used for NCP assessment in real-time applications, e.g. the detection of driver drowsiness.

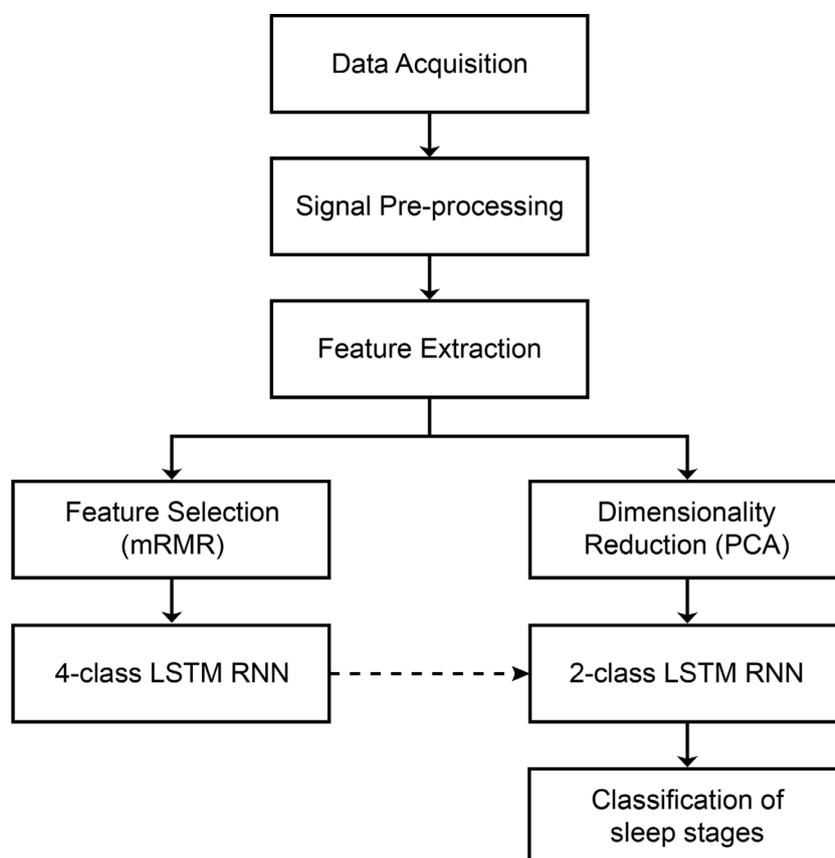
As described in Chapter 1, various research studies have been proposed with the aim of developing ASSC systems to support sleep specialists in this difficult and time-consuming task. The critical issue in most of these works is the accuracy of stage N1 which is the most difficult phase to detect. In this chapter, a novel cascaded recurrent neural network (RNN) with long short-term memory (LSTM) units is presented with the aim of outperforming the limitations of multiple channel recordings and improving the accuracy for N1 sleep. A single EEG channel approach was employed; thus, the choice of the channel was very important to detect most of the EEG patterns for the classification of different sleep stages. To the best of our knowledge, this was the first study published in literature which employed LSTM blocks for the automatic scoring of sleep.

The question related the N1 stage detection emerges in the research field of NCP evaluation; in fact, the first step in testing the capabilities to track subtle NCP changes should be the analysis of the hypnagogic state of the subject, i.e. the transition between wakefulness and the first stage of sleep (i.e. stage N1). In this context, Hori et al. proposed in 1994 a novel classification system for the hypnagogic state and defined stage N1 as the center of the SOP [6]. In their work, they introduced nine EEG hypnagogic stages, where the first two stages belonged to the awake state, the subsequent six stages (i.e. 3-8) were referred to stage N1 and

the last one was the stage N2 according to the AASM gold standard. In fact, from a clinical point of view, the SOP represents a potential physiological indicator of sleep quality which can be compromised by sleep-related disorders such as sleep apnea, depression, insomnia, narcolepsy, breathing-related and circadian rhythm disorders [13]. Therefore, the goal of this work is to define a general workflow to enhance the low performance in staging N1 sleep without reducing the performances in the other four sleep stages (i.e. W, N2, N3 and REM).

## 2.2 Materials and Methods

The proposed approach follows in the first part the standard ASSC workflow, as described in Chapter 1, while in the final part a cascaded network architecture was implemented. The workflow of the proposed study is reported in Figure 2.2.



**Fig. 2.2** Workflow of the proposed automatic sleep stage classification system (mRMR: minimum redundancy maximum relevance, PCA: principal component analysis, LSTM RNN: recurrent neural network with long short-term memory units).

After the first steps of data acquisition, signal pre-processing and feature extraction, two different approaches were developed: for a 4-class classification problem (i.e. W vs N1-REM vs N2 vs N3) with the merging of stages N1 and REM into a single sleep phase N1-REM, the minimum redundancy maximum relevance (mRMR) algorithm was employed for feature selection, while for a 2-class classification problem (N1 vs REM), a dimensionality reduction technique named principal

component analysis (PCA) was used. For both problems, RNNs with LSTM blocks were used for classification purposes. Finally, for the scoring of five sleep stages according to AASM, a cascaded structure for the connection of two RNNs was defined: only the epochs N1-REM correctly classified from the 4-class RNN were fed to the 2-class RNN which performed binary classification. In the following sections, all these steps are described in detail.

### 2.2.1 Data acquisition

The publicly available Sleep-EDF database v1.0.0 published in 2002 and updated in 2007 was used for this study [14], [15]. This database is a collection of PSG signals with corresponding hypnograms annotated by sleep experts. The hypnogram files contained annotations of sleep stages referred to 30 s epochs. The annotation could be sleep stage W, 1, 2, 3, 4 or REM, movement time and unscored. Trained sleep experts manually produced the hypnograms according to the R&K's manual. Sleep EDF signals were relative to full night PSG recording, including EEG (from the Fpz-Cz and Pz-Oz electrode locations according to the international 10-20 system [11]), EOG, oro-nasal respiration, submental chin EMG and rectal body temperature. The EEG signals were sampled with a sampling rate of 100 Hz. The SC (Sleep Cassette) files (*SC4ssnE0-PSG.edf*) contained the PSG signals of subject ss ( $00 \leq ss \leq 19$ ) for night n ( $n=1$  or  $2$ ). In this study, 12 recordings (*SC4001E0*, *SC4002E0*, *SC4011E0*, *SC4012E0*, *SC4021E0*, *SC4031E0*, *SC4051E0*, *SC4061E0*, *SC4112E0*, *SC4122E0*, *SC4131E0*, *SC4182E0*) of 10 healthy Caucasians for age groups between 26 and 33 were selected. The remaining subjects were removed from the analysis due to the presence of several movement and unscored epochs in their recordings. In addition, stage 3 and 4 were merged into stage N3, as the AASM recommends. In this work, only the Fpz-Cz EEG channel was chosen to perform automatic sleep classification tasks, because EEG patterns such as sleep spindles, vertex sharp waves and K-complexes can be noted in frontal/central brain regions as reported in the AASM manual. In addition, the majority of research studies has reported better performances for this channel. In Table 2.1, the number of epochs of 30 seconds fixed duration for each sleep stage, is shown.

**Table 2.1** Dataset composition in various sleep stages.

Sleep stage	W	N1	N2	N3	REM	Total
# Epochs	850	920	4960	1690	1860	10280

As can be seen in Table 2.1, the number of W epochs is relatively small (8.3%) compared to other sleep stages. These epochs were chosen randomly between the subjects and this choice was made since this stage presents EEG patterns and especially EEG frequency content very different with respect to other sleep phases and for this reason, they are easy to classify without the need for selecting too many epochs.

### 2.2.2 Signal pre-processing

After signal acquisition, a pre-processing step is typically employed to reduce the signal frequency bandwidth in the range of 0.5-45 Hz, which, from a clinical point of view, is the most significant. For this purpose, an infinite impulse response (IIR) Butterworth band-pass filter was used. Subsequently, a time-segmentation step was employed to split each EEG epoch into 30 segments of one second duration and in the following feature extraction, a set of features was extracted from these sub-signals. The time duration of the EEG signals from which to extract information was defined by searching a window where the EEG can be at least considered a wide-sense stationary process and, at the same time, satisfying the frequency resolution (i.e. 1 Hz) according to Heisenberg's uncertainty principle. The number of timesteps for the RNN was set equal to 30, accordingly.

### 2.2.3 Feature extraction

The feature extraction process consists in computing mathematical parameters which contain quantitative information about the signal in different domains. In this work we extracted 29 time-domain and 26 frequency-domain features (in total 55 parameters) from the EEG signals relative to Fpz-Cz channel. Features extracted in time domain are parameters extracted from the pre-processed EEG signals without any transformation in a new space but as function of time. Frequency-domain features are parameters extracted after estimating the power spectral density (PSD) which associates the relative power to each frequency component included in the original signal.

Features concerning **time domain** are the following. Statistical moments were firstly extracted from the data, i.e. the 1<sup>st</sup> raw moment or the mean, mathematically defined as

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (\text{Eq. 2.1})$$

the 2<sup>nd</sup> central moment or the variance, defined as

$$\text{var}(x) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (\text{Eq. 2.2})$$

the normalized 3<sup>rd</sup> central moment or the skewness, defined as

$$\text{skew}(x) = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3}{\text{var}(x)^{3/2}} \quad (\text{Eq. 2.3})$$

the normalized 4<sup>th</sup> central moment or the kurtosis, defined as

$$\text{kurt}(x) = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{\text{var}(x)^2} \quad (\text{Eq. 2.4})$$

and the median value. Other parameters were the peak-to-peak amplitude defined as the difference between the wave peak and the wave trough, the absolute maximum value, the number of zero crossings (i.e. the number of points where the

signal has a zero amplitude after changing the sign from negative to positive and vice versa), the root mean square (RMS), defined as

$$RMS(x) = \sqrt{\frac{1}{N} \sum_{i=1}^N |x_i|^2} \quad (\text{Eq. 2.5})$$

the average rectified value (ARV), defined as

$$ARV(x) = \frac{1}{N} \sum_{i=1}^N |x_i| \quad (\text{Eq. 2.6})$$

and the parameters proposed by Hjorth in 1970 [16] called Hjorth mobility (HM), defined as

$$HM(x) = \sqrt{\frac{\text{var}(dx/dt)}{\text{var}(x)}} \quad (\text{Eq. 2.7})$$

and Hjorth complexity (HC), defined as

$$HC(x) = \frac{HM(dx/dt)}{HM(x)} \quad (\text{Eq. 2.8})$$

Subsequently, as discussed in Chapter 1, the interesting band of EEG signals is between 0.5 and 45 Hz and is split into six psychological sub-bands: delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma or the sleep spindle sub-band (12-16 Hz), beta (16-30 Hz) and gamma sub-band (30-45 Hz). Therefore, six different IIR Butterworth band-pass filters were implemented to decompose the 1 s signal into 6 sub-signals relative to the corresponding sub-bands. For each sub-signal was computed the peak-to-peak amplitude obtaining  $A_{delta}$ ,  $A_{theta}$ ,  $A_{alpha}$ ,  $A_{sigma}$ ,  $A_{beta}$  and  $A_{gamma}$  respectively and the same process was repeated for the signal energy, according to this expression

$$Energy = \sum_{i=1}^N |x_i|^2 \quad (\text{Eq. 2.9})$$

obtaining  $E_{delta}$ ,  $E_{theta}$ ,  $E_{alpha}$ ,  $E_{sigma}$ ,  $E_{beta}$  and  $E_{gamma}$  respectively. The last time-domain features were five energy ratios computed as the signal energy in each sub-band (except gamma sub-band) normalized to the energy related to gamma sub-band. A total of 29 time-domain features were extracted for the subsequent analysis.

**Frequency-domain** features are described in the following paragraph. The first step was the computation of PSD starting with the signal in time domain. The non-parametric method proposed by Welch for spectrum estimation [17] was used for this purpose. Then, the mean frequency (MNF) or the spectral centroid defined as

$$MNF = \frac{\sum_i P_i \cdot f_i}{\sum_i P_i} \quad (\text{Eq. 2.10})$$

and the following two entropies called spectral entropy (SE) and Renyi entropy (RE) [18] were employed. The mathematical definitions are reported here:

$$SE = \sum_i p_i \cdot \log\left(\frac{1}{p_i}\right) \quad (\text{Eq. 2.11})$$

$$RE = -\log\left(\sum_i p_i^2\right) \quad (\text{Eq. 2.12})$$

In the previous equations,  $P$  denotes the PSD vector and  $p$  is the power spectrum normalized with respect to its area.

**Table. 2.2** List of 55 time- and frequency-domain parameters extracted from the EEG signal.

N	Feature	N	Feature	N	Feature	N	Feature
1	Peak-to-peak amplitude	15	Amplitude (alpha sub-band)	29	$E_{beta}/E_{gamma}$	43	$P_{delta}/P_{gamma}$
2	Arithmetic mean	16	Amplitude (sigma sub-band)	30	Mean frequency	44	$P_{theta}/P_{alpha}$
3	Absolute maximum value	17	Amplitude (beta sub-band)	31	Spectral entropy	45	$P_{theta}/P_{sigma}$
4	Median	18	Amplitude (gamma subband)	32	Renyi entropy	46	$P_{theta}/P_{beta}$
5	Variance	19	Energy (delta sub-band)	33	Power (delta sub-band)	47	$P_{theta}/P_{gamma}$
6	Skewness	20	Energy (theta sub-band)	34	Power (theta sub-band)	48	$P_{alpha}/P_{sigma}$
7	Kurtosis	21	Energy (alpha sub-band)	35	Power (alpha sub-band)	49	$P_{alpha}/P_{beta}$
8	Root mean square	22	Energy (sigma sub-band)	36	Power (sigma sub-band)	50	$P_{alpha}/P_{gamma}$
9	Average rectified value	23	Energy (beta sub-band)	37	Power (beta sub-band)	51	$P_{sigma}/P_{beta}$
10	Number of zero crossings	24	Energy (gamma subband)	38	Power (gamma sub-band)	52	$P_{sigma}/P_{gamma}$
11	Hjorth mobility	25	$E_{delta}/E_{gamma}$	39	$P_{delta}/P_{theta}$	53	$P_{beta}/P_{gamma}$
12	Hjorth complexity	26	$E_{theta}/E_{gamma}$	40	$P_{delta}/P_{alpha}$	54	$P_{delta} \cdot P_{theta}$
13	Amplitude (delta sub-band)	27	$E_{alpha}/E_{gamma}$	41	$P_{delta}/P_{sigma}$	55	$P_{alpha} \cdot P_{beta}$
14	Amplitude (theta sub-band)	28	$E_{sigma}/E_{gamma}$	42	$P_{delta}/P_{beta}$		

In addition, six relative spectral powers, defined as the ratio between the absolute power in a specific sub-band and the total signal power, were evaluated for the six EEG sub-bands previously described, obtaining  $P_{delta}$ ,  $P_{theta}$ ,  $P_{alpha}$ ,  $P_{sigma}$ ,  $P_{beta}$  and  $P_{gamma}$ . From these relative spectral powers, fifteen power ratios were

computed. The last features in frequency domain were defined as the products of spectral powers in low and high frequency sub-bands, respectively. A total of 26 frequency-domain features were extracted.

All extracted features are listed in Table 2.2. At the end of the feature extraction process, a feature matrix of size 55×30 (i.e. 55 features and 30 timesteps) for each 30 s EEG epoch, was obtained and used as input for the following steps.

### 2.2.4 Feature selection

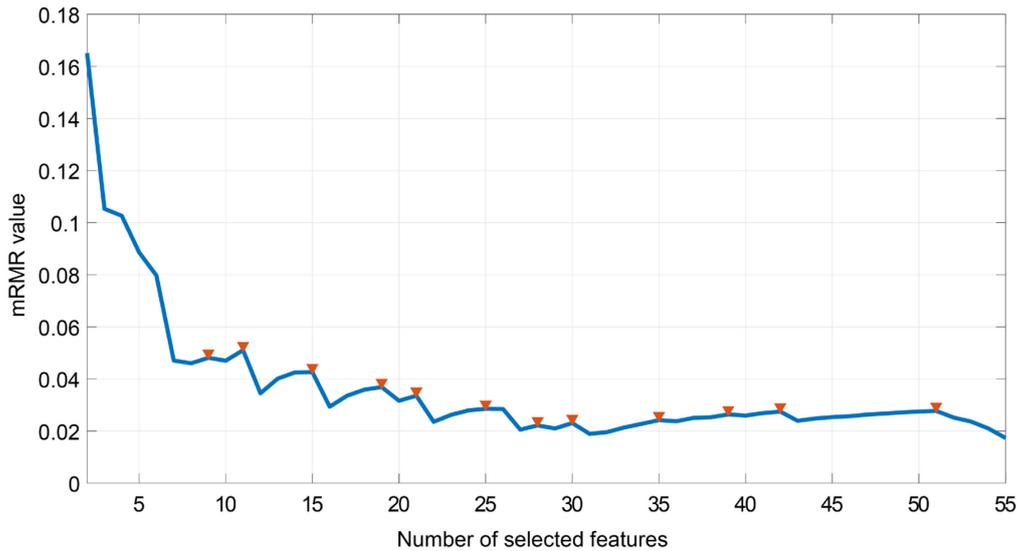
The objective of this step is to come up with the smallest set of features that best captures the characteristics of the problem being addressed. The minimum redundancy maximum relevance (mRMR) algorithm was employed for this purpose with the aim of retaining all significant features for sleep scoring and removing all features strongly cross-correlated with each other. This strategy is based on the information theory according to which the mutual information is defined by considering the probability mass functions  $p(x)$  and  $p(y)$  of two random variables  $X$  and  $Y$  respectively and the joint probability mass function  $p(x, y)$ . Therefore, the mutual information  $I(X; Y)$  is expressed as:

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \cdot \log_2 \frac{p(x, y)}{p(x) \cdot p(y)} \quad (\text{Eq. 2.13})$$

A value equals to zero means that the joint probability can be simply expressed as the product of marginal probabilities with the consequence that the two random variables are independent, while a higher value (the mutual information is not bounded above) denotes an increasing statistical dependence [19]. The goal of mRMR algorithm is to find a subset  $S$  of features which maximizes the following mutual information-based cost function:

$$\phi_{mRMR} = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) - \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \quad (\text{Eq. 2.14})$$

where the operator  $|\cdot|$  denotes the feature cardinality. In Eq. 2.14, the first term, called relevance, is the mean of the mutual information values between each feature  $x_i$  and the label class  $c$  (i.e. the sleep stage), while the second term, called redundancy denotes the cross mutual information between features. The maximization of this function means to have a high relevance and a low redundancy of the feature subset [20]. In this work, the input matrices (features on the rows and timesteps on the columns) for all epochs in the training set were concatenated and with the corresponding class label vector were fed to mRMR algorithm. The most relevant features were selected according to this strategy for ascending numbers of subset (i.e. from 2 to 55) and the optimal feature cardinality was chosen as the global maximum point, as shown in Figure 2.3. As can be seen in Figure 2.3, eleven features were selected for the 4-class classification problem. These features are the following: median value, skewness, RMS, HM, peak-to-peak amplitudes of theta, sigma, beta and gamma sub-band, the energy ratio  $E_{beta}/E_{gamma}$  and the relative spectral powers in theta and beta sub-bands.



**Fig. 2.3** The mRMR value as a function of number of feature subset. The local maxima are highlighted with a red downward-pointing triangle.

### 2.2.5 Dimensionality reduction

The dimensionality reduction strategy tries to define a combination of the previously extracted feature set with the goal of finding a new lower dimensional subset that can effectively summarize the data. In this study, the unsupervised technique called principal component analysis (PCA) was performed for the 2-class classification problem. PCA is a statistical procedure that transforms several correlated variables into uncorrelated *principal components*. This procedure allows to select a few new variables which contain most of the variation of the data and for this reason can be defined as a dimensionality reduction technique. The first step was the data centering: given an input feature matrix for the  $i$ -th 30s EEG epoch and after having computed the mean matrix  $\mu$  and the standard deviation matrix  $\sigma$ , the following feature scaling can be performed:

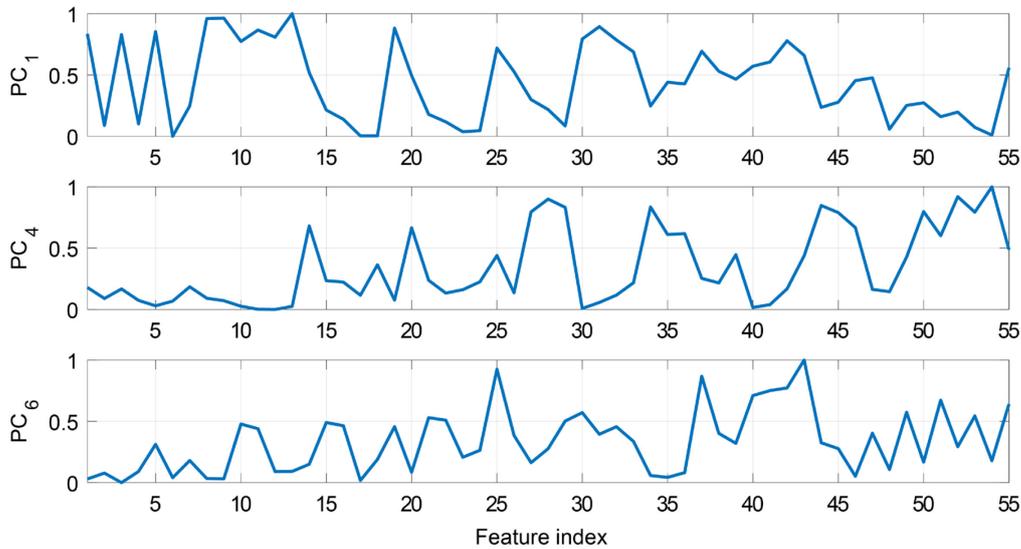
$$X_{norm}^{(i)} = \frac{X^{(i)} - \mu}{\sigma} \quad (\text{Eq. 2.15})$$

Subsequently, this normalized matrix was used as input for PCA, which is based on the singular value decomposition applied to the covariance matrix of data [21]. Dimensionality reduction consists in the selection of a smaller number of principal components, by retaining the eigenvectors corresponding to highest eigenvalues. In this work the 95% was set for the variance criterion and consequently 27 principal components were retained and stored in the transformation matrix  $U$ ; hence, the new data matrix for the  $i$ -th 30s epoch was expressed as [21]:

$$X_{pca}^{(i)} = U^T \cdot X^{(i)} \quad (\text{Eq. 2.16})$$

PCA, differently from the mRMR algorithm, is an unsupervised strategy because it uses as input unlabeled data. The reduced format matrix was fed to the classifier which performed 2-class classification (N1 vs REM). PCA was applied to our dataset and the principal components coefficients with a highest correlation with

the target are reported in Figure 2.4; the normalized value (between 0 and 1) of the weight coefficients of the linear combination of input features is reported.



**Fig. 2.4** Weight coefficients of the three most significant principal components with respect to the original extracted features.

## 2.2.6 Recurrent Neural Networks

In recent years, with the advent of deep learning, which is a subset of artificial intelligence (AI), new algorithms have been proposed with the aim of simulating human like decision making. Deep learning approaches exploit deep neural networks for the processing and recognition of nonlinear information inside input data for pattern analysis and classification. These types of networks are often constructed with a layer-by-layer method and their architecture is much more complex and deeper than traditional networks [22]. The main applications of deep learning are speech recognition, natural language processing (NLP) [23] and image object detection [24]. Recurrent neural networks (RNNs) are powerful deep learning models especially used to process and classify sequential data such as text, genomes, handwriting and also time series [25], [26]. The RNN represents the temporal extension of the basic artificial neural network (ANN) and it is capable of handling time series as digital signals and interpreting structural dependencies for long range applications [27].

The structure of a standard ANN is generally composed of one input layer, several interposed hidden layers and one output layer. Each layer is activated by a transfer function (called *activation function*), that may be linear or, more frequently nonlinear, and it is composed by several hidden units (also called *neurons*) which receive inputs, change their internal activation state and produce output. This process is called *forward propagation*. The ANN is trained by computing the error between the prediction and the desired output, also known as the cost function. In the training phase, the weights and bias term (parameters of each neuron) are iteratively adjusted to minimize the error between the true label and the prediction,

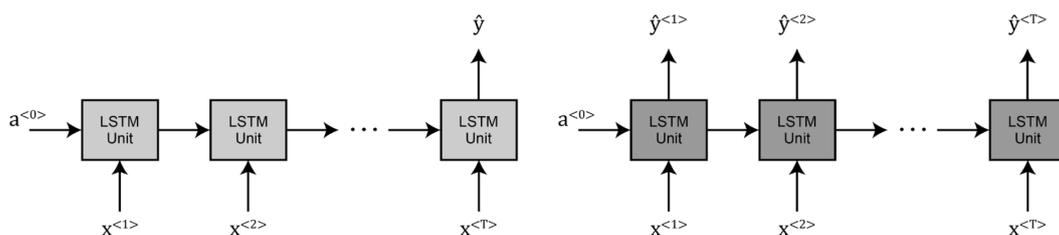
by computing the partial derivatives of the cost function with respect to weights and bias term, starting from the output and propagating back to the input; this process is called *backward propagation* [28].

In the RNN structure, an input sequence  $\{x^{<1>}, \dots, x^{<t>}, \dots, x^{<T>}\}$  with the previously selected features for each time step (from 1 to the last timestep  $T$ ) is used as input from which the model computes the hidden activation sequence  $\{a^{<1>}, \dots, a^{<t>}, \dots, a^{<T>}\}$  and the predicted output sequence  $\{\hat{y}^{<1>}, \dots, \hat{y}^{<t>}, \dots, \hat{y}^{<T>}\}$  for each timestep. The first activation  $a^{<0>}$  is conventionally set to zero. Then, the activation and the output prediction for each timestep are computed using the following equations:

$$a^{<t>} = g(W_a \cdot [a^{<t-1>}, x^{<t>}] + b_a) \quad (\text{Eq. 2.17})$$

$$\hat{y}^{<t>} = g(W_y \cdot a^{<t>} + b_y) \quad (\text{Eq. 2.18})$$

where the operator  $g$  denotes the activation function (which may be different in the two equations),  $W_a, b_a$  and  $W_y, b_y$  are the weight matrices and bias terms for the activation and the output, respectively. As can be seen in Eq. 2.17, the new activation is a function not only of the input at current timestep but also of the activation computed at the previous timestep and for this reason the advantage of RNN respect to standard ANN is to retain information of past temporal series. This structure is called unidirectional RNN which is the first form of RNN proposed for research studies, but also exists the bidirectional form of RNN which combines the forward and backward connection in time for the output prediction [29]. The forward propagation in RNN is described by Eq. 2.17 and Eq. 2.18. The backpropagation process is similar to that of standard ANN where the partial derivatives of cost function are computed in order to update the weights matrices and bias terms for each hidden layer propagating back until the input layer. The difference in RNN architecture is that the backpropagation is performed also going back in time for all timesteps and therefore is called *backpropagation through time* (BPTT). There are two different types of RNN models: sequence-to-sequence and sequence-to-label architecture. In the first model, the network receives a temporal sequence as input and produces a temporal sequence as output (i.e. one output for each timestep) as previously described, while in the second model the input is the same, but the network produces in output a single label after processing all timesteps. Figure 2.5 shows the basic representation of these two types of RNN models.



**Fig. 2.5** Sequence-to-label (left) and sequence-to-sequence (right) architecture.

The main issue related to the network training phase is due to the vanishing gradient, i.e. the partial derivatives through time assume very small values and the unknown variables (i.e. weights and the bias) remains similar to the previous state, the network cannot change its learning performance and consequently the training phase is stopped [30]. The solution to this problem is to replace the standard gated recurrent units (GRUs) of the RNN with long short-term memory (LSTM) units introduced by Hochreiter in 1997 [31]. The LSTM block is composed of five main elements: the memory cell  $c^{<t>}$ , the candidate value  $\tilde{c}^{<t>}$  and three new gates denoted as  $\Gamma_f, \Gamma_u, \Gamma_o$  and called forget, update and output gate, respectively. The function of memory cell is to store for a long-time information useful for the training process; the candidate value at each time step replaces the memory cell and the gates, normalized between 0 and 1, are updated by using specific weight matrices and bias terms, as follows:

$$\Gamma_f = \sigma(W_f \cdot [a^{<t-1>}, x^{<t>}] + b_f) \quad (\text{Eq. 2.19})$$

$$\Gamma_u = \sigma(W_u \cdot [a^{<t-1>}, x^{<t>}] + b_u) \quad (\text{Eq. 2.20})$$

$$\Gamma_o = \sigma(W_o \cdot [a^{<t-1>}, x^{<t>}] + b_o) \quad (\text{Eq. 2.21})$$

where the operator  $\sigma$  denotes the sigmoid activation function. Next, the LSTM equations to update the candidate value, the memory cell and the activation state are the following:

$$\tilde{c}^{<t>} = \tanh(W_c \cdot [a^{<t-1>}, x^{<t>}] + b_c) \quad (\text{Eq. 2.22})$$

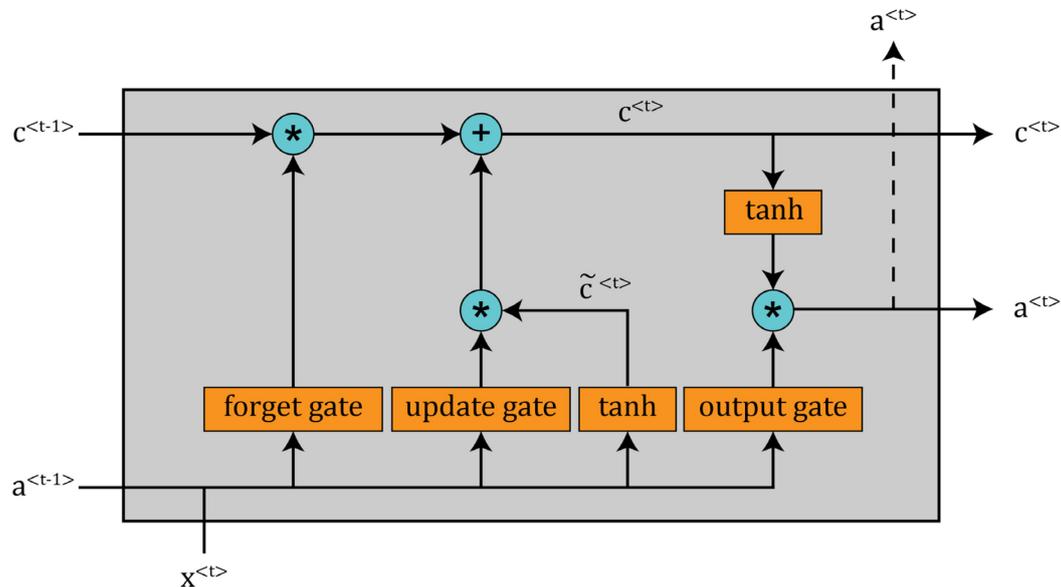
$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>} \quad (\text{Eq. 2.23})$$

$$a^{<t>} = \Gamma_o * \tanh(c^{<t>}) \quad (\text{Eq. 2.24})$$

where the operator  $*$  is the element-wise product and  $W_c, b_c$  are the weight matrix and bias term relative to the cell, respectively. The graphical representation of LSTM block is reported in Figure 2.6.

The EEG is considered a highly nonlinear and non-stationary random process and the recognition of specific patterns in time domain is very complex [32]; thus the choice of using RNNs with LSTM units becomes relevant in this context for their capability of retaining past temporal information. In this study, cascaded LSTM RNNs were implemented to perform 4-class classification and 2-class classification tasks. The first RNN employed the mRMR selected features to classify stage W, N1-REM, N2 and N3, while the second RNN employed as input the principal components to distinguish N1 epoch vs REM epochs. The second network became more specific to classify the two most complex sleep stages, while in the first phase of the process these two classes were merged into the N1-REM class. The two LSTM RNNs had the same architecture: a sequence input layer with 30 timesteps, the hidden LSTM layers which analyzed through time the selected features; the fully connected (FC) layer used for the conversion of previous layer information into the correct output dimension (i.e. the number of sleep stages); the

softmax layer which estimated the probability map for each predicted output and finally the classification layer which quantified the cost function. Figure 2.7 represents the structure for both RNNs employed in this work.



**Fig. 2.6** Graphical representation of long short-term memory (LSTM) operations.

As can be seen in Figure 2.7, the 2-class LSTM RNN is more complex since it is composed of a sequence-to-sequence LSTM layer connected to the input layer, differently from the 4-class LSTM RNN which consists only of a sequence-to-label architecture. The softmax activation function has the useful mathematical property to provide values between 0 and 1 as a cumulative distribution function; its mathematical definition is:

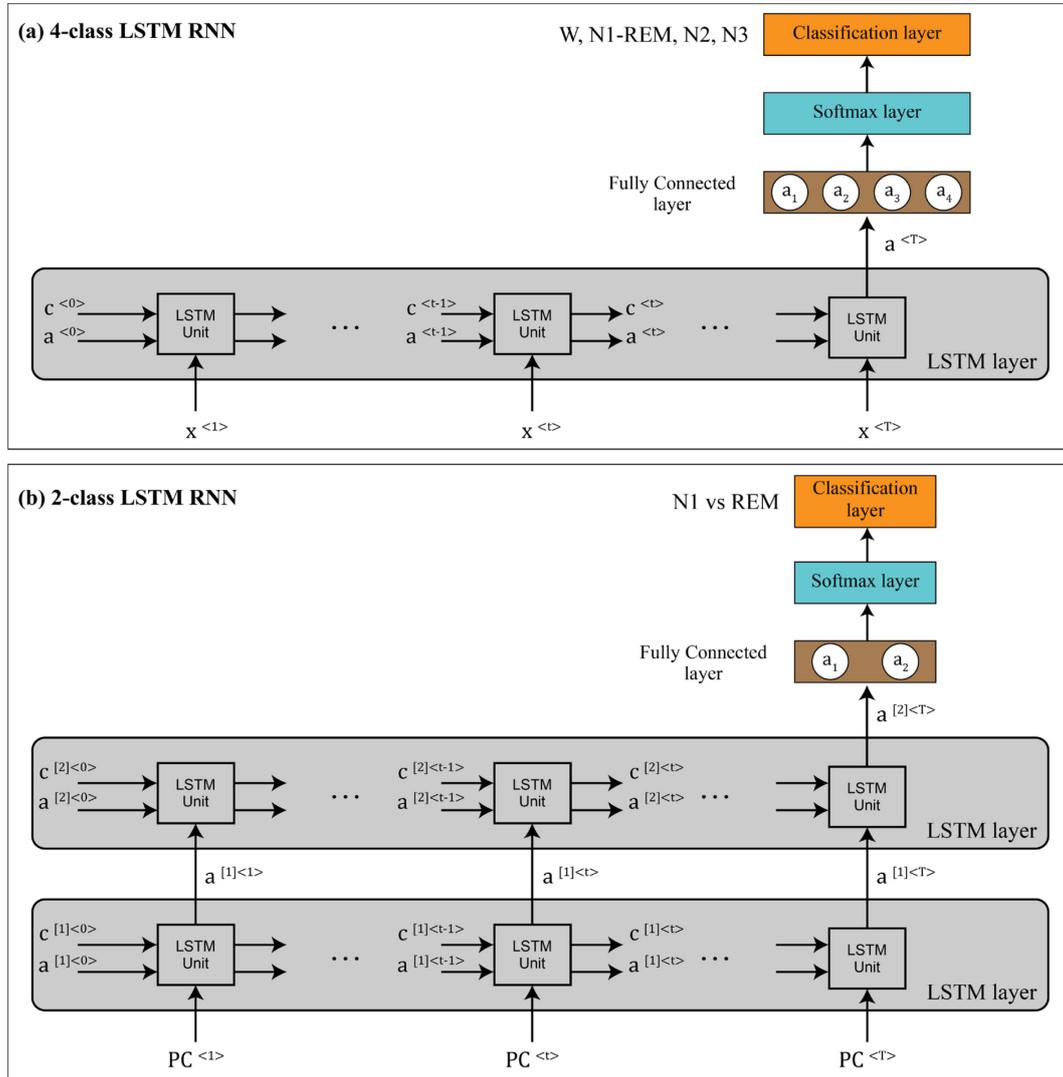
$$\hat{y}_j^{(i)} = \frac{e^{z_j^{(i)}}}{\sum_{j=1}^C e^{z_j^{(i)}}} \quad (\text{Eq. 2.25})$$

where the superscript  $i$  denotes the generic training epoch,  $z_j$  denotes the output of the  $j$ -th neuron of the FC layer and  $C$  is the total number of classes. In this study, the loss function to minimize was the cross-entropy function for  $C$  mutually exclusive classes, defined as follows:

$$J(W, b) = -\frac{1}{M} \sum_{i=1}^M \sum_{j=1}^C y_j^{(i)} \cdot \log(\hat{y}_j^{(i)}) \quad (\text{Eq. 2.26})$$

where  $M$  is the total number of epochs used for training,  $y$  is the true class and  $\hat{y}$  is the output value predicted by the network. As can be seen in Eq. 2.26, the objective expression depends on the unknown parameters  $W$  and  $b$ , which were updated by using the adaptive moment estimation (ADAM) optimization method [33]. ADAM algorithm combines the properties of gradient descent with momentum method and the advantage of root mean square propagation (RMSProp) to work well for non-stationary processes. The ADAM hyperparameters employed in this study were:

learning rate ( $\alpha = 0.001$ ), gradient decay and squared gradient decay factors ( $\beta_1 = 0.9, \beta_2 = 0.999$ ) and, for numerical stability, epsilon ( $\epsilon = 10^{-8}$ ).



**Fig. 2.7** Structure of the LSTM RNN which performs 4-class classification (a) and LSTM RNN which performs binary classification (b).

## 2.3 Results

The processing was performed in MATLAB (The MathWorks, Inc., Natick, MA, USA) environment on a workstation with 16 GB of RAM, 2.5 GHz quad-core CPU and 64-bit Windows version. Several RNNs with LSTM blocks were tested for 4-class and 2-class classification problems. The input size was defined for the two networks according to the number of mRMR features (i.e. 11) and the number of principal components (i.e. 27), respectively. The numbers of LSTM layers and units in each layer were different in all tested architectures and only for the second network, different thresholds for the softmax activation function were tested to assign the predicted class to N1 or REM stage. The default value, of course, was 0.5. The following quantitative metrics were computed to assess the performance of the classifiers: the percentage of correct classification (PCC) defined as the ratio

between the correctly classified data and the total amount of data, and the well-known per-class metrics, i.e. recall or sensitivity (Se), selectivity or specificity (Sp) and accuracy (Acc) for each EEG sleep stage. The best RNN for 4-class classification was chosen in terms of the highest sensitivity for the merged class N1-REM, while the best second network was selected with the aim of obtaining the highest classification performance in stage N1. The best architectures for the 4-class and 2-class LSTM RNN are reported in Table 2.3 and Table 2.4, respectively.

**Table. 2.3** Layer structure for the RNN which identified 4 classes (W vs N1-REM vs N2 vs N3).

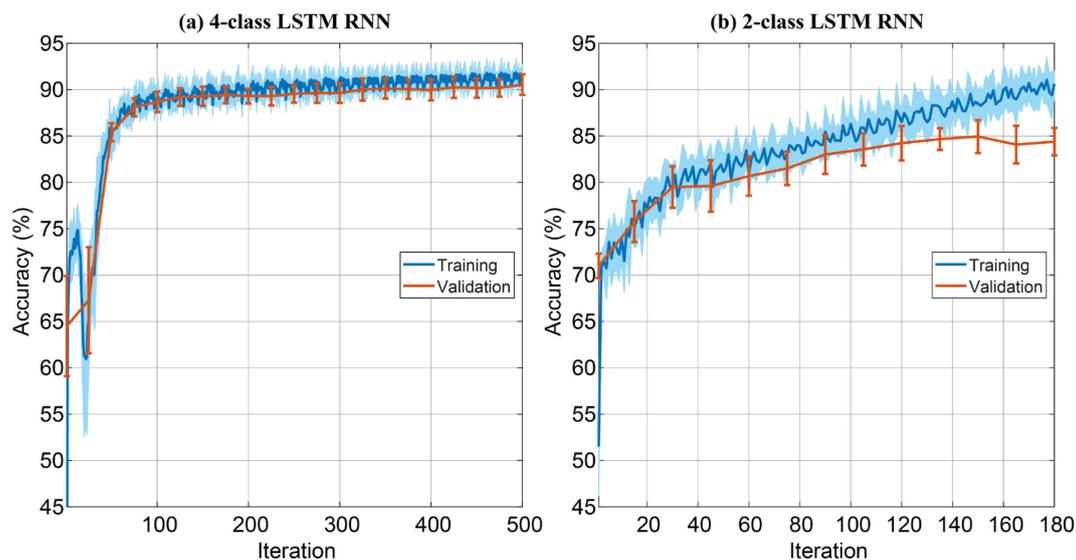
Layer number	Layer type	Properties
Layer 1	Sequence input layer	11 input features
Layer 2	LSTM layer	101 hidden units, sequence-to-label architecture
Layer 3	Fully connected layer	4 units
Layer 4	Softmax layer	softmax activation function
Layer 5	Classification output layer	4 classes

**Table. 2.4** Layer structure for the RNN which identified 2 classes (N1 vs REM).

Layer number	Layer type	Properties
Layer 1	Sequence input layer	27 principal components
Layer 2	LSTM layer	125 hidden units, sequence-to-sequence architecture
Layer 3	LSTM layer	98 hidden units, sequence-to-label architecture
Layer 4	Fully connected layer	2 units
Layer 5	Softmax layer	softmax activation function (threshold = 0.6)
Layer 6	Classification output layer	2 classes

The entire dataset was partitioned into three different parts called training, validation or development and test set. The network learnt using only the training set, then the validation set was used to assess performances and the training was stopped when a maximum number of iterations was reached or when validation error started to increase to avoid data overfitting [28]. The final performance was then assessed on the test set, which the network had never previously seen. To test the performance of the RNN models, we assessed the classification metrics by 10-fold stratified cross-validation strategy. The simplest form of cross-validation, called “hold-out” validation consists in a random separation of the available data into a single training and test set. This approach is not a good choice because an unlucky subdivision of the dataset could cause an ineffective network. In k-fold cross-validation, the data are randomly split into k mutually exclusive partitions of equal size, and the classification process is repeated k times using in turn different

data partitions. The stratification is performed to ensure that each class is proportionally represented in the training, validation, and test set. In this study by applying 10 folds, the 80% of data (i.e. 8224 epochs) was used in the training set, the 10% of data (i.e. 1028 epochs) was used in the validation set and the remaining 10% of data in the test set (i.e. 1028 epochs) randomly. Smaller subsets, named mini batches, were used in the ADAM optimization algorithm to speed up the training process. A value of 512 and 256 for the mini-batch size was set in the training phase for the first and second network, respectively. The plots for training and validation accuracy as function of iterations is reported in Figure 2.8 for both networks.



**Fig. 2.8** Training (in blue) and validation (in red) accuracy of RNN which performs 4-class classification (a) and RNN which performs binary classification (b). The solid line is used to indicate the average value and the shaded area for training and the symmetrical error bars for validation denote the standard deviation within 10 folds.

The feature selection and dimensionality reduction techniques were performed only on training and validation set and subsequently the mRMR selected features and the PCA coefficients were applied to the test data for each fold. The confusion matrix for each network obtained as the sum of the performances in each fold and the per-class metrics are reported in Table 2.5 and Table 2.6. In Table 2.6, only the N1-REM epochs correctly classified by the first network are considered for the performance assessment. The PCC of the first and second RNN with LSTM units are 90.8% and 83.56% respectively. Finally, considering the cascading connection of the two networks, the performances in terms of sensitivity for the five sleep stages are: 95.29% for stage W, 61.09% for stage N1, 89.48% for stage N2, 91.66% for stage N3 and 83.76% for stage REM. The final performances for N1 and REM are computed as the ratio between the number of epochs correctly classified by the 2-class RNN and the total number of epochs in the dataset (i.e. 920 and 1860 for stage N1 and REM, respectively). The overall PCC is 86.74%.

**Table. 2.5** Classification performances for 4-class RNN.

	Predicted				Per-class [%]		
True	W	N1-REM	N2	N3	Se	Sp	Acc
W	810	28	3	9	95.29	99.77	99.40
N1-REM	10	2537	231	2	91.26	95.21	94.14
N2	4	331	4438	187	89.48	93.10	91.35
N3	8	0	133	1549	91.66	97.69	96.70
<b>Percentage of correct classification (PCC) = 90.80%</b>							

**Table. 2.6** Classification performances for 2-class RNN.

	Predicted		Per-class [%]		
True	N1	REM	Se	Sp	Acc
N1	562	202	73.56	87.87	83.56
REM	215	1558	87.87	73.56	83.56
<b>Percentage of correct classification (PCC) = 83.56%</b>					

## 2.4 Discussion and Conclusion

Sleep scoring, considered as a tedious, complex and time-consuming task carried out by sleep technologists, consists in classifying the 30 s epochs contained in a subject's full night recording into five classes according to the guidelines proposed by AASM. The proposed approach is a novel ASSC system based on the first traditional steps of data acquisition, pre-processing and feature extraction, while in the final part of the workflow a cascaded network architecture was implemented. Most of published works in literature are based on single-channel architecture in order to limit the number of wire connections during the recording process with the consequence of reducing sleep disturbances. In addition, in the majority of studies, the publicly available Sleep-EDF database [14], [15] was used for performance comparison. In Table 2.7, the per-class sensitivity and the total PCC of the proposed method and other state-of-the-art single-channel based approaches are reported for comparison (in bold are highlighted the best performances for each sleep stage).

A brief description of each state-of-the-art method is reported in the following. Hsu et al. [26], employed Elman RNN for automatically scoring sleep stages and features based on signal energy were computed from the Fpz-Cz EEG signals. They obtained the best performances in stage N2 and REM. Hassan et al. in their first study [34], used bootstrap aggregating (Bagging) as classifier and time-domain statistical parameters and frequency-domain spectral features were extracted from single-channel EEG signals. They performed feature selection using the non-parametric Kruskal-Wallis statistic and obtained a PCC of 86.53%. In their second study [35], the classifier was the same but, with the application of signal

decomposition technique, which decomposed the EEG signals into intrinsic mode functions (IMFs) using a novel modified version of empirical mode decomposition (EMD), they reached a PCC of 90.69%. Fraiwan et al. [36], extracted entropy features from time-frequency distributions, which were fed to random forest classifier and obtained a PCC of 82.57%. Seifpour et al. [37], employed a multi-class support vector machine (SVM) and a novel statistical parameter, called statistical behavior of local extrema, was extracted from Fpz-Cz signals. They obtained the best classification rate in stage W and consequently the best overall PCC since the percentage of W epochs used in their work was 53% of the entire dataset. Sors et al. [38], proposed a 1D convolutional neural network (CNN) which took as inputs the raw EEG signals without the feature extraction process and a PCC of 86.79% was obtained. Wei et al. [39] estimated time-frequency spectra which were directly fed to CNN for classification purposes and they reached a PCC of 83.93%. In the work of Sharma et al. [40], a feature set based on wavelet decomposition was extracted and the SVM was used as classifier. The performances were robust only for stage W while in the other sleep stages the results were not satisfactory.

**Table. 2.7** Comparison table of state-of-the-art single-channel based methods (results are reported in terms of per-class sensitivity and overall percentage of correct classification in the last column).

Single-channel methods		Results [%]					
Authors	Classifier	W	N1	N2	N3	REM	PCC
Hsu et al. [26]	Elman RNN	70.80	36.70	<b>97.30</b>	89.70	<b>89.50</b>	87.20
Hassan et al. [34]	Bagging	96.60	27.48	82.93	76.92	69.57	86.53
Hassan et al. [35]	Bagging	95.28	47.02	92.38	90.00	80.87	90.69
Fraiwan et al. [36]	Random forest	93.33	43.22	84.76	68.37	76.41	82.57
Seifpour et al. [37]	Multi-class SVM	<b>98.76</b>	40.07	90.94	85.08	83.98	<b>91.82</b>
Sors et al. [38]	CNN	91.40	34.92	89.24	85.08	85.82	86.79
Wei et al. [39]	CNN	92.70	26.66	87.40	87.05	82.74	83.93
Sharma et al. [40]	Multi-class SVM	95.41	17.39	76.38	57.11	36.46	83.92
<b>Proposed method</b> [41]	LSTM RNN	95.29	<b>61.09</b>	89.48	<b>91.66</b>	83.76	86.74

In the proposed work [41], the goal was to define a general workflow based on single-channel EEG signals to improve the low performance in staging N1 sleep without reducing the performances in the other four sleep stages (i.e. W, N2, N3 and REM). After the first traditional steps of data acquisition, signal pre-processing and feature extraction in time and frequency domain, two different approaches were developed: for a 4-class classification problem (i.e. W vs N1-REM vs N2 vs N3) with the merging of stages N1 and REM into a single sleep phase N1-REM, the mRMR algorithm was employed for feature selection, while for a 2-class classification problem (N1 vs REM), the PCA was used as dimensionality reduction

technique. For both problems, RNNs with LSTM blocks were used for classification purpose. Finally, for the sleep scoring of five sleep stages according to AASM, a cascading connection of two previous RNNs was defined. To the best of our knowledge, this was the first study published in literature which employed a cascaded RNN architecture based on LSTM units for the automatic scoring of sleep.

In the pre-processing step, time-segmentation was applied to EEG signals in order to obtain 30 segments of 1s time duration for each 30 s EEG epoch, because of the non-stationary nature of the EEG signals. The feature extraction process in time and frequency domain was performed for each time segment. The main crucial point is the link between the extracted handcrafted features and the clinical problem. Independently on the used classifier, there could be a set of features that are more significant. Such features contain the “signature” of the physiological system and, if present, of the pathology. Therefore, the feature selection process is useful to select only significant and relevant variables for sleep scoring. For the 4-class classification problem (W vs N1-REM vs N2 vs N3) the mRMR method selected the following significant features: the peak-to-peak amplitude, signal energy and relative spectral power in beta sub-band which is a high frequency band typical of stage W, the peak-to-peak amplitude of sleep spindles or sigma sub-band which occurs during N2 sleep, the peak-to-peak amplitude and relative spectral power of theta sub-band which is present during REM, N1 and N2 sleep and finally the Hjorth mobility parameter which assumes a low value for N3 sleep since the high-pass effect of time derivative attenuates the low-frequency components of delta sub-band. For the 2-class classification problem (N1 vs REM) the unsupervised strategy of PCA was implemented and in Figure 2.4 are reported the most significant principal coefficients as function of original feature set. The highest values of weight coefficients in the linear combination can be noted for indices 8, 9 and 13 for PC1, index 54 for PC4 and index 43 for PC6. According to Table 2.2, these indices correspond to RMS value, ARV, peak-to-peak amplitude of delta sub-band, the product  $P_{\delta} \cdot P_{\theta}$  and the ratio  $P_{\delta}/P_{\gamma}$  respectively. The presence of delta sub-band features in the linear combination of PCA is justified by the low-frequency vertex waves that can occur during N1 sleep with respect to the waves typical of stage REM. The quantitative performances, as can be seen in Table 2.5 and Table 2.6, show low classification rates in stage N1 and REM. The issue is that stage N1 is the most difficult phase to detect since no specific and repeatable EEG patterns can be observed in this phase and, for a healthy human subject, the N1 sleep is the shortest phase in sleep cycle (about 2-5%) [42]. On the other hand, REM sleep is characterized by high eye movement activity recorded by EOG and hence with a single EEG channel, the detection process becomes more complex [43]. In addition, the muscle tone recorded by EMG could be useful for the detection of REM sleep since during this stage the muscle activity is lower than other sleep stages. In a future work, a higher number of N1 epochs might be considered for the training but there is a risk that the classification performances of the network on the test set were reduced since in the typical human sleep cycle the occurrence of this phase is the lowest. Moreover, features in time-frequency or nonlinear/complexity

domain [44] could be extracted from the EEG signals, in order to reveal specific patterns in N1 or REM sleep. Table 2.7 shows that the proposed approach obtained the highest value in the detection of N1 and N3 sleep with respect to other single-channel based state-of-the-art methods and at the same time the performances in the other three sleep stages are superior to 83%. In addition, very satisfactory results were obtained in the stage W detection, by using a smaller number of epochs for the training process with respect to other studies.

The limitation of the proposed method is due to the cascaded architecture: from the point of view of N1 detection, the cascaded approach is useful, but on the other hand, the classification error for REM stage can propagate through the second RNN. Another shortcoming with respect to machine learning techniques, is that the computational times of training deep networks can be higher. Nevertheless, the advantage of RNNs is to handle temporal feature sequences and add nonlinear functions in the classification model. In fact, in recent works, several authors have employed the RNNs with LSTM units for the automatic classification of other biomedical signals (e.g. electrocardiographic signal), obtaining high performances [45]–[47].

The question related to N1 stage detection emerges in the context of neurocognitive performance (NCP) evaluation; in fact, the first step in testing the capabilities to track subtle NCP changes should be the analysis of the hypnagogic state of the subject, i.e. the transition between wakefulness and the first stage of sleep (i.e. stage N1). In this context, a novel classification system for the hypnagogic state becomes crucial [6]. In a future work, the proposed approach based on RNNs could be employed for the detection of sleep onset period (SOP) which is the time that elapses between the awake state and the drowsiness and becomes relevant for studies about alertness and reaction times. Therefore, the classification results obtained for stages W and N1 should be improved in order to define an automated system for the assessment of NCP.

## References

- [1] A. Gevins, “Non-invasive human neurocognitive performance capability testing method and system,” *U.S. Patent*, no. 5,295,491, 1994.
- [2] C. F. Reynolds and R. O’Hara, “DSM-5 Sleep-Wake Disorders Classification: Overview for Use in Clinical Practice,” *Am. J. Psychiatry*, vol. 170, no. 10, pp. 1099–1101, 2013, doi: 10.1176/appi.ajp.2013.13010058.
- [3] J. S. Higgins *et al.*, “Asleep at the Wheel-The Road to Addressing Drowsy Driving,” *Sleep*, vol. 40, no. 2, Feb. 2017, doi: 10.1093/sleep/zsx001.
- [4] N. Goel, H. Rao, J. S. Durmer, and D. F. Dinges, “Neurocognitive Consequences of Sleep Deprivation,” *Semin. Neurol.*, vol. 29, no. 4, pp. 320–339, Sep. 2009, doi: 10.1055/s-0029-1237117.
- [5] A. Garcés Correa, L. Orosco, and E. Laciari, “Automatic detection of drowsiness in EEG records based on multimodal analysis,” *Med. Eng. Phys.*, vol. 36, no. 2, pp. 244–249, 2014, doi: <https://doi.org/10.1016/j.medengphy.2013.07.011>.
- [6] H. Tanaka, M. Hayashi, and T. Hori, “Topographical characteristics and principal component structure of the hypnagogic EEG,” *Sleep*, vol. 20, no. 7, pp. 523–534, Jul. 1997, doi: 10.1093/sleep/20.7.523.
- [7] C. Iber, S. Ancoli-Israel, A. L. Chesson, and S. Quan, “The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications,” *Westchester, Am. Acad. Sleep Med.*, Jan. 2007.
- [8] 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, doi: 10.1046/j.1440-1819.2001.00810.x.
- [9] A. A. Putilov and O. G. Donskaya, “Association of an individual’s ability to overcome desire to fall asleep with a higher anterior-posterior gradient in electroencephalographic indexes of sleep pressure,” *Int. J. Psychophysiol.*, vol. 113, pp. 23–28, 2017, doi: <https://doi.org/10.1016/j.ijpsycho.2017.01.002>.
- [10] R. Boostani, F. Karimzadeh, and M. Nami, “A comparative review on sleep stage classification methods in patients and healthy individuals,” *Comput. Methods Programs Biomed.*, vol. 140, pp. 77–91, 2017, doi: <https://doi.org/10.1016/j.cmpb.2016.12.004>.
- [11] R. K. Malhotra and A. Y. Avidan, “Chapter 3 - Sleep Stages and Scoring Technique,” S. Chokroverty and R. J. B. T.-A. of S. M. (Second E. Thomas, Eds. St. Louis: W.B. Saunders, 2014, pp. 77–99.
- [12] T. Hastie, R. Tibshirani, and J. Friedman, “Model Assessment and Selection BT - The Elements of Statistical Learning: Data Mining, Inference, and Prediction,” T. Hastie, R. Tibshirani, and J. Friedman, Eds. New York, NY: Springer New York, 2009, pp. 219–259.
- [13] R. D. Ogilvie, “The process of falling asleep,” *Sleep Med. Rev.*, vol. 5, no. 3, pp. 247–270, Jun. 2001, doi: 10.1053/smr.2001.0145.
- [14] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. C. Kamphuisen, and J. J. L. Obery, “Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG,” *IEEE Trans. Biomed. Eng.*, vol. 47, no. 9, pp. 1185–1194, 2000, doi: 10.1109/10.867928.

- [15] A. L. Goldberger *et al.*, “PhysioBank, PhysioToolkit, and PhysioNet,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000, doi: 10.1161/01.CIR.101.23.e215.
- [16] B. Hjorth, “EEG analysis based on time domain properties,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, 1970, doi: [https://doi.org/10.1016/0013-4694\(70\)90143-4](https://doi.org/10.1016/0013-4694(70)90143-4).
- [17] P. Welch, “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,” *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, 1967, doi: 10.1109/TAU.1967.1161901.
- [18] 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, 2005, doi: <https://doi.org/10.1016/j.cmpb.2005.06.012>.
- [19] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley-Interscience, 2006.
- [20] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, 2005, doi: 10.1109/TPAMI.2005.159.
- [21] I. Jolliffe, “Principal Component Analysis BT - International Encyclopedia of Statistical Science,” M. Lovric, Ed. Springer Berlin Heidelberg, 2011, pp. 1094–1096.
- [22] L. Deng and D. Yu, “Deep learning: methods and applications,” *Found. trends signal Process.*, vol. 7, no. 3–4, pp. 197–387, 2014.
- [23] Y. Goldberg, “Neural Network Methods for Natural Language Processing,” *Synth. Lect. Hum. Lang. Technol.*, vol. 10, no. 1, pp. 1–309, Apr. 2017, doi: 10.2200/S00762ED1V01Y201703HLT037.
- [24] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
- [25] N. F. Güler, E. D. Übeyli, and İ. Güler, “Recurrent neural networks employing Lyapunov exponents for EEG signals classification,” *Expert Syst. Appl.*, vol. 29, no. 3, pp. 506–514, 2005, doi: <https://doi.org/10.1016/j.eswa.2005.04.011>.
- [26] Y.-L. Hsu, Y.-T. Yang, J.-S. Wang, and C.-Y. Hsu, “Automatic sleep stage recurrent neural classifier using energy features of EEG signals,” *Neurocomputing*, vol. 104, pp. 105–114, 2013, doi: <https://doi.org/10.1016/j.neucom.2012.11.003>.
- [27] A. Graves, “Generating Sequences With Recurrent Neural Networks,” *eprint arXiv:1308.0850*, Aug. 2013, Accessed: Jul. 25, 2018. [Online]. Available: <http://arxiv.org/abs/1308.0850>.
- [28] I. A. Basheer and M. Hajmeer, “Artificial neural networks: fundamentals, computing, design, and application,” *J. Microbiol. Methods*, vol. 43, no. 1, pp. 3–31, 2000, doi: [https://doi.org/10.1016/S0167-7012\(00\)00201-3](https://doi.org/10.1016/S0167-7012(00)00201-3).
- [29] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997, doi: 10.1109/78.650093.
- [30] S. Hochreiter, “The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions,” *Int. J. Uncertainty, Fuzziness Knowledge-Based Syst.*, vol. 06, no. 02, pp. 107–116, Apr. 1998, doi:

- 10.1142/S0218488598000094.
- [31] S. Hochreiter and J. Schmidhuber, “Long short-term memory.,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–80, Nov. 1997, Accessed: Jul. 18, 2018. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9377276>.
  - [32] M.-T. Lo, P. H. Tsai, P.-F. Lin, C. Lin, and Y.-L. Hsin, “The Nonlinear and nonstationary Properties in EEG Signals: Probing the Complex Fluctuations by Hilbert-Huang Transform.,” *Adv. Adapt. Data Anal.*, vol. 1, pp. 461–482, Jul. 2009, doi: 10.1142/S1793536909000199.
  - [33] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv Prepr. arXiv1412.6980*, 2014, Accessed: Jul. 26, 2018. [Online]. Available: <http://arxiv.org/abs/1412.6980>.
  - [34] A. R. Hassan, S. K. Bashar, and M. I. H. Bhuiyan, “On the classification of sleep states by means of statistical and spectral features from single channel Electroencephalogram,” in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2015, pp. 2238–2243, doi: 10.1109/ICACCI.2015.7275950.
  - [35] A. R. Hassan and M. I. H. Bhuiyan, “Computer-aided sleep staging using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and bootstrap aggregating,” *Biomed. Signal Process. Control*, vol. 24, pp. 1–10, 2016, doi: <https://doi.org/10.1016/j.bspc.2015.09.002>.
  - [36] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, “Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier,” *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 10–19, 2012, doi: <https://doi.org/10.1016/j.cmpb.2011.11.005>.
  - [37] S. Seifpour, H. Niknazar, M. Mikaeili, and A. M. Nasrabadi, “A new automatic sleep staging system based on statistical behavior of local extrema using single channel EEG signal,” *Expert Syst. Appl.*, vol. 104, pp. 277–293, 2018, doi: <https://doi.org/10.1016/j.eswa.2018.03.020>.
  - [38] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J.-F. Payen, “A convolutional neural network for sleep stage scoring from raw single-channel EEG,” *Biomed. Signal Process. Control*, vol. 42, pp. 107–114, 2018, doi: <https://doi.org/10.1016/j.bspc.2017.12.001>.
  - [39] L. Wei, Y. Lin, J. Wang, and Y. Ma, “Time-Frequency Convolutional Neural Network for Automatic Sleep Stage Classification Based on Single-Channel EEG,” in *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2017, pp. 88–95, doi: 10.1109/ICTAI.2017.00025.
  - [40] M. Sharma, D. Goyal, P. V Achuth, and U. R. Acharya, “An accurate sleep stages classification system using a new class of optimally time-frequency localized three-band wavelet filter bank,” *Comput. Biol. Med.*, vol. 98, pp. 58–75, 2018, doi: <https://doi.org/10.1016/j.compbimed.2018.04.025>.
  - [41] N. Michielli, U. R. Acharya, and F. Molinari, “Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals,” *Comput. Biol. Med.*, vol. 106, pp. 71–81, 2019, doi: 10.1016/j.compbimed.2019.01.013.
  - [42] H. R. Colten and B. M. Altevogt, *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. National Academies Press (US), 2006.
  - [43] S. Charbonnier, L. Zoubek, S. Lesecq, and F. Chapotot, “Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging,” *Comput. Biol. Med.*, vol. 41, no. 6, pp. 380–389, 2011, doi:

- <https://doi.org/10.1016/j.combiomed.2011.04.001>.
- [44] U. R. Acharya *et al.*, “Nonlinear Dynamics Measures for Automated EEG-Based Sleep Stage Detection,” *Eur. Neurol.*, vol. 74, no. 5–6, pp. 268–287, 2015, doi: 10.1159/000441975.
- [45] Ö. Yildirim, “A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification,” *Comput. Biol. Med.*, vol. 96, pp. 189–202, 2018, doi: <https://doi.org/10.1016/j.combiomed.2018.03.016>.
- [46] J. H. Tan *et al.*, “Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals,” *Comput. Biol. Med.*, vol. 94, pp. 19–26, 2018, doi: <https://doi.org/10.1016/j.combiomed.2017.12.023>.
- [47] O. Faust, A. Shenfield, M. Kareem, T. R. San, H. Fujita, and U. R. Acharya, “Automated detection of atrial fibrillation using long short-term memory network with RR interval signals,” *Comput. Biol. Med.*, vol. 102, pp. 327–335, 2018, doi: <https://doi.org/10.1016/j.combiomed.2018.07.001>.

## Chapter 3

# Analysis of Signal Decomposition methods and timestep Feature Selection

### 3.1 Introduction

As discussed in Chapter 2, recurrent neural networks (RNNs) can be considered a powerful approach for the automatic classification of biomedical signals, in particular non-stationary and nonlinear time series, because of their property of handling temporal feature vectors and employing nonlinear activation functions in the model architecture. In addition, the main crucial point is the link between the extracted handcrafted features and the clinical problem. Independently on the used classifier, there could be a set of features that are more significant. Such features contain the “signature” of the physiological system and, if present, of the pathology. On the other hand, in the convolutional neural networks (CNNs), which are the most powerful deep learning networks suitable for image recognition and classification, the problem of feature extraction is bypassed. In fact, the fundamental building block of CNN is the convolution operation: several filters with different filter sizes are employed to detect feature maps from the input image and the filter coefficients are treated as parameters which are updated using backward propagation, as explained in Chapter 2. The features are learned by the network directly from the data and the big advantage is the parameter sharing and the sparsity of connections [1]. CNNs are employed for image segmentation and classification and in the context of sleep staging, time-frequency transforms can be treated as bidimensional images for CNN-based applications, but there is not an explainable relation between learning patterns detected by the network during the training process and the subject’s conditions [2]. In addition, the number of hyperparameters to set is higher with respect to RNN with a consequence of a higher

computational time. Moreover, the advantage of RNNs is to handle directly temporal sequences such as time series, as discussed in Chapter 2.

In this chapter, a novel timestep feature selection approach is proposed to select only significant and relevant variables for classification purposes. Several features in time, frequency and time-frequency domain are fed to RNN architectures for the problem of sleep scoring, already described in detail in Chapter 1 and 2. This approach should be seen as a general workflow for the automatic classification of non-stationary biomedical signals.

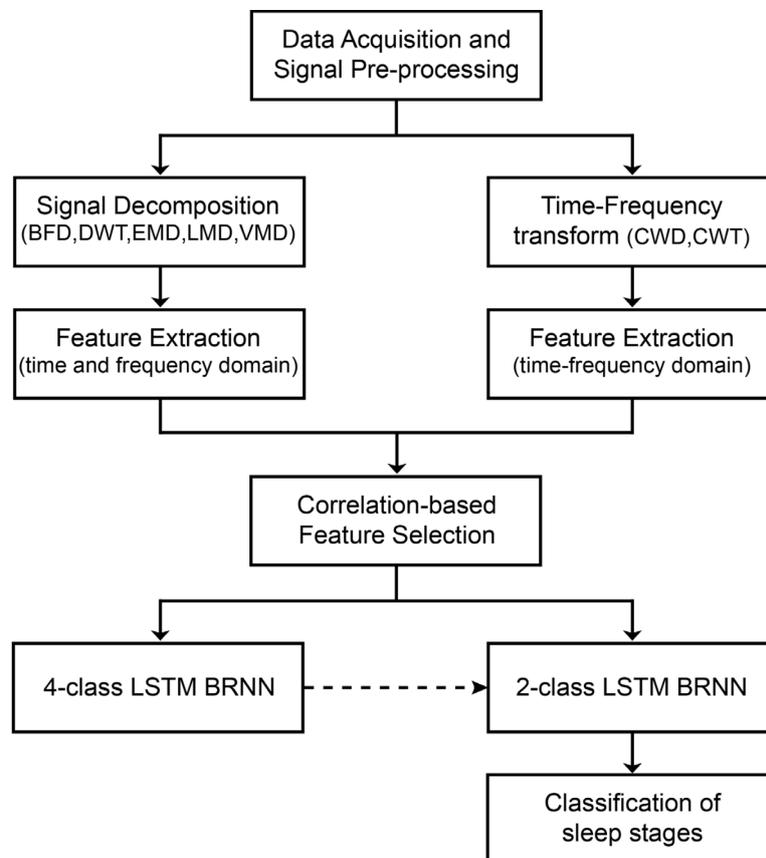
The sleep scoring process is briefly recalled in the following. Sleep scoring is a complex process performed by sleep experts who, according to the American Academy of Sleep Medicine (AASM) criteria [3], [4], assign a score (between W, N1, N2, N3 or REM) to a 30 s epoch, after having visually examined the electroencephalographic (EEG) signals of a human subject. The time employed by the expert to score a full night is about 1-3 hours. In addition, the expert may sometimes request a second opinion for scoring agreement or to diagnose a sleep-related disease [5]. Hence, sleep scoring is time-consuming, often tedious and suffers from intra- and inter-operator variability. For this reason, automatic sleep stage classification (ASSC) systems with low computational times are intended to support sleep experts in this difficult and complex task. The main structure of these automated systems consists of four modules for the analysis and classification of EEG signals: *i*) data acquisition and signal pre-processing, *ii*) feature extraction, *iii*) feature selection and *iv*) classification [6]. From the study of the sequence and duration of the various sleep stages throughout the night, the sleep expert obtains the hypnogram, which is a graphical representation of the sleep stages as function of time and allows to obtain quantitative data related to subject's sleep [4]. A critical issue is that, during sleep, there are several epochs scored as wakefulness in addition to the initial and final part of the hypnogram; thus, the subject wakes up during the night and the number of nocturnal wake epochs is very important for the diagnosis of sleep-related disorders [7].

The EEG signal can be employed for the monitoring of brain activity since it can be accessible in a non-invasive way and due to its high temporal resolution. The information extracted in time domain from the EEG signals is not as significant as that contained in frequency domain since EEG has a limited frequency bandwidth in the range of 0.3-80 Hz. From a clinical point of view and for the sleep analysis, the interesting band is between 0.5 and 45 Hz and is split into six psychological sub-bands: delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma or the sleep spindle sub-band (12-16 Hz), beta (16-30 Hz) and gamma sub-band (> 30 Hz). The lower the frequency range of a sub-band, the higher the amplitude of the EEG signal, and the more mental activity associated to that sub-band is reduced. Therefore delta and theta sub-bands are related to sleep, drowsiness or pathological conditions, instead alpha sub-band occurs when the subject is relaxed or starts to fall asleep and finally beta and gamma sub-bands are recorded during concentration or attention related processes [8]. Nevertheless, these sub-bands are not to be considered physiologically separate and well distinct and the frequency range may

vary between subjects, hence, in this work, several signal decomposition techniques were analyzed and employed in order to extract significant features from sub-signals with different frequency content. In addition, two time-frequency distributions were used for the analysis of single-channel EEG signals. The segmentation of time-frequency transforms has been performed based on the frequency-bands of the rhythms of EEG and for each sub-band were extracted several features in time-frequency domain. A novel correlation-based timestep feature selection method was proposed and finally the selected features were fed to bidirectional RNN model. In the discussion section of this chapter, future directions about this novel approach are discussed.

### 3.2 Materials and Methods

The proposed workflow is reported in Figure 3.1. After the data acquisition and pre-processing step, several decomposition techniques were applied to single-channel EEG signals and features in time and frequency domain were computed for each sub-signal; on the other hand, two time-frequency transforms were employed and features in time-frequency domain were extracted after frequency segmentation of physiological EEG sub-bands. Finally, a novel correlation-based feature selection was performed, and the selected features were fed to a cascaded architecture based on bidirectional RNNs (BRNNs) with long short-term memory (LSTM) units for the automatic scoring of sleep.



**Fig. 3.1** Workflow of the proposed automatic sleep stage classification system.

### 3.2.1 Data acquisition and pre-processing

The publicly available Sleep-EDF database expanded v1.0.0 published in 2013 and updated in 2018 was used for this study [9], [10]. This database is a collection of polysomnographic (PSG) signals with corresponding hypnograms annotated by sleep experts. Trained sleep technicians manually produced hypnograms with annotations of sleep stages referred to 30 s epochs, according to the guidelines introduced by Rechtschaffen and Kales (R&K) [11]. Sleep signals in European data format (EDF) were whole-night PSG recordings including EEG (from the Fpz-Cz and Pz-Oz electrode locations according to the international 10-20 system [12]), electrooculogram (EOG), oro-nasal respiration, submental chin electromyogram (EMG) and rectal body temperature. The EEG signals were sampled with a frequency of 100 Hz. The *SC* (Sleep Cassette) files (*SC4ssnE0-PSG.edf*) contained the PSG signals of subject *ss* ( $00 \leq ss \leq 19$ ) for night *n* ( $n=1$  or  $2$ ). In this study thirty-nine recordings of 20 healthy young subjects (age:  $28.6 \pm 3$  years, range: 25-34 years, 10 males and 10 females) were used. In addition, stage 3 and 4 were merged into stage N3, as AASM recommended. The channel chosen for this study was the Fpz-Cz for the same reason discussed in Chapter 2 and for performance comparison with other state-of-the-art methods. This extended dataset contained more epochs with respect to the database described in Chapter 2, and there was also a higher inter-subject variability for the non-stationary characteristics of the EEG signal. For each subject's recording, in addition to nocturnal W epochs which occurred during the night, only the first 50 epochs (i.e. 25 minutes) of stage W were retained in the first part of the hypnogram, in order to assess the performances of the classifier in W epochs related to the transition between wakefulness and the first stage of sleep (i.e. stage N1) [13]. In Table 3.1, the number of epochs of 30 s fixed duration and the relative sleep percentage for each sleep stage, is shown.

**Table. 3.1** Dataset composition in various sleep stages.

Sleep stage	W	N1	N2	N3	REM	Total
# Epochs	6507	2804	17799	5703	7717	40530
Sleep percentage (%)	16	7	44	14	19	100

After signal acquisition, a pre-processing step was employed to reduce the EEG signal frequency bandwidth in the range of 0.5-45 Hz, which, from a clinical point of view, is the most significant. In addition, a time-segmentation with an overlap of 50% was employed to split each EEG epoch into 11 overlapped segments of 5 s duration. The time duration of the EEG signals from which to extract information was defined by searching a window where the EEG can be at least considered a wide-sense stationary process and, at the same time, satisfying the frequency resolution (i.e. 200 mHz) according to Heisenberg's uncertainty principle. The number of timesteps for the BRNN was set equal to 11, accordingly.

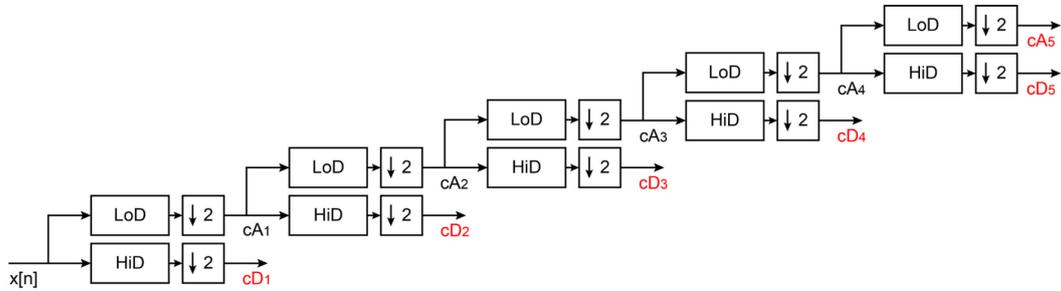
### 3.2.2 Signal decomposition

In this study, five different signal decomposition techniques were employed: Butterworth filter decomposition (BFD), discrete wavelet transform (DWT), empirical mode decomposition (EMD), local mean decomposition (LMD) and variational mode decomposition (VMD).

The first decomposition strategy was based on the implementation of six infinite impulse response (IIR) 8<sup>th</sup>-order band-pass Butterworth digital filters [14] in order to decompose each 30 s EEG signal into six sub-signals in delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma (12-16 Hz), beta (16-30 Hz) and gamma sub-band (30-45 Hz), respectively. An example of the application of BFD to EEG signal is reported in Figure 3.2 (a).

The second decomposition strategy was the discrete wavelet transform (DWT). DWT is a powerful multiresolution strategy which employs filter banks for the decomposition and subsequent reconstruction of a non-stationary signal. The decomposition scheme consists in the implementation of low-pass and high-pass filters and consequently the signal is simultaneously passed through these filters in order to obtain the approximation and detail coefficients of the wavelet decomposition [15], [16]. These filters are finite impulse response (FIR) filters with the property of quadrature mirroring: the impulse response of the high-pass filter is obtained by reversing and alternating the signs of the low-pass filter impulse response. The decomposition and reconstruction filters are associated with several wavelet (for high-pass) and scaling (for low-pass) functions and are obtained using a lifting scheme implementation [17]. Each type of wavelet function (e.g. the most commonly used Daubechies wavelets) can be defined with an order (e.g. second, fourth, tenth order Daubechies wavelet). A function with order  $N$  has the first  $N$  moments equal to zero (i.e. vanishing moments) and a sequence length of the corresponding filter impulse response equal to  $2N$  [16]. In general, a wavelet with a greater number of vanishing moments is a more complex function and has the ability to represent a complex signal more accurately (it is capable of representing a polynomial signal of degree equal or greater than  $N$ ). The disadvantage is that a wavelet with a higher number of vanishing moments has a higher temporal support, therefore a worse temporal localization. In this study, Daubechies wavelet and scaling functions were employed since were commonly used for the analysis of EEG signals [18]–[20] and an order equal to ten (i.e. “db10”) was selected to find a compromise between vanishing moments and compact support. Starting from the original signal, two set of coefficients were evaluated: the approximation coefficients (cA) were computed by convolving the signal with the impulse response of the low-pass decomposition (LoD) filter, while the detail coefficients (cD) were computed by convolving the signal with the impulse response of the high-pass decomposition (HiD) filter. During each filtering process, the coefficient sequence must be dyadic decimated due to the output size of each convolution operation without increasing the sampling rate. The process can be recursively repeated until the desired number of decomposition levels is reached. In this study

a fifth level decomposition scheme was adopted, in order to separate delta, theta, alpha, sigma, beta and gamma sub-bands. In Figure 3.3, the fifth level filter bank structure is reported for DWT. After the decomposition,  $cA_5, cD_5, cD_4, cD_3, cD_2, cD_1$  were retained for their frequency content in delta, theta, alpha, sigma, beta, and gamma sub-band, respectively. An example of the application of DWT to EEG signal is reported in Figure 3.2 (b).



**Fig. 3.3** Wavelet decomposition scheme: approximation (cA) and detail coefficients (cD) using a 5<sup>th</sup> level decomposition. The coefficients used for the EEG decomposition are highlighted in red.

The third decomposition strategy was the empirical mode decomposition (EMD). The goal of EMD is to decompose an input signal  $x(t)$  into a finite number ( $N$ ) of sub-signals called *intrinsic mode functions* (IMFs) plus a residual term  $r_N(t)$ . Subsequently, the signal can be reconstructed as:

$$x(t) = \sum_{i=1}^N IMF_i(t) + r_N(t) \quad (\text{Eq. 3.1})$$

IMFs are signal components based on the local temporal features of the input signal, which must satisfy the following properties: *i*) the number of local maxima and minima and the number of zero crossings must be equal or must vary no more than one; *ii*) the average value of the lower and upper envelope must be locally equal to zero. A single IMF can be considered a non-stationary signal since its amplitude and frequency can vary as function of time [21]. The decomposition procedure in EMD is based on the sifting process: there is an outer loop which iterates until either a fixed number of IMFs is obtained, or the residual becomes a monotonic function (i.e. no local maxima or minima can be defined for the computation of the upper and lower envelope) and an inner loop which iterates until a stopping criterion based on the difference between two consecutive sifting components is reached. The stopping criterion is necessary to create a true IMF and at the same time to avoid generating constant amplitude components [21]. The first IMFs, obtained with the sifting process, capture the high-frequency waves of the signal, while the last IMFs are referred to low-frequency components. The final residual is typically the signal trend. In this study, five IMFs were sufficient to decompose the EEG signals, since the residual term assumed very small amplitude values. In the EMD, there was no relation between IMFs and physiological EEG sub-bands (i.e. delta, theta, alpha, sigma, beta, and gamma sub-band), since the power spectrum of an IMF could include more than one sub-band or could be interposed between two sub-bands. An example of the application of EMD to EEG signal is reported in Figure 3.2 (c).

The fourth decomposition technique was the local mean decomposition (LMD). LMD is a decomposition method employed for the analysis of amplitude and frequency modulated biomedical signals. This strategy consists in the derivation of a pure frequency modulated (FM) signal and the corresponding amplitude modulated (AM) envelope signal in order to obtain the first *product function* (PF). The process is repeated until a fixed number ( $N$ ) of mono-component PFs is reached and consequently the original signal can be expressed as:

$$x(t) = \sum_{i=1}^N PF_i(t) + u_N(t) \quad (\text{Eq. 3.2})$$

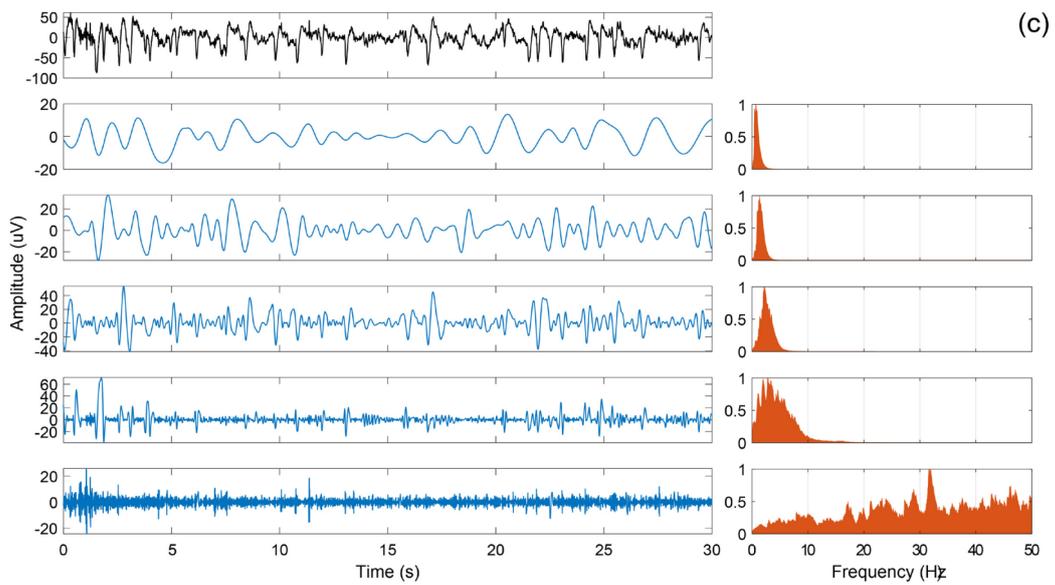
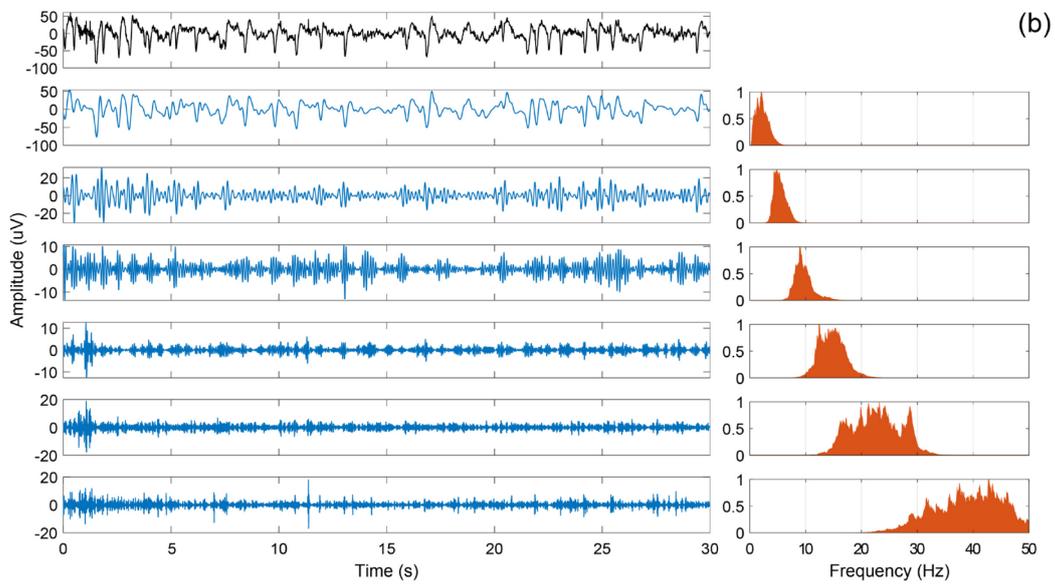
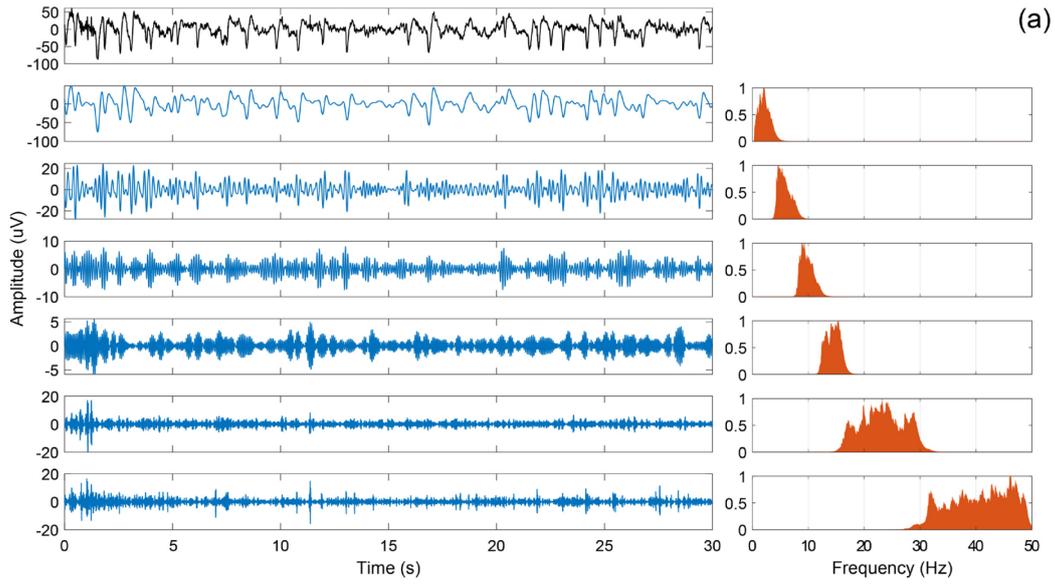
$$PF_i(t) = AM_i(t) \cdot FM_i(t) \quad (\text{Eq. 3.3})$$

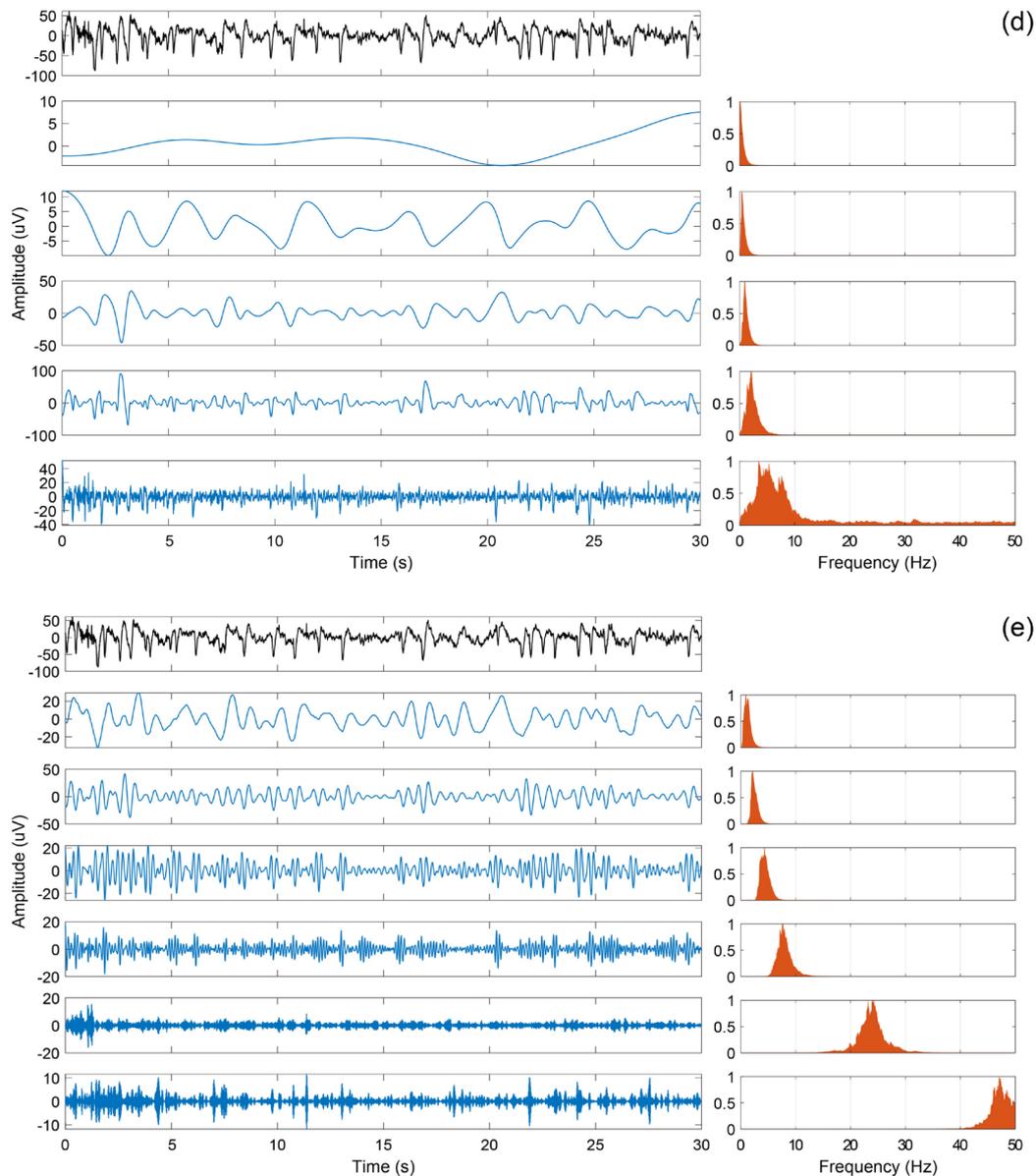
where  $u_N(t)$  represents the residual signal [22]. The implementation strategy is similar to the sifting algorithm used in EMD: there is an outer loop which iterates until a fixed number of PFs is reached or when the residual presents no oscillations, and an inner loop which iterates until a pure FM component (i.e. until a flat envelope is satisfied) is extracted from the input signal. A recent improved version of the LMD was proposed and implemented by Liu et al. [23]. In this study, a number of five PFs was sufficient to decompose the EEG signals, since the residual term assumed very small amplitude values. In the LMD, just as in EMD, there was no relation between PFs and physiological EEG sub-bands (i.e. delta, theta, alpha, sigma, beta, and gamma sub-band), since the power spectrum of a PF could include more than one sub-band or could be interposed between two sub-bands. An example of the application of LMD to EEG signal is reported in Figure 3.2 (d).

The last decomposition method analyzed in this chapter is the variational mode decomposition (VMD). VMD is a novel adaptive method proposed by Dragomiretskiy and Zosso [24] with the aim of decomposing an input signal into sub-signals compactly centered around their characteristic frequency, named *principal modes*. The authors proposed an iterative optimization algorithm to solve the following minimization problem:

$$\min_{u_k, \omega_k} \left\{ \sum_{k=1}^K \left\| \frac{d}{dt} \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] \cdot e^{-j\omega_k t} \right\|_2^2 \right\} \quad (\text{Eq. 3.4})$$

subjected to the following constraint:  $x(t) = \sum_{k=1}^K u_k(t)$ , where  $x(t)$  is the input signal,  $u_k$  is the  $k$ -th mode and  $K$  is the number of all modes involved in the decomposition. In Eq. 3.4, the objective is to minimize the sum of the bandwidths of each mode; in fact, firstly the analytic form (that has a unilateral spectrum with no negative frequency components) of each mode is computed using the Hilbert transform, next the exponential factor is useful to shift the Fourier transform to center pulsation  $\omega_k$  and finally the squared norm of the time derivative estimates the mode bandwidth. The convergence of this method was proved, and the optimal principal modes and the corresponding central frequencies were obtained. In this study, a number of six modes was sufficient to decompose the EEG signals, since the residual term assumed very small amplitude values. An example of the application of VMD to EEG signal is reported in Figure 3.2 (e).





**Fig. 3.2** Signal decomposition techniques: the original EEG signal (in black), the decomposed sub-signals (in blue) and the corresponding normalized power spectra (in red) using Butterworth filter decomposition (a), discrete wavelet transform (b), empirical mode decomposition (c), local mean decomposition (d) and variational mode decomposition (e).

Subsequently, for each sub-signal obtained by using the previously described signal separation methods, 12 features were extracted in time domain and 7 features in frequency domain. The features were extracted from each 5 s time segment, repeating the procedure for 11 windows (50% of overlap) of the 30 s original input EEG signal and the decomposed sub-signals obtained with different techniques (i.e. BFD, DWT, EMD, LMD and VMD). The time-domain features were the following: average rectified value (ARV), min and max amplitude, root mean square (RMS) value, median value, standard deviation, skewness, kurtosis, energy ratio (i.e. the energy of each sub-signal normalized to the total energy of the input signal), number of zero crossings, Hjorth mobility (HM) and Hjorth complexity (HC). The frequency-domain features were the following: power ratio (i.e. the power of each

sub-signal normalized to the total power of the input signal), mean frequency (MNF), dominant or center frequency (i.e. the frequency associated to the highest power), spectral edge frequency (SEF) at 50% (i.e. the lowest frequency below which 50% of the total power is present), SEF at 95% (i.e. the lowest frequency below which 95% of the total power is present), spectral and Renyi entropy. The mathematical expressions of these parameters are reported in Chapter 2. The total number of features extracted from each signal decomposition method is reported in Table 3.2.

**Table. 3.2** Number of time- and frequency-domain features for each decomposition technique.

<b>Signal Decomposition</b>	<b>BFD</b>	<b>DWT</b>	<b>EMD</b>	<b>LMD</b>	<b>VMD</b>
Time domain	84	84	72	72	84
Frequency domain	49	49	42	42	49
Total	133	133	114	114	133

### 3.2.3 Time-Frequency distributions

In this study, two time-frequency representations (TFRs) were employed for the analysis of EEG signals: Choi-Williams distribution (CWD) and continuous wavelet transform (CWT). Time-frequency distributions are specific tools of spectral analysis useful to treat non-stationary signals since they represent the signal power spectrum instantaneously, i.e. as a function of time. The most frequently used TFR is the short-time Fourier transform (STFT) but it suffers from the problem that the maximum frequency resolution cannot be guaranteed otherwise the time window should be too large and the signal can no longer be considered stationary. On the other hand bilinear time-frequency distributions were introduced with the aim of working at full frequency resolution without breaking the Wiener-Khinchin theorem [25]. The Wigner-Ville distribution (WVD) was the first time-frequency distribution which maintained the time dependence of the autocorrelation function, introducing the concept of instantaneous autocorrelation function of a signal  $x(t)$ , using the following mathematical relation:

$$WVD(t, f) = \int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right) \cdot x^*\left(t - \frac{\tau}{2}\right) \cdot e^{-j2\pi f\tau} d\tau \quad (\text{Eq. 3.5})$$

where  $\tau$  is the time delay and the asterisk denotes the complex conjugate [25]. The problem of WVD is that suffers from the interference or cross terms (i.e. terms related to frequency mixing). The solution to this problem was solved by the Choi-Williams distribution (CWD) which employed a kernel in the ambiguity function domain with the purpose of attenuating the interference terms. The mathematical expression of CWD is reported in the following:

$$CWD(t, f) = \iiint_{-\infty}^{+\infty} x\left(t' + \frac{\tau}{2}\right) x^*\left(t' - \frac{\tau}{2}\right) g(\theta, \tau) e^{-j2\pi\theta(t-t')} e^{-j2\pi f\tau} dt' d\theta d\tau \quad (\text{Eq. 3.6})$$

$$g(\theta, \tau) = e^{-\frac{\theta^2 \tau^2}{\sigma}} \quad (\text{Eq. 3.7})$$

where  $\theta$  is the frequency delay,  $g(\theta, \tau)$  is the Choi-Williams kernel defined in the ambiguity function domain and  $\sigma$  is the kernel width. In this study, the kernel width of CWD was set equal to 1.

The second TFR used in this study was the continuous wavelet transform (CWT). The STFT was found to be inadequate for processing highly non-stationary signals since it is not possible to locate rapidly and slowly evolving phenomena with the same precision. Therefore, in order to obtain an analysis with variable resolution, a set of basic functions called wavelet family was introduced by translation and scaling of a mother wavelet  $\psi(t)$ :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (\text{Eq. 3.8})$$

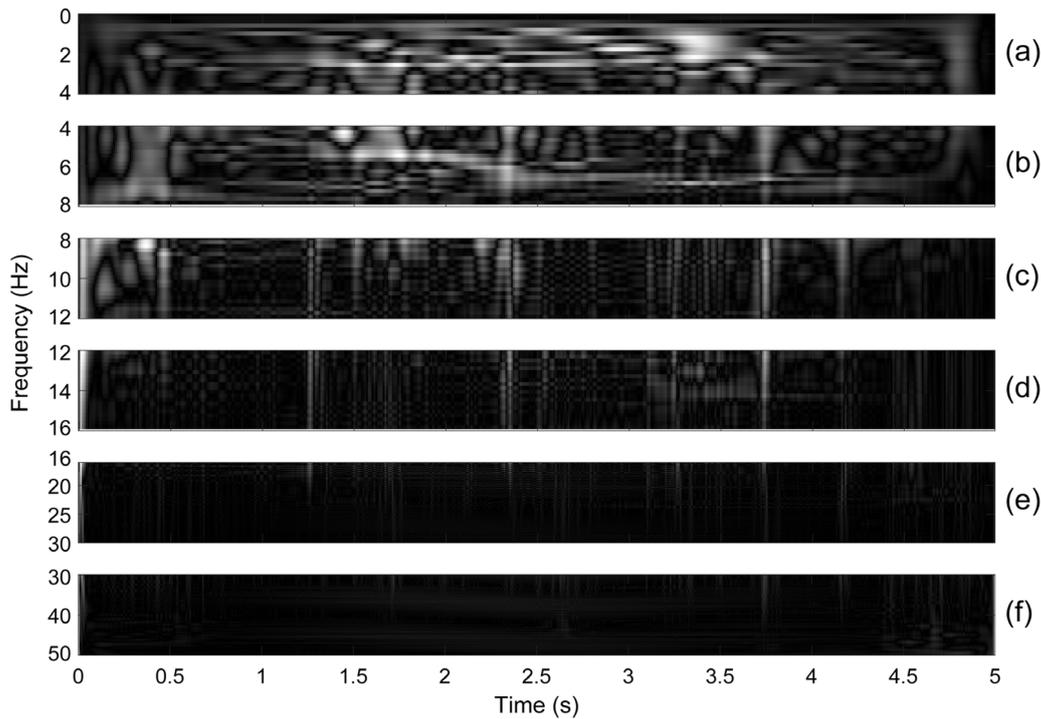
where  $a, b$  are the scale and shift parameters, respectively. The coefficient  $1/\sqrt{a}$  is useful to normalize the different wavelets which must have the same energy [26]. The CWT can be defined as the inner product between the signal  $x(t)$  and the wavelet family as:

$$CWT(a, b) = \int_{-\infty}^{+\infty} x(t) \cdot \psi_{a,b}^*(t) dt \quad (\text{Eq. 3.9})$$

The squared magnitude of the CWT is called scalogram. The scalogram is a time-scale representation (since the parameter  $b$  is expressed in time units) that can be converted in a TFR by remembering that the relation between scale and frequency is related to the central or dominant frequency of the mother wavelet [26]. The wavelet functions present a good frequency localization at low frequencies and, at the same time, a good temporal localization at high frequency [27]. The classical implementation of Eq. 3.9 is computationally expensive, for this reason, in recent years, an efficient and fast algorithm for the CWT computation has been proposed. The definition in Eq. 3.9 can also be treated as the cross-correlation of the signal and the wavelet, and according to the convolution theorem, the cross-correlation can be expressed in frequency domain; consequently the CWT can also be expressed in terms of inverse Fourier transform which can be implemented using the fast Fourier transform (FFT) algorithm. All the mathematical derivations of the fast algorithm for the computation of CWT are reported in the work of Komorowski and Pietraszek [26]. Different mother wavelet can be employed for the analysis, in this study the analytic Morlet wavelet was employed since has obtained the best performances in the processing of EEG signals [28], [29] and other biomedical signals [26].

Subsequently, both CWD and CWT were computed for each EEG epoch; then the TFRs were segmented in order to separate delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma (12-16 Hz), beta (16-30 Hz) and gamma (30-45 Hz) EEG sub-bands in the time-frequency plane. The TFRs were subdivided into six sub-images corresponding to each frequency sub-band and 25 time-frequency domain features

were extracted from each sub-image. An example of the subdivision of TFR is reported in Figure 3.4.



**Fig. 3.4** Time-frequency sub-images corresponding to delta (a), theta (b), alpha (c), sigma (d), beta (e) and gamma (f) EEG rhythms.

In the following, a detailed description of the time-frequency features is reported. Let  $TF[m, n]$  be a generic TFR, where  $1 < m < M$  represents the index of frequency vector and  $1 < n < N$  represents the index of time vector. Firstly, the relative power was computed as the ratio between the power (averaged over time) in each sub-band and the total power of the entire TFR. Then, the following four statistical features for each sub-band were computed as the average over frequency: RMS value, standard deviation, skewness, and kurtosis. The marginal mean time (MNT) and mean frequency (MNF) can be defined as:

$$MNT = \frac{\sum_{n=1}^N (\sum_{m=1}^M TF[m, n]) \cdot t[n]}{\sum_{m=1}^M \sum_{n=1}^N TF[m, n]} \quad (\text{Eq. 3.10})$$

$$MNF = \frac{\sum_{m=1}^M (\sum_{n=1}^N TF[m, n]) \cdot f[m]}{\sum_{m=1}^M \sum_{n=1}^N TF[m, n]} \quad (\text{Eq. 3.11})$$

The time-frequency fluxes can be defined along three directions:

$$TF_{i,j}^{flux} = \sum_{m=1}^{M-i} \sum_{n=1}^{N-j} TF[m+i, n+j] - TF[m, n] \quad (\text{Eq. 3.12})$$

where  $i = 0, j = 1$  denotes the flux along the time direction,  $i = 1, j = 0$  denotes the flux along the frequency direction and  $i = 1, j = 1$  is the flux along the diagonal axis. The time-frequency flatness is defined as the ratio between the geometric mean and the arithmetic mean of the TFR (values lower than 10% of the maximum

were excluded from the computation) [30]. The time-frequency energy concentration is defined as:

$$TF^{conc} = \left( \sum_{m=1}^M \sum_{n=1}^N \sqrt{TF[m, n]} \right)^2 \quad (\text{Eq. 3.13})$$

The Shannon entropy (SE) and Renyi entropy (RE) can be computed as:

$$SE = - \sum_{m=1}^M \sum_{n=1}^N \frac{TF[m, n]}{\sum_{m=1}^M \sum_{n=1}^N TF[m, n]} \cdot \log_2 \left( \frac{TF[m, n]}{\sum_{m=1}^M \sum_{n=1}^N TF[m, n]} \right) \quad (\text{Eq. 3.14})$$

$$RE = - \frac{1}{2} \log_2 \left( \sum_{m=1}^M \sum_{n=1}^N \left( \frac{TF[m, n]}{\sum_{m=1}^M \sum_{n=1}^N TF[m, n]} \right)^3 \right) \quad (\text{Eq. 3.15})$$

In the computation of time-frequency entropies, the zero values were excluded from the analysis. Next, the following four instantaneous frequency (IF) related statistical features were computed: mean, variance, skewness, and kurtosis. The mathematical definition is reported here [30]. Finally, each TFR was considered as an image and histogram-based features and binary image features were extracted. The gray-level histogram of each time-frequency sub-image was computed and the following features were extracted: the Renyi entropy (the entropy order was equal to 2) computed on the histogram distribution [31], maximum count of pixel intensity and spread in the histogram, as reported here [29]. The binary image was obtained by applying Otsu's threshold [32] to the original TFR and the following features were extracted: aspect ratio [29] and area, perimeter and compactness of the largest connected component in the binary image [30].

To summarize, 25 time-frequency domain features were extracted from each time-frequency sub-image related to each of the six EEG sub-bands, i.e. a total of 150 features both for CWD and CWT.

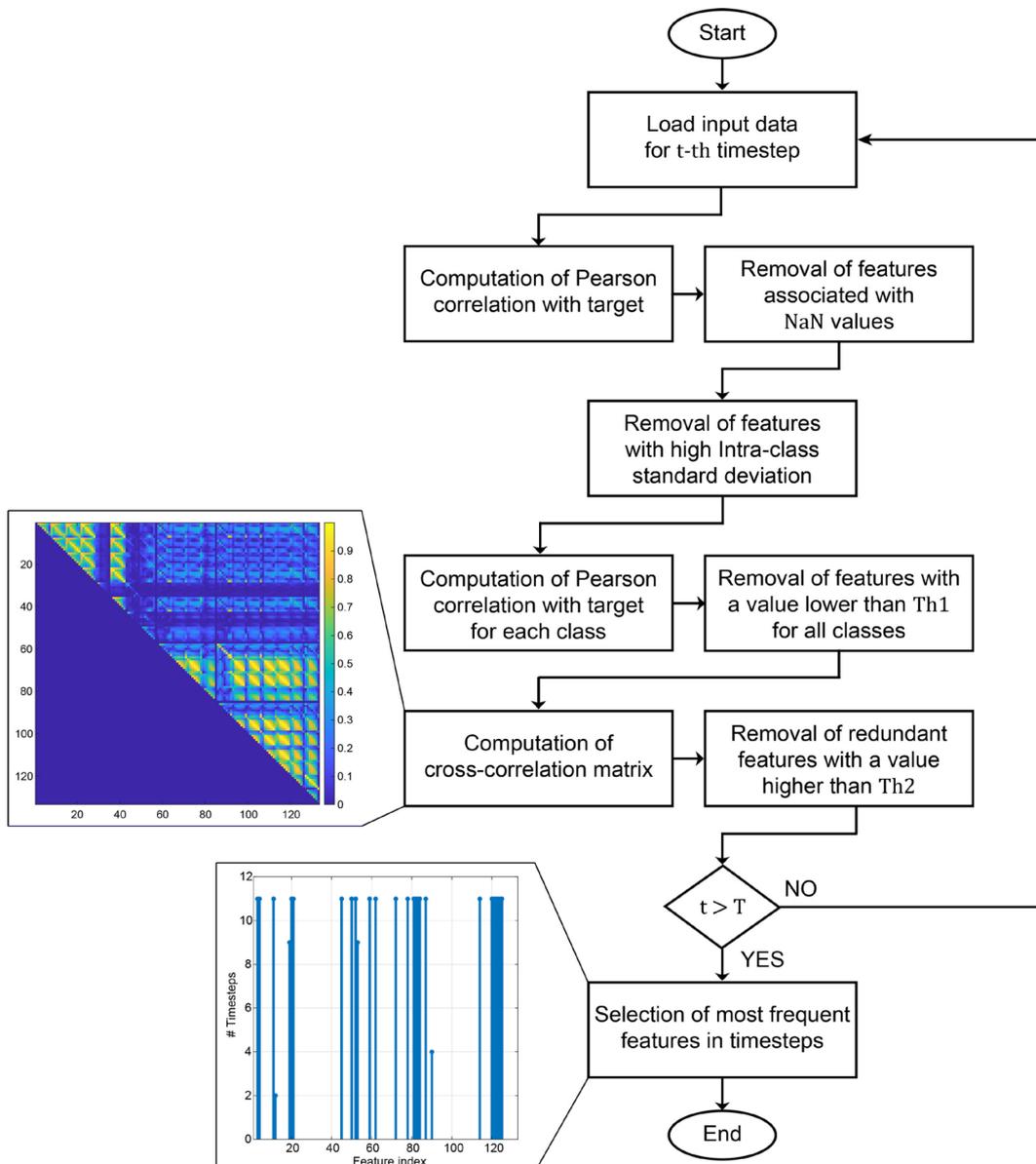
### 3.2.4 Correlation-based timestep feature selection

In this section, a novel supervised correlation-based feature selection is proposed for the selection of features in a multiclass problem. The strategy is based on a timestep implementation, since each set of features was computed for a 5 s EEG epoch and the procedure was repeated for 11 time-segments (with an overlap of 50%) for each 30 s signal. Consequently, the number of timesteps  $T$  for the classification network was set to 11. The flowchart of the proposed feature selection strategy is reported in Figure 3.5.

The proposed strategy was based on the computation of Pearson correlation coefficient defined as the ratio between the covariance of two random variables  $X$  and  $Y$  and the product of their standard deviations:

$$r = \frac{cov(X, Y)}{\sigma_X \cdot \sigma_Y} \quad (\text{Eq. 3.16})$$

This coefficient is bounded between -1 and 1 and an increasing positive or negative value means a higher positive or negative linear correlation respectively, while values close to zero denote the absence of linear correlation.



**Fig. 3.5** Flowchart of the correlation-based timestep feature selection.

The first step was the loading of input feature data matrix for each timestep and the corresponding target vector. The feature data matrix contained on the rows the features and on the columns the observations. The target was a column vector with the corresponding label class for each observation. Next, Pearson correlation with the target was computed and features which presented a NaN (not a number) correlation value were removed since this means that the covariance was zero and at the same time the standard deviation of the feature was zero. In addition, all the features with a high intra-class standard deviation (in this work a threshold value of 5 was set) were removed since they are not representative of a specific class. Then, Pearson correlation with target for each class was computed and features with a

value lower than a threshold  $Th_1$  in all classes were removed since they were considered not relevant. After that, the cross-correlation matrix (i.e. Pearson correlation between features) was obtained and all pairs of features with a value higher than a threshold  $Th_2$  were analyzed only in the upper triangular part since the matrix is symmetric. For each pair of features, that one with a highest correlation with target was retained and the other was removed since it was considered redundant. The values of thresholds  $Th_1$ ,  $Th_2$  are heavily dependent on the application and are determined empirically. Finally, this process was repeated for all timesteps (in this study, the number of timesteps  $T = 11$ ) and only the features selected according to the previous procedure in most of timesteps were retained for the subsequent classification process. In the result section, the thresholds values employed in this study are reported.

### 3.2.5 Bidirectional Recurrent Neural Networks

The classification structure employed in this chapter was the same as that described in Chapter 2. A cascaded architecture based on RNNs with LSTM units, in order to reduce the vanishing gradient problem, was implemented: the first network employed the selected features to classify four stages (i.e. W, N1-REM, N2 and N3), while N1 vs REM epochs were discriminated by the second network. The second network became more specific to classify the two most complex sleep stages, while in the first phase of the process these two classes were merged into the N1-REM class. In this chapter, the only difference with respect to the architecture reported in Chapter 2, is that a sequence-to-sequence architecture based on bidirectional RNNs (BRNNs) was used. The BRNN consists of forward and backward recurrent components; thus it takes information both earlier and later in the sequence and has been proved to provide better results than unidirectional RNN [33], [34]. In the BRNN, there is a forward layer with LSTM units based on the mathematical equations reported in Chapter 2 and a backward layer with LSTM units which employs the same updating equations in the opposite direction. The basic structure of BRNN with LSTM blocks is reported here [33]. Finally the outputs of the forward and backward pass are combined and fed to subsequent fully connected (FC) layers [34]. In the sequence-to-sequence architecture, the network receives a temporal sequence as input and produces a temporal sequence as output; thus the FC layer produces a number of predictions equal to the number of timesteps of the network. The softmax activation function computes the probability of belonging to a given class and the final output was defined as the statistical mode (i.e. the value that most often appears) of the outputs predicted by each block. The cost function to minimize during the training process was the cross-entropy function for mutually exclusive classes and the optimization method was the adaptive moment estimation (ADAM) [35], as described in Chapter 2.

### 3.3 Results

The processing was performed in MATLAB (The MathWorks, Inc., Natick, MA, USA) environment on a workstation with 16 GB of RAM, 2.5 GHz quad-core CPU and 64-bit Windows version. Several BRNN configurations were tested and the performances for each signal decomposition technique were analyzed. The best performances were obtained for VMD and the selected features according to the previously described correlation-based technique are reported in Table 3.3 and Table 3.4. For the 4-class classification problem, 25 features were selected using the following thresholds:  $Th_1 = 0.3, Th_2 = 0.8$  and only the features which occurred in at least 75% of timesteps were retained. For the 2-class classification problem, 33 features were selected using the following thresholds:  $Th_1 = 0.15, Th_2 = 0.9$  and only the features which occurred in at least 65% of timesteps were retained. The thresholds for the second network were less selective since the stages to classify (N1 vs REM) were more complex and the features showed less relevance and a higher correlation between each other. The best architectures for the 4-class and 2-class LSTM BRNN are reported in Table 3.5 and Table 3.6, respectively. A value of 256 for the mini-batch size was set in the training phase. The entire dataset was partitioned into three different sets called training, validation, and test set. The network was trained using only the training set, then the validation set was used to assess the performances and the training was stopped when a maximum number of iterations was reached or when validation error started to increase to avoid data overfitting. The final performance was then evaluated using the test set, which the network had never previously seen. The following quantitative metrics were computed to assess the performance of the classifiers: the percentage of correct classification (PCC) defined as the ratio between the correctly classified data and the total amount of data, and the well-known per-class metrics, i.e. recall or sensitivity (Se), selectivity or specificity (Sp) and accuracy (Acc) for each EEG sleep stage. To test the performance of the BRNN models, we assessed the classification metrics by 10-fold stratified cross-validation strategy; thus in each fold, 80% of data were used in the training set, the 10% of data were used in the validation set and the remaining 10% of data in the test set, randomly. The confusion matrix for each network obtained as the sum of the performances in each fold and the per-class metrics are reported in Table 3.7 and Table 3.8. In Table 3.8, only the epochs correctly classified by the first network are considered for the performance assessment.

**Table 3.3** Number of selected features for the 4-class classification network.

VMD	raw signal	1 <sup>st</sup> mode	2 <sup>nd</sup> mode	3 <sup>rd</sup> mode	4 <sup>th</sup> mode	5 <sup>th</sup> mode	6 <sup>th</sup> mode
Time	1	0	2	1	2	0	0
Frequency	1	0	0	0	0	1	0
TFR	delta	theta	alpha	sigma	beta	gamma	
CWD	1	0	1	2	3	2	
CWT	1	1	1	2	1	2	

(VMD: variational mode decomposition, TFR: time-frequency representation, CWD: Choi-Williams distribution, CWT: continuous wavelet transform)

**Table 3.4** Number of selected features for the 2-class classification network.

VMD	raw signal	1 <sup>st</sup> mode	2 <sup>nd</sup> mode	3 <sup>rd</sup> mode	4 <sup>th</sup> mode	5 <sup>th</sup> mode	6 <sup>th</sup> mode
Time	0	0	0	0	0	1	0
Frequency	0	0	1	1	0	0	0
TFR	delta	theta	alpha	sigma	beta	gamma	
CWD	3	4	2	2	1	1	
CWT	3	3	3	2	2	4	

(VMD: variational mode decomposition, TFR: time-frequency representation, CWD: Choi-Williams distribution, CWT: continuous wavelet transform)

**Table 3.5** Layer structure for the BRNN which identifies 4 classes (W vs N1-REM vs N2 vs N3).

Layer number	Layer type	Properties
Layer 1	Sequence input layer	25 input features
Layer 2	Bidirectional LSTM layer	123 hidden units, sequence-to-sequence architecture
Layer 3	Fully connected layer	4 units
Layer 4	Softmax layer	softmax activation function
Layer 5	Classification output layer	4 classes

**Table 3.6** Layer structure for the BRNN which identifies 2 classes (N1 vs REM).

Layer number	Layer type	Properties
Layer 1	Sequence input layer	33 input features
Layer 2	Bidirectional LSTM layer	146 hidden units, sequence-to-sequence architecture
Layer 3	Fully connected layer	48 units
Layer 4	Fully connected layer	2 units
Layer 5	Softmax layer	softmax activation function (threshold = 0.7)
Layer 6	Classification output layer	2 classes

**Table 3.7** Classification performances for 4-class BRNN.

True	Predicted				Per-class [%]		
	W	N1-REM	N2	N3	Se	Sp	Acc
W	5753	682	50	22	88.41	98.12	96.56
N1-REM	465	9329	718	9	88.67	92.07	91.19
N2	135	1688	15078	898	84.71	94.88	90.42
N3	39	10	395	5259	92.21	97.33	96.61
<b>Percentage of correct classification (PCC) = 87.39%</b>							

**Table 3.8** Classification performances for 2-class BRNN.

True	Predicted		Per-class [%]		
	N1	REM	Se	Sp	Acc
N1	1566	539	74.39	86.41	83.70
REM	982	6242	86.41	74.39	83.70
<b>Percentage of correct classification (PCC) = 83.70%</b>					

The PCC of the first and second BRNN with LSTM units are 87.39% and 83.7% respectively. Finally, considering the cascading connection of the two networks, the performances, in terms of sensitivity, for the five sleep stages are: 88.41% for stage W, 55.85% for stage N1, 84.71% for stage N2, 92.21% for stage N3 and 80.89% for stage REM. The final performances for N1 and REM are computed as the ratio between the number of epochs correctly classified by the 2-class BRNN and the total number of epochs in the dataset (i.e. 2804 and 7717 for stage N1 and REM, respectively). The overall PCC is 83.64%.

### 3.4 Discussion and Conclusion

The proposed approach should be seen as a general workflow for the automatic classification of non-stationary biomedical signals. The common objective is to extract mono-component sub-signals from multicomponent input signal using signal separation or signal decomposition methods. In addition, a novel timestep feature selection approach is proposed to select only significant and relevant variables for classification purposes. Several features in time, frequency and time-frequency domain were fed to BRNN architectures for the problem of sleep scoring. In this context, recent studies have tested their performances using the publicly available Sleep-EDF database expanded [9], [10]. In Table 3.9, the per-class sensitivity and the total PCC of the proposed method and other state-of-the-art approaches are reported for comparison (in bold are highlighted the best performances for each sleep stage).

**Table 3.9** Performance comparison of state-of-the-art methods (results are reported in terms of per-class sensitivity and overall percentage of correct classification in the last column).

Methods		Results [%]					
Authors	Classifier	W	N1	N2	N3	REM	PCC
Zhu et al. [36]	CNN with attention layer	88.37	53.39	82.01	88.41	84.55	82.68
Phan et al. [37]	CNN	75.47	31.86	86.84	86.75	<b>90.56</b>	82.25
Humayun et al. [38]	Residual CNN	96.38	44.23	79.72	76.58	75.97	<b>91.40</b>
Cai et al. [39]	Graph-temporal fused CNN	<b>98.13</b>	33.11	<b>90.57</b>	72.35	79.74	89.10
<b>Proposed method</b>	LSTM BRNN	88.41	<b>55.85</b>	84.71	<b>92.21</b>	80.89	83.64

A brief description of each state-of-the-art method is reported in the following. Zhu et al. [36], analyzed the same subjects used in this chapter and single-channel Fpz-Cz EEG signals were fed to a deep CNN and feature learning was performed by using attention mechanism for sleep scoring. Phan et al. [37], employed the Fpz-Cz EEG channel and the horizontal EOG signals to compute time-frequency STFTs which were used as input images for multi-task CNN. They obtained the best performances in stage REM due to the use of EOG signals. Humayun et al. [38], implemented a 1D CNN-based residual architecture with skip connections which

took as input raw single-channel (Fpz-Cz) EEG signals. They reported the best overall PCC, but the per-class metrics were not satisfactory except for stage W since their dataset was composed of 68% of W epochs. Cai et al. [39], employed a graph representation based on limited penetrable visibility graph (LPVG) and both the signal-channel (Fpz-Cz) raw EEG signals and the graph representations were fed to a dual-input CNN for automatic sleep staging. They reported the best performances in stage W and N2. The high performance in stage W was due to the highest number of W epochs (53% of dataset) used for training.

In this study, a cascaded architecture based on BRNN with LSTM blocks was employed for classification purposes and a large feature set was analyzed. Time- and frequency-domain parameters were computed for different sub-signals obtained through several signal decomposition techniques. In addition time-frequency domain features were extracted from each of six EEG rhythms (i.e. delta, theta, alpha, sigma, beta, and gamma). Finally, a novel correlation-based feature selection strategy, based on a timestep implementation, was proposed for the selection of most relevant features. The cascading connection was composed of a first BRNN which classified four classes (W vs N1-REM vs N2 vs N3) and a second BRNN which performed binary classification (N1 vs REM).

The classification performances between different signal decomposition methods were quite close, since most of features were selected in time-frequency domain as can be shown in Table 3.3 and 3.4, but the highest performance was obtained for VMD. For the 4-class classification problem, features were mostly selected in high frequency sub-bands for the presence of stage W, while in the 2-class problem, this difference between high and low frequency features was not observed due to the complex EEG patterns present in stage N1 and REM according to AASM rules. Table 3.9 shows that the proposed approach obtained the highest value in the detection of N1 and N3 sleep with respect to other state-of-the-art methods and at the same time the performances in the other three sleep stages are superior to 80%. In addition, except in the works which used a huge number of W epochs for training [38], [39], this method obtained the highest classification rate in wake detection if only nocturnal W epochs, which occurred during the night, were considered. These epochs are very different in frequency content with respect to wake epochs in the first and last part of a human subject's sleep cycle, since they are related to the transition between wakefulness and the first stages of sleep [13]. The loss of attention that accompanies the transition from wakefulness to sleep is associated with changes in other physiological parameters in addition to the EEG signal. Among these, there are the so-called slow eye movements (SEMs), whose presence is maximum in the final phases of wakefulness, when alpha rhythms are still present and a reduction in neurocognitive performance (NCP) is observed. In addition, the phenomenon of eye blinking changes with the variation of attention, as it has been observed that, with the decrease of alertness, it becomes less frequent and slower, but with blinking bursts, probably due to the mechanisms for trying to stay awake [40]. This can also be explained with the difference between the sensitivity in stage W reported in Chapter 2 and in this chapter (i.e. 95% vs 88%).

Another issue is related to REM sleep, which is characterized by high eye movement activity recorded by EOG and hence with a single EEG channel, the detection process becomes more complex [41].

The problem of low performances in stage N1 is mainly due to the absence of specific and repeatable EEG patterns that can be observed in this phase and to the low number of epochs in the dataset compared to the other sleep stages, since for a healthy human subject, the N1 sleep is the shortest phase in sleep cycle (about 2-5%) [7]. For this reason, the performances of the classification model were also tested after having balanced the dataset: the results were negatively affected by the operation of class balancing since the number of training epochs in the other sleep stages was reduced to balance the number of N1 epochs and especially, when the network was validated on the test set, which represents a generic subject's recording with the typical physiological percentages of each sleep phase in the human sleep cycle, the performances were not satisfactory. This problem may be addressed by using data augmentation techniques to address class imbalance.

In the context of data augmentation, a novel strategy could be the development of generative adversarial networks (GANs). GANs are an emergent class of deep learning networks with the aim of generating artificial signals to feed to learning algorithms. There are two models behind the GAN, the generator and the discriminator. The generator model is learned to synthesize realistic data samples, while the discriminator is a model typically used for classification in machine and deep learning with the task of distinguishing between different classes. Generative models try to learn how to make a realistic representation of a certain class and discriminative models are simultaneously trained to distinguish real training data samples from artificially generated ones. These models compete against each other which is why they are called adversarial networks [42], [43]. To the best of our knowledge, no robust ASSC methods based on GANs have been published in literature. Recently, a novel method which employs GANs and RNNs with LSTM blocks, has been published with the aim of detecting sleepiness during driving by employing EEG and EOG signals [44]. Another strategy of data augmentation is the application of Fourier transform-based methods for the generation of surrogate data that approximate real examples [45]. In literature, has also been proposed a novel cross-entropy loss function, named focal loss (FL), employed as the loss function between the true labels and the network predictions, where the inliers (easy examples) are down-weighted such that their contribution to the total loss is small even if their number is large [46]. Our research group is currently working on the implementation, in a deep learning framework, of architectures based on GANs in Python environment both for biomedical signal processing and medical image classification.

Another unresolved issue emerges in the following context: single-channel vs multichannel approach. From one side, with the use of single-channel EEG signals remains very difficult to improve the performances of stages N1 and REM due to their heterogeneity and as relevant information is hidden in EOG and EMG signals, but with this approach the acquisition setup is more simple with less noise

interference and it may be used for NCP assessment in real-time applications, e.g. the detection of driver drowsiness [47]. On the other side, multichannel recording can be used only as a support for clinical diagnosis but the performances in sleep scoring may be improved, especially with the advent of deep learning frameworks. In a future study the same workflow described in this chapter may be used by taking into account also several EEG channels, EOG and EMG signals, especially for nocturnal W, stage N1 and REM sleep detection. All the signal decomposition techniques and time-frequency distributions could also be evaluated for the analysis and interpretation of other biomedical signals (e.g. EMG, ECG, heart rate variability or evoked potentials).

In this moment, the advantage of using CNNs over RNNs for biomedical signal processing and classification is not seen for several reasons: the performances are not superior, the number of hyperparameters to set is higher and the relation between handcrafted features and subject's condition cannot be taken into account into a CNN framework. Moreover, the advantage of RNNs is to handle directly temporal sequences such as time series. The disadvantage of BRNNs with respect to unidirectional RNNs is that the entire sequence of data is taken into account to make predictions and for this reason is not intended for real-time applications. A recent technique belonging to the class of XAI (eXplainable Artificial Intelligence) is the layer-wise relevance propagation (LRP) which aims to explain which features are responsible of class prediction [48]. LRP technique can be considered a post-processing method since it can be applied on already trained classifiers and will be evaluated in a future study.

## References

- [1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
- [2] L. Wei, Y. Lin, J. Wang, and Y. Ma, “Time-Frequency Convolutional Neural Network for Automatic Sleep Stage Classification Based on Single-Channel EEG,” in *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2017, pp. 88–95, doi: 10.1109/ICTAI.2017.00025.
- [3] R. S. Rosenberg and S. Van Hout, “The American Academy of Sleep Medicine Inter-scorer Reliability Program: Sleep Stage Scoring,” *J. Clin. Sleep Med.*, vol. 9, no. 1, pp. 81–87, Jan. 2013, doi: 10.5664/jcsm.2350.
- [4] C. Iber, S. Ancoli-Israel, A. L. Chesson, and S. Quan, “The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications,” *Westchester, Am. Acad. Sleep Med.*, Jan. 2007.
- [5] K. M. Gunnarsdottir *et al.*, “A novel sleep stage scoring system: Combining expert-based features with the generalized linear model,” *J. Sleep Res.*, vol. 29, no. 5, p. e12991, Oct. 2020, doi: <https://doi.org/10.1111/jsr.12991>.
- [6] R. Boostani, F. Karimzadeh, and M. Nami, “A comparative review on sleep stage classification methods in patients and healthy individuals,” *Comput. Methods Programs Biomed.*, vol. 140, pp. 77–91, 2017, doi: <https://doi.org/10.1016/j.cmpb.2016.12.004>.
- [7] H. R. Colten and B. M. Altevogt, *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. National Academies Press (US), 2006.
- [8] M. Roohi-Azizi, L. Azimi, S. Heysiattalab, and M. Aamidfar, “Changes of the brain’s bioelectrical activity in cognition, consciousness, and some mental disorders,” *Med. J. Islam. Repub. Iran*, vol. 31, p. 53, Sep. 2017, doi: 10.14196/mjiri.31.53.
- [9] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. C. Kamphuisen, and J. J. L. Obery, “Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG,” *IEEE Trans. Biomed. Eng.*, vol. 47, no. 9, pp. 1185–1194, 2000, doi: 10.1109/10.867928.
- [10] A. L. Goldberger *et al.*, “PhysioBank, PhysioToolkit, and PhysioNet,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000, doi: 10.1161/01.CIR.101.23.e215.
- [11] 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, doi: 10.1046/j.1440-1819.2001.00810.x.
- [12] R. K. Malhotra and A. Y. Avidan, “Chapter 3 - Sleep Stages and Scoring Technique,” S. Chokroverty and R. J. B. T.-A. of S. M. (Second E. Thomas, Eds. St. Louis: W.B. Saunders, 2014, pp. 77–99.
- [13] H. Tanaka, M. Hayashi, and T. Hori, “Topographical characteristics and principal component structure of the hypnagogic EEG,” *Sleep*, vol. 20, no. 7, pp. 523–534, Jul. 1997, doi: 10.1093/sleep/20.7.523.
- [14] S. Butterworth, “On the theory of filter amplifiers,” *Wirel. Eng.*, vol. 7, no. 6, pp. 536–541, 1930.
- [15] I. Daubechies, *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.
- [16] S. Mallat, *A wavelet tour of signal processing*. Elsevier, 1999.

- [17] P. J. Van Fleet, *Discrete wavelet transformations: An elementary approach with applications*. John Wiley & Sons, 2011.
- [18] E. Oropesa, H. L. Cycon, and M. Jobert, "Sleep Stage Classification using Wavelet Transform and Neural Network," in *Proceedings of the Joint Conference on Information Sciences*, 2000, vol. 5, no. 1, pp. 811–814.
- [19] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "A Wavelet-Chaos Methodology for Analysis of EEGs and EEG Subbands to Detect Seizure and Epilepsy," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 2, pp. 205–211, 2007, doi: 10.1109/TBME.2006.886855.
- [20] F. Ebrahimi, M. Mikaeili, E. Estrada, and H. Nazeran, "Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients.," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2008, pp. 1151–1154, 2008, doi: 10.1109/IEMBS.2008.4649365.
- [21] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. London A Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998, doi: 10.1098/rspa.1998.0193.
- [22] J. S. Smith, "The local mean decomposition and its application to EEG perception data," *J. R. Soc. Interface*, vol. 2, no. 5, pp. 443–454, Dec. 2005, doi: 10.1098/rsif.2005.0058.
- [23] Z. Liu, Y. Jin, M. J. Zuo, and Z. Feng, "Time-frequency representation based on robust local mean decomposition for multicomponent AM-FM signal analysis," *Mech. Syst. Signal Process.*, vol. 95, pp. 468–487, 2017, doi: <https://doi.org/10.1016/j.ymsp.2017.03.035>.
- [24] K. Dragomiretskiy and D. Zosso, "Variational Mode Decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, 2014, doi: 10.1109/TSP.2013.2288675.
- [25] L. Stanković, M. Daković, and T. Thayaparan, "Chapter 3 - Non-Stationary Signal Analysis Time-Frequency Approach," in *Academic Press Library in Signal Processing: Volume 3*, vol. 3, A. M. Zoubir, M. Viberg, R. Chellappa, and S. B. T.-A. P. L. in S. P. Theodoridis, Eds. Elsevier, 2014, pp. 27–142.
- [26] D. Komorowski and S. Pietraszek, "The Use of Continuous Wavelet Transform Based on the Fast Fourier Transform in the Analysis of Multi-channel Electrogastrography Recordings," *J. Med. Syst.*, vol. 40, no. 1, pp. 1–15, 2016, doi: 10.1007/s10916-015-0358-4.
- [27] A. Mertins, *Signal analysis: wavelets, filter banks, time-frequency transforms and applications*. John Wiley & Sons, Inc., 1999.
- [28] T. K. Ugur and A. Erdamar, "An efficient automatic arousals detection algorithm in single channel EEG," *Comput. Methods Programs Biomed.*, vol. 173, pp. 131–138, 2019, doi: <https://doi.org/10.1016/j.cmpb.2019.03.013>.
- [29] V. Bajaj and R. B. Pachori, "Automatic classification of sleep stages based on the time-frequency image of EEG signals," *Comput. Methods Programs Biomed.*, vol. 112, no. 3, pp. 320–328, 2013, doi: <https://doi.org/10.1016/j.cmpb.2013.07.006>.
- [30] B. Boashash and S. Ouelha, "Automatic signal abnormality detection using time-frequency features and machine learning: A newborn EEG seizure case study," *Knowledge-Based Syst.*, vol. 106, pp. 38–50, 2016, doi: <https://doi.org/10.1016/j.knosys.2016.05.027>.

- [31] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 10–19, 2012, doi: <https://doi.org/10.1016/j.cmpb.2011.11.005>.
- [32] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66, 1979, doi: 10.1109/TSMC.1979.4310076.
- [33] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Comput. Biol. Med.*, vol. 96, pp. 189–202, 2018, doi: <https://doi.org/10.1016/j.compbiomed.2018.03.016>.
- [34] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked bidirectional and unidirectional lstm recurrent neural network for forecasting network-wide traffic state with missing values," *arXiv*. arXiv, May 23, 2020, Accessed: Nov. 26, 2020. [Online]. Available: <http://arxiv.org/abs/2005.11627>.
- [35] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv Prepr. arXiv1412.6980*, 2014, Accessed: Jul. 26, 2018. [Online]. Available: <http://arxiv.org/abs/1412.6980>.
- [36] T. Zhu, W. Luo, and F. Yu, "Convolution-and attention-based neural network for automated sleep stage classification," *Int. J. Environ. Res. Public Health*, vol. 17, no. 11, pp. 1–13, Jun. 2020, doi: 10.3390/ijerph17114152.
- [37] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "Joint Classification and Prediction CNN Framework for Automatic Sleep Stage Classification," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1285–1296, 2019, doi: 10.1109/TBME.2018.2872652.
- [38] A. I. Humayun, A. S. Sushmit, T. Hasan, and M. I. H. Bhuiyan, "End-To-end sleep staging with raw single channel EEG using deep residual convnets," in *2019 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2019 - Proceedings*, 2019, pp. 1–5, doi: 10.1109/BHI.2019.8834483.
- [39] Q. Cai, Z. Gao, J. An, S. Gao, and C. Grebogi, "A Graph-Temporal fused dual-input Convolutional Neural Network for Detecting Sleep Stages from EEG Signals," *IEEE Trans. Circuits Syst. II Express Briefs*, pp. 1–1, 2020, doi: 10.1109/tcsii.2020.3014514.
- [40] L. Torsvall and T. Åkerstedt, "Extreme sleepiness: Quantification of EOG and spectral EEG parameters," *Int. J. Neurosci.*, vol. 38, no. 3–4, pp. 435–441, Jan. 1988, doi: 10.3109/00207458808990704.
- [41] S. Charbonnier, L. Zoubek, S. Lesecq, and F. Chapotot, "Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging," *Comput. Biol. Med.*, vol. 41, no. 6, pp. 380–389, 2011, doi: <https://doi.org/10.1016/j.compbiomed.2011.04.001>.
- [42] A. M. Shaker, M. Tantawi, H. A. Shedeed, and M. F. Tolba, "Generalization of Convolutional Neural Networks for ECG Classification Using Generative Adversarial Networks," *IEEE Access*, vol. 8, pp. 35592–35605, 2020, doi: 10.1109/ACCESS.2020.2974712.
- [43] Y. Luo, L. Z. Zhu, Z. Y. Wan, and B. L. Lu, "Data augmentation for enhancing EEG-based emotion recognition with deep generative models," *J. Neural Eng.*, vol. 17, no. 5, p. 56021, 2020, doi: 10.1088/1741-2552/abb580.
- [44] Y. Jiao, Y. Deng, Y. Luo, and B.-L. Lu, "Driver sleepiness detection from EEG and EOG signals using GAN and LSTM networks," *Neurocomputing*,

- vol. 408, pp. 100–111, 2020, doi: <https://doi.org/10.1016/j.neucom.2019.05.108>.
- [45] J. T. C. Schwabeda, J. C. Snyder, A. Cakmak, S. Nemati, and G. D. Clifford, “Addressing class imbalance in classification problems of noisy signals by using fourier transform surrogates,” *arXiv*, 2018.
- [46] J. Gao, H. Zhang, P. Lu, and Z. Wang, “An Effective LSTM Recurrent Network to Detect Arrhythmia on Imbalanced ECG Dataset,” *J. Healthc. Eng.*, vol. 2019, p. 6320651, 2019, doi: 10.1155/2019/6320651.
- [47] I. Belakhdar, W. Kaaniche, R. Djmel, and B. Ouni, “Detecting driver drowsiness based on single electroencephalography channel,” in *13th International Multi-Conference on Systems, Signals and Devices, SSD 2016*, 2016, pp. 16–21, doi: 10.1109/SSD.2016.7473671.
- [48] L. Arras, G. Montavon, K. R. Müller, and W. Samek, “Explaining recurrent neural network predictions in sentiment analysis,” *arXiv*. arXiv, Jun. 22, 2017, doi: 10.18653/v1/w17-5221.

# Chapter 4

## Introduction - Part II

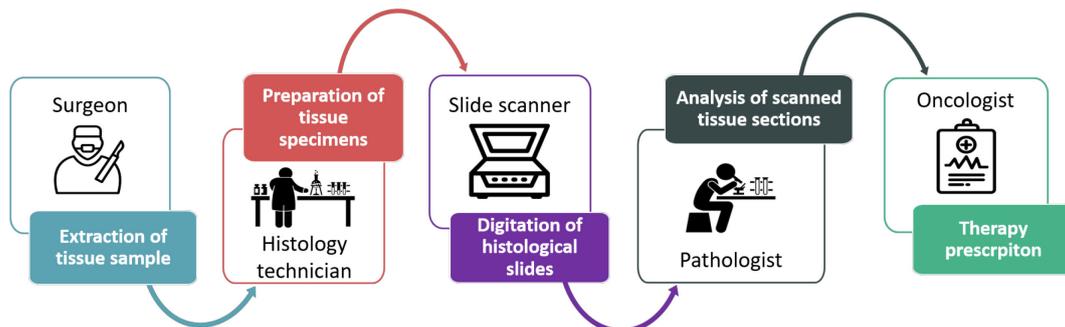
### 4.1 Digital Pathology

Anatomic pathology is the branch of medicine with the aim to detect and identify human diseases by macroscopic examination of organs and microscopic examination of tissues and cells. Within the diagnostic process, it plays a fundamental role and aims to develop a complete and accurate clinical diagnosis. The identification of abnormalities is the primary role of anatomic pathology since it can help to diagnose diseases and plan treatment. Anatomic pathology is split into two main categories: histology (or histopathology) and cytology (or cytopathology). In histopathological images several specimens are visible: isolated cells, cell nuclei, cell clusters and more extensive structures such as epithelium, gland lumen and dense connective tissue or stroma. On the other hand, cytology, also called cell biology, is the branch of biology which studies the structure and the chemistry of cells from a functional point of view. The difference is reflected in the image content which consists of simpler structures compared to histopathology [1].

The traditional process based on the evaluation of glass slide samples under a light microscope is time consuming and includes a risk of damaging cell structures of biological tissue associated with the transfer of the physical slides. In addition, there is a lack of concordance, between different pathologists, in the final diagnosis of cancer and the cost of physical shipping slides for a second opinion or external consultation is high. A report of the UK's National Patient Safety Agency (March 2010) found that delays in decision making in pathology account for 41% of delays in cancer diagnosis [2].

Digital pathology is one of the new declinations of digital healthcare. It is proposed to use digital technology to improve the diagnostic and predictive process of anatomic pathology. The advent of whole slide digital scanners has allowed the digitalization of the optically scanned tissue slides which can now be stored in digital image form, visualized on desktop computers, laptops, tablets, and even

smartphones and interpreted using advanced software for diagnostic or research purposes. Digital images can be analyzed with software to quantify histopathological parameters (e.g. the percentage of cells that are proliferating) with the advent of computational pathology. In addition, digital pathology exploits algorithms, based on Artificial Intelligence (AI), with a growing capability in managing big data and speeding up the pathologist's work but, in perspective, they may be able to formulate diagnostics to be presented directly to the oncologist [3]. The workflow of digital pathology is summarized in Figure 4.1.



**Fig. 4.1** Workflow of Digital Pathology.

In digital pathology, the process consists generally of the following steps: firstly, the tissue removal is performed by the surgeon, then the sample is prepared for the microscopic examination in the histology laboratory; this phase involves a series of complex operations. Tissue block thick about 1 cm are subjected to *i*) chemical fixation with 4% aqueous solution of formaldehyde in order to stabilize and maintain the structure of the tissue, kill bacteria and inactivate enzymes that otherwise might degrade the tissue. The aim is to leave tissues as close as their living state as possible. Subsequently, specimen is *ii*) dehydrated by gradually replacing water in the sample with alcohol and cleared (i.e. the alcohol is replaced with xylene which is a paraffin solvent). Then, the tissue is *iii*) embedded in warm paraffin wax; the aim of this process is the tissue embedding in a solid medium stiff enough to support the tissue during the subsequent step of section cutting. The following phase is *iv*) section cutting using a microtome, which is a mechanical instrument used to cut biological specimens into very thin slices (2-10  $\mu\text{m}$ ) to allow light to pass through. Finally, the sections are *v*) stained to make cell structures visible and mounted on a glass slide to keep specimens pressed flat and protect them from accidental contact or dust [4]. The histology technician, which is a specialized medical lab worker, performs all the previous tasks involving the acquisition, processing and preparation of patient tissue specimens. Subsequently, the pathologist analyzes the optically scanned tissue slides and finally the oncologist may proceed with therapy accordingly.

The human eye can distinguish two points as distinct only if they are at least 0.2 mm apart; hence the visual examination of biological tissues under light microscope is necessary. In histological specimens, erythrocytes are commonly used as a “histological ruler” to measure the size of other cells and structures found

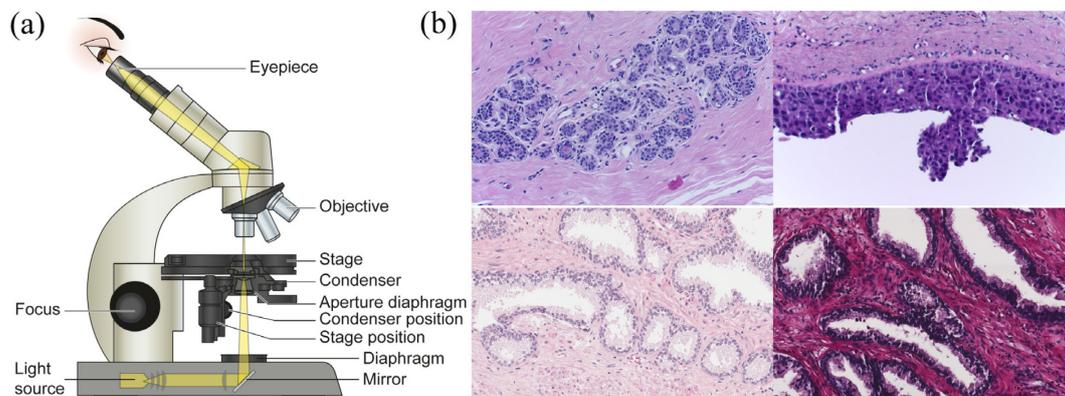
in the tissue section, since they are uniform in size, 7 to 8  $\mu\text{m}$  in diameter [4]. The basic optical microscope elements are shown in Figure 4.2 (a). The modern microscope has also a digital camera connected to a workstation. The resolution of light microscope is 0.2  $\mu\text{m}$ , while the transmission electron microscope (TEM) has a three orders of magnitude higher resolution. Electrons have shorter wavelength than visible light (i.e. from about 380 to 740 nm) and therefore permit several million times magnification. TEM is employed only for specific applications which require a higher level of microscopy (e.g. to visualize very small tissue structures such as some organelles, membranes, macromolecular complexes or viruses). The analysis and interpretation of most histopathological images can be performed at different magnifications. Most commonly, scanned tissue slides in routine histology are viewed at 10 $\times$ , 20 $\times$  or 40 $\times$  magnification, regardless of the scanning and focusing method [3].

The staining process is necessary since cell structures are transparent and colorless, thus tissue sections are prepared using colored stains that bind selectively to cellular components. The techniques used can either be non-specific, i.e. most of the cells are stained in the same way, or specific, i.e. particular cell components are stained with a bright color and the rest of the tissue with a different color by using a counterstain [5]. Histological staining techniques are split into two main categories: histochemistry and immunohistochemistry. Histochemistry refers to histological stains which bind selectively to cell constituents and highlight chemical features. Hematoxylin and Eosin (H&E), periodic acid–Schiff (PAS) and trichrome staining (e.g. Mallory's and Masson's trichrome) are the most frequently used stains which belong to this first group [6]. In routine histology, the most widely used staining method is H&E. Hematoxylin is a basic and positively charged dye which stains acid molecules such as nuclei of cells with a blue or purple color, while eosin is an anionic and acidic dye which stains with a pinkish color cytoplasmatic structures such as cell stroma [7].

On the other hand, immunohistochemistry (IHC) is an immunology-based method which employs tissue-based biomarkers (i.e. antibodies marked with specific reagents) for the localization of specific tissue components (generally these are proteins) that act as antigens. Thus, IHC technique is based on antigen-antibody interaction combined with enzymatic or fluorescent detection systems that make the reaction visible under a light microscope [8]. For this reason, pathologists select specific antibodies in order to identify cellular structures or elements that are associated with certain human diseases for diagnostic, prognostic, and therapeutic purposes. The tissue stains differently in the presence of cancer, inflammatory response, infections or autoimmune diseases. IHC staining is widely used in the diagnosis of cancerous tumors, especially for breast cancer [9].

The manual processes currently in place within the laboratories may be subjected to errors as they depend solely on the control of individual operators. The process is complex and consists of several steps that involve the transfer of the biological sample in different forms, each of which represents a potential source of error. In addition, the histological tissue can take on different staining intensities

depending on the following factors: *i*) the concentration, the pH and the degree of deterioration of the staining dye, *ii*) the staining exposure time; *iii*) the ability of histology technician, *iv*) the illumination condition and *v*) the scanner performances and specifications [10]. An example of stain variability in H&E stained slides is reported in Figure 4.2 (b).



**Fig. 4.2** (a) The light microscope. (b) Staining variability of tissue specimens stained with Hematoxylin and Eosin (H&E).

Thus, there is a need for an automated system that can help the operator in the management and standardization of the process by eliminating possible errors. The diagnosis based on histological images is often influenced by the staining quality of the histological preparation. If the quality of the staining is not satisfactory for the pathologist, an additional portion of the tissue will be sectioned and stained with a consequent loss of time and financial resources. This variability in the color of the histological preparation affects both the pathologist's diagnostic process and the performance of computer-assisted diagnosis (CAD) systems [1]. In this context, a standardization procedure of histological slides becomes relevant. To address this problem, a color (stain) normalization method is proposed in the second part of this thesis work. This method allows to standardize color of an original image with respect to that of a reference image (gold standard) which contains correct and uniform stain intensities according to an expert pathologist. The color standardization and normalization of histological slides may be useful to increase the speed and accuracy of the diagnosis made by the pathologist, eliminate delays due to a potential re-staining and increase the performances of an automated algorithm employed for the processing of histological images.

## 4.2 Color Deconvolution

Stain separation assumes great importance in the development of automatic solutions for digital histopathology image analysis, as it allows a correct separation of cell structures. As discussed before, staining dye molecules bind selectively to cell components which are highlighted during the staining process. This is useful for the pathologist who performs the subsequent analysis for diagnosis purposes. The structures that have bound to one stain are separated from other cell

components which have bound to another stain. More specifically, the different contributions related to two or more stains are isolated in a digital histological image. This process is called stain separation and is based on the mathematical technique defined as Color Deconvolution [11].

The tissue slides exposed to incident light are converted in digital form by using a slide scanner and stored as RGB (red (R), green (G) and blue (B) color) images, by using a 3-sensor RGB camera. Color deconvolution allows to isolate in an RGB image, two or more channels related to different stain concentrations which assume a specific color after light exposure. Color deconvolution is based on a similar technical basis to that of signal decomposition in the field of signal processing. A generic signal can be decomposed using a digital filter or a specific decomposition technique in different sub-bands related to frequency content, while an RGB stained image can be decomposed into different channels related to stain concentration.

The stain color intensity is related to the stain concentration through the well-known physical Beer-Lambert (BL) law. According to BL law, the absorbance  $A$  of an absorbing medium can be computed as the ratio of the transmitted light intensity  $I$  (i.e. the number of photons who pass through the sample along the incident direction without being completely absorbed or diffused) with respect to the incident light intensity  $I_0$ . In optics, the absorbance is also called decadic absorbance, since a common logarithm is used in the mathematical expression:

$$A = -\log_{10} \left( \frac{I}{I_0} \right) = \varepsilon \cdot c \cdot l \quad (\text{Eq. 4.1})$$

The absorbance can also be expressed as the product of the molar extinction coefficient  $\varepsilon$ , which quantifies how strongly a chemical specimen absorbs light at a specific wavelength, the molar concentration  $c$ , i.e. the amount of absorbing species dissolved in a solution and the optical path length  $l$  [12].

From a numerical point of view, the input RGB histological image with dimensions  $n_H \times n_W \times 3$ , where  $n_H$ ,  $n_W$ , 3 are the number of rows, columns and RGB channels respectively is reshaped in a vectorized form where each column represents all the grayscale pixel intensities for a specific channel (R, G or B). Thus, the input image can be treated as a matrix  $I \in \mathbb{R}^{N \times 3}$  with  $N = n_H \cdot n_W$ . The incident light intensity  $I_0$  is set to a constant, which is 255 for images with a dynamic range of 8 bits [13].

The absorption of the tissue depends on the tissue itself but also on its instantaneous characteristics. The BL law is reported in a standard form without the term related to the scattering phenomenon, which is not relevant in digital histology image analysis, since tissue sections are very thin slices (i.e. 2-10  $\mu\text{m}$ ). The image intensity or equivalently the transmission of light, as shown in Eq. 4.1, is related to the absorbance in a nonlinear way, while the absorbance is directly proportional to the molar concentration of absorbing species [14]. In routine histology, the thickness of the sample assumes a very small value and is almost the same for all tissue sections, thus the concept of optical density has been introduced [15]. The

Optical Density (OD) is defined as the absorbance per unit path length and it can be expressed as the product of two non-negative matrices:

$$V = H \cdot W \quad (\text{Eq. 4.2})$$

where  $V$  is the OD matrix,  $W$  is called stain color appearance matrix, which contains the molar extinction coefficients and  $H$  is the stain density map or concentration map which contains the molar concentration values. Both matrices cannot assume negative values due to their real physical significance. The stain color appearance matrix  $W \in \mathbb{R}^{r \times 3}$ , the stain density map  $H \in \mathbb{R}^{N \times r}$  and  $r$  is the number of stains (e.g. in H&E stained images  $r$  equals two). Of course,  $V$  is the vectorized image converted in OD space, hence  $V \in \mathbb{R}^{N \times 3}$  [13]. In Figure 4.3, the structure of matrices  $W$  and  $H$  involved in the image decomposition in OD space is reported.

$$W = \begin{bmatrix} 2 \times 3 & R & G & B \\ 1^{st} \text{stain} & \cdot & \cdot & \cdot \\ 2^{nd} \text{stain} & \cdot & \cdot & \cdot \end{bmatrix} \quad H = \begin{bmatrix} N \times 2 & 1^{st} \text{stain} & 2^{nd} \text{stain} \\ 1 & \cdot & \cdot \\ 2 & \cdot & \cdot \\ \vdots & \vdots & \vdots \\ N & \cdot & \cdot \end{bmatrix}$$

$W \in \mathbb{R}^{r \times 3}$

$H \in \mathbb{R}^{N \times r}$

**Fig. 4.3** Structure of stain color appearance matrix (left) and stain density map (right) when the number of stains is equal to 2.

The rows of matrix  $W$  are also called stain vectors since each row contains the RGB triplet of a specific stain. These rows are normalized in order to obtain stain vectors with unit Euclidean norm to achieve the correct relative factor for each separate stain [14]. On the other hand, each column of matrix  $H$  refers to the concentration of a specific stain, for all pixels in the image.

In the color deconvolution technique, the first step is the conversion of grayscale intensity of the input image, for each channel (R, G or B), into OD space using the following equation:

$$V(:, k) = -\log_{10} \left( \frac{I(:, k)}{I_0} \right), \quad k = 1, 2, 3 \quad (\text{Eq. 4.3})$$

Subsequently the problem is the estimation of the stain color appearance matrix  $W$ . In the literature, several state-of-the-art methods have been proposed with the purpose of estimating this matrix and they are discussed in detail in Chapter 5. The matrix  $H$  can be computed through the inversion of matrix  $W$  in the linear system of Eq. 4.2. If the number of stains is different from 3 (e.g. in H&E stained images the number of stains is 2), matrix  $W$  results in a non-square matrix and for this reason a Moore-Penrose pseudoinverse must be computed or equivalently a third stain vector must be defined as the cross product of the first two rows (i.e. the first

and second stain vector) [16]. The mathematical expression of matrix  $H$  is obtained as follows:

$$\begin{aligned} V^T &= W^T \cdot H^T \\ H^T &= (W^T)^{-1} \cdot V^T \\ H &= ((W \cdot W^T)^{-1} \cdot W \cdot V^T)^T \end{aligned} \quad (\text{Eq. 4.4})$$

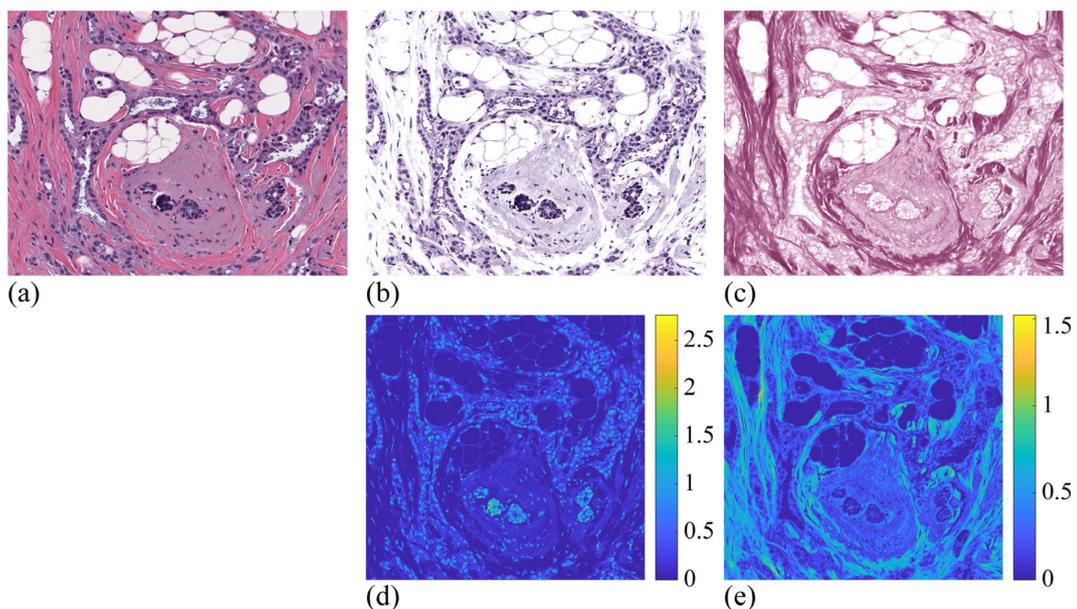
where the superscript  $T$  denotes the transpose operator. The issue is that the linear system reported in Eq. 4.2 is unconstrained, hence  $H$ , in Eq. 4.4, can assume negative values but this is impossible due to its physical meaning (in the previous works these negative values were simply set to zero). In this thesis work a solution to this problem was proposed and discussed in Chapter 5. After estimating and computing respectively the two matrices  $W$  and  $H$ , the input image is decomposed into two separate channels (if the number of stains is equal to 2) using the following equation:

$$V_j = H(:, j) \cdot W(j, :), \quad j = 1, 2 \quad (\text{Eq. 4.5})$$

Finally, the separate channels in RGB color space are computed by inverting the BL law reported in Eq. 4.3 as:

$$I_j = I_0 \cdot 10^{-V_j}, \quad j = 1, 2 \quad (\text{Eq. 4.6})$$

In Figure 4.4 an example of color deconvolution technique applied to an H&E stained histological image is reported. On the first row (a-c) the original image and the two decomposed channels are shown; while on the second row, the concentration values of matrix  $H$  are displayed with their corresponding colorbar. The stain density values are bounded below, as can be seen in the colorbar, since negative value for concentration are not physically allowed.



**Fig. 4.4** Example of color deconvolution technique: (a) input RGB H&E stained image; (b) first decomposed channel (i.e. cell nuclei); (c) second decomposed channel (i.e. cell stroma); (d-e) stain density map for hematoxylin and eosin respectively.

Color deconvolution is useful to convert multiple stained biological samples into images representing the individual stain concentrations (e.g. structures stained only with hematoxylin and structures stained only with eosin dye) in OD space since the absorbance (normalized to the optical path length) assumes a linear relationship with concentration under monochromatic conditions. However, the concentrations of multiple stains are not mutually exclusive since certain cell structures absorb both hematoxylin and eosin dye at some wavelengths. The problem of stain mixtures represents an open issue for this image decomposition technique [17]. Stain separation performed by color deconvolution is the basic for the more advanced strategy of stain normalization proposed in this thesis and described in the following chapters.

### 4.3 Aim of the Thesis - Part II

The aim of the work presented here is to develop and validate a set of multi-tissue and multiscale automated solutions for the separation and normalization of histological stains in the field of medical imaging and more specifically in digital histopathology image analysis.

In digital pathology, the diagnosis of histological images is based on the visual examination of small tissue portions under an optical microscope. Cellular structures, which are transparent to visible light wavelength, are made visible by using histological stains which bind selectively to specific biological structures and components within the tissue with the aim to create a relevant contrast. The reproducibility of histological specimens is affected by color intensity variation of stained images due to several factors: *i*) the concentration, the pH and the degree of deterioration of the staining dye, *ii*) the staining exposure time, *iii*) the ability of histology technician, *iv*) the illumination condition and *v*) the scanner specifications.

The solution to this problem is the standardization of color appearance in digital pathology. This concept can be addressed through a stain normalization method which adapts the color intensity of histological images to a reference template, chosen by an expert pathologist, which presents optimal characteristics in terms of stain distribution, illuminant conditions and color intensity saturation. The aim of the stain normalization strategy is twofold: firstly the standardization of stain color appearance in digital pathology could improve the pathologist's work in the diagnosis of biological diseases; secondly this method could be used as a pre-processing step for subsequent computer-assisted diagnosis (CAD) systems for accurate cellular structure segmentation, classification and quantification of histological prognostic parameters. In this second part (Part II), two completely automated stain separation and normalization strategies, which were both quantitatively and qualitatively assessed and compared to other state-of-the-art methods in the same research field yielding promising results, are proposed.

In Chapter 5, a novel image processing approach of stain separation and normalization for histological images stained with histochemical Hematoxylin and

Eosin (H&E), is proposed. The algorithm called SCAN (Stain Color Adaptive Normalization) consists in the development of segmentation and clustering strategies for cellular structures detection and standardizes the color of a histological image with respect to the stain color of a target image with a consequent increase of the contrast between tissue regions and background. In addition, the cellular structures of the original image are not distorted and preserved.

In Chapter 6, an automated strategy for the estimation of stain color appearance matrix for immunohistochemical (IHC) staining dyes, is proposed and a stain normalization approach, applied to IHC stained histopathological images, is developed. The method is based on stain estimation using chromaticity plane and blue ratio image and the normalization is performed on stain channels, separately. Finally, the normalized image is reconstructed using a novel proposed mathematical technique named Inverse Color Deconvolution (ICD) and the unstained structures such as gland lumen and background are preserved.

The conclusions and final remarks of this work are reported in the last section of the thesis.

## References

- [1] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological Image Analysis: A Review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009, doi: 10.1109/RBME.2009.2034865.
- [2] N. T. Omorodion, "Review on Digitalization of Basic Routine Procedures in Histopathology Laboratories," *Int. J. Biosens. Bioelectron.*, vol. 3, no. 5, Dec. 2017, doi: 10.15406/ijbsbe.2017.03.00085.
- [3] F. Aeffner *et al.*, "Digital Microscopy, Image Analysis, and Virtual Slide Repository," *ILAR J.*, vol. 59, no. 1, pp. 66–79, Dec. 2018, doi: 10.1093/ilar/ily007.
- [4] A. Rehfeld, M. Nylander, K. Karnov, A. Rehfeld, M. Nylander, and K. Karnov, "Histological Methods," in *Compendium of Histology*, Springer International Publishing, 2017, pp. 11–24.
- [5] M. T. McCann, J. A. Ozolek, C. A. Castro, B. Parvin, and J. Kovacevic, "Automated Histology Analysis: Opportunities for signal processing," *IEEE Signal Process. Mag.*, vol. 32, no. 1, pp. 78–87, 2015, doi: 10.1109/MSP.2014.2346443.
- [6] V. Varone, C. Bellevicine, and G. Troncone, "Biopsic sampling (cancer)," in *Comprehensive Sampling and Sample Preparation*, vol. 1, Elsevier Inc., 2012, pp. 413–439.
- [7] A. H. Fischer, K. A. Jacobson, J. Rose, and R. Zeller, "Hematoxylin and eosin staining of tissue and cell sections.," *CSH Protoc.*, vol. 2008, no. 5, p. pdb.prot4986, May 2008, doi: 10.1101/pdb.prot4986.
- [8] J. A. Ramos-Vara and M. A. Miller, "When Tissue Antigens and Antibodies Get Along: Revisiting the Technical Aspects of Immunohistochemistry—The Red, Brown, and Blue Technique," *Vet. Pathol.*, vol. 51, no. 1, pp. 42–87, Oct. 2013, doi: 10.1177/0300985813505879.
- [9] E. Gudlaugsson *et al.*, "Comparison of the effect of different techniques for measurement of Ki67 proliferation on reproducibility and prognosis prediction accuracy in breast cancer," *Histopathology*, vol. 61, no. 6, pp. 1134–1144, Dec. 2012, doi: 10.1111/j.1365-2559.2012.04329.x.
- [10] J. D. Bancroft and M. Gamble, *Theory and practice of histological techniques*. Elsevier health sciences, 2008.
- [11] T. C. Cornish and M. K. Halushka, "Color deconvolution for the analysis of tissue microarrays.," *Anal. Quant. Cytol. Histol.*, vol. 31, no. 5, pp. 304–312, Oct. 2009.
- [12] D. F. Swinehart, "The Beer-Lambert Law," *J. Chem. Educ.*, vol. 39, no. 7, p. 333, 1962, doi: 10.1021/ed039p333.
- [13] A. Vahadane *et al.*, "Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images," *IEEE Trans. Med. Imaging*, vol. 35, no. 8, pp. 1962–1971, 2016, doi: 10.1109/TMI.2016.2529665.
- [14] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution.," *Anal. Quant. Cytol. Histol.*, vol. 23, no. 4, pp. 291–299, Aug. 2001.
- [15] M. Yi, Z. Shen, X. Zhang, and S. Ma, "Achieving concentrated graphene dispersions in water/acetone mixtures by the strategy of tailoring Hansen solubility parameters," *J. Phys. D: Appl. Phys.*, vol. 46, no. 2, p. 25301, 2012, doi: 10.1088/0022-3727/46/2/025301.

- [16] R. Penrose, “A generalized inverse for matrices,” *Math. Proc. Cambridge Philos. Soc.*, vol. 51, no. 3, pp. 406–413, 1955, doi: DOI: 10.1017/S0305004100030401.
- [17] P. Haub and T. Meckel, “A Model based Survey of Colour Deconvolution in Diagnostic Brightfield Microscopy: Error Estimation and Spectral Consideration,” *Sci. Rep.*, vol. 5, p. 12096, Jul. 2015, doi: 10.1038/srep12096.

## Chapter 5

# Stain Separation and Normalization of Histological Images in Digital Pathology

*Part of this chapter has been published as:*

M. Salvi, N. Michielli and F. Molinari. "**Stain Color Adaptive Normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology.**" *Computer Methods and Programs in Biomedicine* 193 (2020): 105506.

## 5.1 Introduction

As discussed in Chapter 4, the diagnosis of histological and histopathological images is based on the visual examination of small tissue portions of under an optical microscope. Cellular structures, which are transparent to visible light wavelength, are made visible by using histological stains which bind selectively to specific biological structures and components within the tissue with the aim to create a relevant contrast. The pathologist analyzes optically scanned tissue glass slides to formulate a diagnosis and subsequently the oncologist prescribes a therapy accordingly. With the advent of digital pathology, tissue glass slides, with different magnifications, can be stored in a digital form and evaluated using advanced software solutions, with the advantage of speeding up the pathologist's work and reducing errors due to the transit of physical slides in the traditional manual process.

The analysis of histopathological images represents the gold standard for several kinds of human diseases including different types of cancer [1], [2]. In fact, the tumor aggressiveness, i.e. the rapidity of growth and spreading to other organs, is evaluated conventionally through a standard scoring. Currently, the prostate cancer malignancy degree is evaluated via visual examination of stained microscope images and the cancer staging is performed according to the Gleason Score. Briefly, this system classifies the architecture of prostate glands into 5 Gleason Grade Groups, according to the recent guidelines from the 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma [3].

A major issue in this process is that intra- and inter-operator variability may negatively affect the manual evaluation of histological images [4], [5]. To solve this problem, automated solutions for digital histopathology image analysis have been proposed with the aim to reduce the human variability and to make the pathologist's work less time-consuming. In addition, the reproducibility of histological specimens is affected by color intensity variation of stained images due to several factors: *i*) the concentration, the pH and the degree of deterioration of the staining dye, *ii*) the staining exposure time, *iii*) the ability of histology technician, *iv*) the illumination condition and *v*) the scanner performances and specifications [6], [7]. The solution to this problem is the standardization of color appearance in digital pathology. In this context, emerges the stain normalization technique which adapts the color intensity of histological images to a reference template, chosen by an expert pathologist, which presents optimal characteristics in terms of stain distribution, illuminant conditions and color intensity saturation. The stain normalization strategy is based on the stain separation method described in detail in Chapter 4.

In routine histology, the most commonly used staining procedure is based on Hematoxylin and Eosin (H&E) stain. Hematoxylin is a basic and positively charged dye which stains acid molecules such as nuclei of cells with a blue or purple color, while eosin is an anionic and acidic dye which stains with a pinkish color cytoplasmic structures such as cell stroma [8]. In H&E stained digital images, the

stain separation method based on the mathematical technique called color deconvolution, as described in Chapter 4, decomposes the histological image into two separate channels: the first is the hematoxylin channel where cell nuclei are highlighted while the second one is the eosin channel where cell stroma is isolated [9]. The concept behind the stain normalization method is the adaptive color transfer between source image and target image. The source image is the original scanned image which presents a non-optimal stain color appearance; on the other hand, the target image is a reference image with uniform stain color distribution, selected by an expert pathologist. The stain normalization method tries to adapt the hematoxylin color of the source image to the color of the hematoxylin stained structures in the target image; the same process is repeated for the eosin channel.

The aim of the stain normalization strategy is twofold: firstly the standardization of stain color appearance in digital pathology could improve the pathologist's work in the diagnosis of biological diseases; secondly this method could be used as a pre-processing step for subsequent computer-assisted diagnosis (CAD) systems for accurate cellular structure segmentation and classification based on deep learning framework [1], [2], [10].

Several state-of-the-art methods have been proposed with the purpose of normalization of histological images stained with H&E, which is the most widely used staining technique. According to the strategy employed, these methods can be classified into three main classes: *i*) color deconvolution-based, *ii*) histogram transformation-based and *iii*) iterative optimization-based methods. Most of these methods exploited the decomposed matrices  $W$  and  $H$ , which are called stain color appearance matrix and stain density map respectively, whose mathematical definition is reported in Chapter 4.

In this chapter a novel image processing approach of stain separation and normalization for H&E stained histological images is proposed. The algorithm called Stain Color Adaptive Normalization (SCAN), standardizes the color of a histological image with respect to the stain color of a target image with a consequent increase of the contrast between tissue regions and background. In addition, the cellular structures of the original image are not distorted and preserved. In the following, a detailed description of the SCAN algorithm is reported.

### 5.1.1 Color deconvolution-based methods

Ruifrok et al. [9] were the first authors to propose a stain color image approach in histology. Their method is based on the assumption that histological images present an average stain color for hematoxylin and eosin with a very small variation from one image to another. Hence, they proposed a fixed value of matrix  $W$  and the matrix  $H$  was computed by inverting the following linear system:

$$V = -\log_{10}\left(\frac{I}{I_0}\right) = H \cdot W \quad (\text{Eq. 5.1})$$

as discussed in Chapter 4. This was the first attempt to separate H&E stains using the mathematical basis of color deconvolution technique. This method is a non-adaptive method and obtains low performances when the source image to be normalized presents a different stain color from the average RGB triplet set by the authors. In routine histology the stain color variation is a condition that is frequently repeated and represents a problem for this method.

Macenko et al. [11] employed the principal component analysis (PCA) [12] in optical density (OD) space, where the relation between absorbance and stain concentration is linear. The idea is to project the OD triplets onto the plane identified by the singular value decomposition (SVD) directions corresponding to the two largest eigenvalues. The directions with angles (with respect to the first SVD direction) closer to robust extreme percentiles (e.g. 1<sup>st</sup> and 99<sup>th</sup> percentile) of whole angle distribution are retained and the final stain vectors are obtained converting back to the OD plane. The first advantage of this method is the adaptive estimation of matrix  $W$ , which is computed by exploiting the SVD-geodesic method in OD space, instead of determining stain vectors experimentally. The second advantage with respect to Ruifrok's method [9] is that all the OD triplets with an intensity value less than a fixed threshold  $\beta$  (in the original manuscript  $\beta=0.15$ ) were removed; thus, with this gimmick, all white values corresponding to unstained regions in the histological image (e.g. gland lumen, background, etc.) are removed from the estimation of matrix  $W$ . In practice, this choice can be too selective and in presence of non-uniform illuminant conditions, the stain estimation may be negatively affected.

Khan et al. [13] employed a stain color descriptor (SCD) based on a supervised method and exploited the advantages of color deconvolution technique using a nonlinear mapping between source and target image. Their work consists of a classification framework based on a training set of quantized histograms. They introduced a SCD useful to estimate stain concentrations and relevance vector machine (RVM) [14] is used to classify stain color. The normalization of the source image with respect to target image is performed using spline-based nonlinear functions. The big limitation of this approach is the high computational time to perform image normalization.

### 5.1.2 Histogram transformation-based methods

Reinhard et al. [15] proposed a method of color transfer based on the computation of image statistics (i.e. mean and standard deviation) of the source image and target image in  $l\alpha\beta$  color space [16]. This method, differently from the previous strategies based on color devolution, performs a stain normalization without the initial step of stain estimation (i.e. computation of matrix  $W$ ). The color correction of the source image was performed using the following expression for each of the three channels in  $l\alpha\beta$  color space.

$$I_j^{norm} = (I_j^{source} - \mu_j^{source}) \cdot \left( \frac{\sigma_j^{target}}{\sigma_j^{source}} \right) + \mu_j^{target}, \quad j = 1, 2, 3 \quad (\text{Eq. 5.2})$$

where the notations  $\mu$  and  $\sigma$  denote mean and standard deviation of each channel, respectively. Finally, the normalized image in the original RGB space, was obtained by performing a  $\lambda\alpha\beta$ -RGB conversion of each channel computed with Eq. 5.2. The limitations of this method are that the contrast after the normalization is not increased and stain colors in  $\lambda\alpha\beta$  color space are not optimally separated as in OD space. In addition, the color matching, based on the computation of basic statistics such as a single mean and standard deviation, performs correctly only when the pixel distribution is unimodal, but this is not true in histological image analysis due to stain color variability. As a result, the unstained regions of the source image (i.e. where the absorbance is zero) assume an incorrect stain color in the normalized image and a low-pass filter effect is caused.

### 5.1.3 Iterative optimization-based methods

Rabinovich et al. [17] compared two unsupervised methods for stain normalization. The first is based on independent component analysis (ICA) [18] while the second strategy employed the algorithm of non-negative matrix factorization (NMF) [19]. ICA is a computational strategy used to separate multivariate data into independent components assuming that there is a mutual statistical independence of signal sources (i.e. stain vectors for this purpose). In histological image analysis, this means that each dye binds tissue components irrespective of all the other dyes, which is wrong. On the other hand, NMF is a more interesting strategy for this purpose, since it is a factorization technique which decomposes the original OD image into the product of two non-negative matrices (i.e. the stain color appearance and the stain density map). In the work of Rabinovich et al. [17], NMF strategy obtained better performances than ICA, in terms of relative percent error between the estimate and the ground truth, applied to histological images derived from human biopsies. The NMF strategy consists of the minimization of the following error norm in terms of matrices  $W$  and  $H$ , by solving a constrained problem:

$$f(W, H) = \frac{1}{2} \|V - H \cdot W\|_F^2 \quad \text{such that} \quad W, H \geq 0 \quad (\text{Eq. 5.3})$$

The definition of the matrices involved in this expression are the same reported in Chapter 4, while the subscript  $F$  denotes the Frobenius error norm between the OD image  $V$  and its low rank approximation (i.e. the product  $H \cdot W$ ). The advantage of this method is to find a solution of a constrained problem: in fact, the non-negative constraint is useful to obtain physically interpreted values for stain color and stain concentration which cannot assume negative values (i.e. negative color or concentration makes no physical sense). The cost function in Eq. 5.3 is not convex in both input variables but only in  $W$  or  $H$ , separately. Therefore, closed-form solutions to find a global minimum are not available, thus numerical iterative

optimization algorithms were proposed to find local minima. The most used are the alternating least squares and the multiplicative update algorithm [20]. In these algorithms, the initialization of unknown matrices is performed by using non-negative random values for matrices  $W$  and  $H$ , and after a certain number of iterations the convergence is proved [19].

Vahadane et al. [21], in their work added a sparseness term in the NMF cost function reported in Eq. 5.3 for the regularization of the stain density map  $H$ , to reduce the solution space. For this reason, this method is called sparse non-negative matrix factorization (SNMF). In addition, they required, in their optimization method, stain vectors in the stain color appearance matrix  $W$  had unit Euclidean norm, in order to obtain a correct balance for each separate stain [9]. The principal advantage of this method is that using the new cost function, the performances of stain normalization and the correctness of the estimate of stain colors are superior than the standard NMF algorithm proposed by Rabinovich et al [17]. In addition, cellular structures of the original image are preserved and not distorted after the normalization. Nevertheless, the problem of local minima remains, and numerical solutions were forced for this purpose.

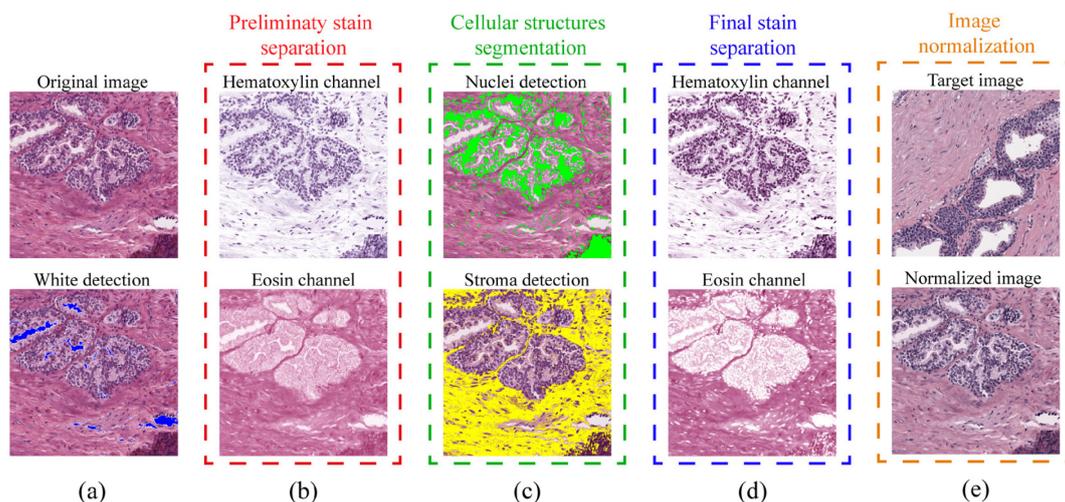
Finally, Zheng et al. [22] introduced a novel adaptive color deconvolution (ACD) strategy for stain normalization of H&E stained slides. An integrated optimization method was employed to find optimal stain vectors by minimizing an objective function which is a linear combination of three terms: the first term minimizes the residual of the stain separation, the second controls the ratio of hematoxylin and eosin proportion in the image and finally the third one controls the overall energy of stains. The normalization is obtained by recombining the solution matrices which minimize the objective function with the target stain color appearance matrix. The shortcoming of this method is the estimate of stain separation matrices which is not robust; indeed, the quantitative metric employed to validate their work is not robust, as argued in the discussion section of this chapter.

## 5.2 Materials and Methods

In this section, a detailed description of the SCAN (Stain Color Adaptive Normalization) algorithm is reported. The proposed technique is a fully automated method which performs stain separation and normalization of histological and histopathological H&E stained tissue slides. This method from one side belongs to color deconvolution-based approaches but from another side is an iterative adaptive refinement which tries to estimate the optimal color appearance and for this reason can also be referred to iterative optimization-based approaches. The flowchart of the proposed strategy is reported in Figure 5.1.

The method is composed of four main modules: *i*) white detection and preliminary stain separation, *ii*) cellular structures segmentation, *iii*) final stain separation and *iv*) image normalization. The proposed strategy is an adaptive refinement of the SVD-geodesic method proposed by Macenko et al. [11] with an

additional pre-processing step useful for unstained regions. This refinement consists of an accurate detection of hematoxylin nuclei and eosin stroma regions in order to compute stain vectors; in addition, a novel strategy for stain density map estimation is proposed. Finally, the input matrices  $W$  and  $H$  of the original image are adapted to target information and the OD image is converted back in RGB space with a result of stain normalization.



**Fig. 5.1** Graphical representation of SCAN algorithm: (a) original image and white mask (in blue); (b) preliminary stain separation based on SVD-geodesic approach; (c) cellular structures segmentation (cell nuclei in green and stroma in yellow); (d) final stain separation; (e) stain normalization of the original image with respect to target image.

The automated method was tested on a multiscale and multi-tissue image dataset which consists of 270 optically scanned H&E stained histological images acquired with Aperio Scanscope (Leica Biosystems). Each digital slide is a RGB image with a size of  $1024 \times 2048$  pixels. The same input size for all images was only used to standardize the procedure, but the image input dimension is not relevant for the processing. The histological sections were extracted from five different tissues (adrenal gland, breast, colon, liver and prostate), scanned at three magnifications ( $10\times$ ,  $20\times$  and  $40\times$ ) and stained with H&E staining method, as described in Chapter 4. The overall dataset distribution is listed in Table 5.1. In addition, five template histological slides, one for each tissue were selected by an expert pathologist as target images. The target image is a reference image with an optimal and reproducible staining distribution (i.e. with a quasi-unimodal distribution of concentration for each stain), with a homogeneous illuminant condition and which has stain intensities as expected by the clinician.

**Table 5.1** Image dataset composition.

Tissue	Magnifications			# Images
Adrenal gland	$10\times$	$20\times$		50
Breast	$10\times$	$20\times$		50
Colon		$20\times$	$40\times$	40
Liver	$10\times$	$20\times$	$40\times$	60
Prostate	$10\times$	$20\times$		70

### 5.2.1 Preliminary stain separation

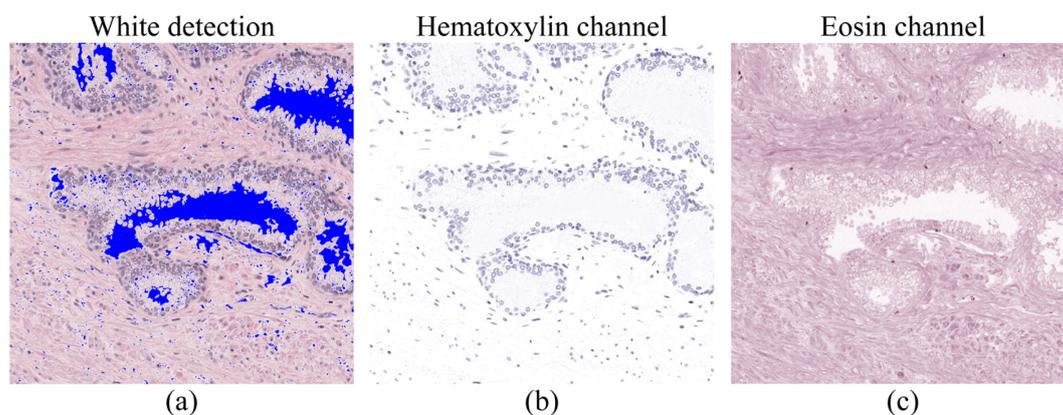
The first step of the proposed algorithm is the detection of unstained regions (e.g. background, gland lumen, etc.) which assume white colors in the original RGB image since the light has passed through without absorption. A white detection strategy is proposed since a simple thresholding used in the work of Macenko et al. [11] is not enough in the presence of non-uniform illuminant conditions or image artifacts. Therefore, the input image is convolved with a 2D Gabor function, expressed as the product of a Gaussian and a cosine function:

$$g_{\lambda,\theta,\varphi,\sigma,\gamma}(x,y) = e^{-\left(\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)} \cdot \cos\left(2\pi \frac{x'}{\lambda} + \varphi\right) \quad (\text{Eq. 5.4})$$

$$x' = x \cos \theta + y \sin \theta; \quad y' = -x \sin \theta + y \cos \theta$$

where  $\gamma$  is the spatial aspect ratio which denotes the kernel ellipticity,  $\sigma$  is the standard deviation of the Gaussian term (i.e. the kernel width),  $\lambda$  and  $\varphi$  are the wavelength and the phase offset of the cosine term respectively and  $\theta$  denotes the normal orientation to the parallel stripes of the Gabor function [23]. In this work the following values for Gabor parameters were set:  $\gamma = 1.2$ ,  $\sigma = 1$ ,  $\lambda = 10$ ,  $\varphi = 0$  and 8 directions ( $\theta$ ) between 0 and 360 degrees were considered. Finally, each filtered image (one for each orientation) was segmented using a threshold of 90% of the image maximum intensity to obtain the background mask; all these masks were combined using a logical union. An example of white detection applied to histological image is reported in Figure 5.2 (a).

The detection of white regions is useful for the subsequent preliminary stain estimation based on the SVD-geodesic method proposed by Macenko et al. [11]. The stain color appearance matrix  $W$  was estimated using this strategy with an additional pre-processing step of removal from the analysis all white regions detected using the previous white detection. Then a color deconvolution was applied to separate hematoxylin and eosin channel, as shown in Figure 5.2 (b-c).



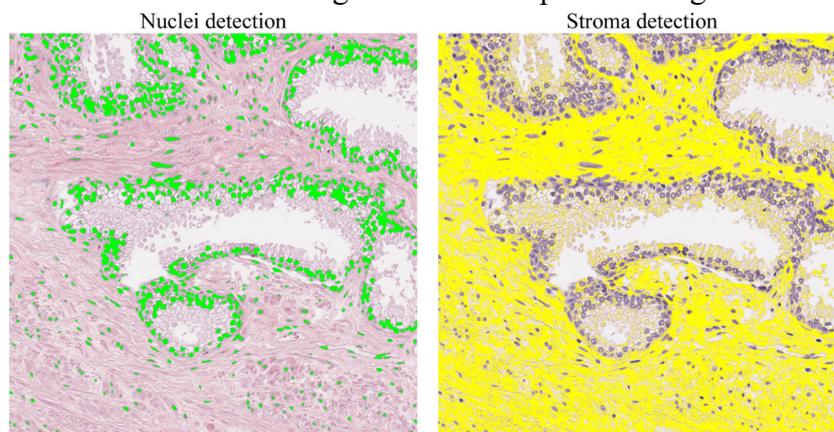
**Fig. 5.2** (a) White detection mask (in blue). Preliminary stain separation of hematoxylin channel (b) and eosin channel (c) using SVD-geodesic method.

## 5.2.2 Cellular structures segmentation

The following step of SCAN algorithm is based on cellular structures segmentation with the aim to estimate optimal stain color of hematoxylin stained nuclei and eosin stained stroma. Two different image processing techniques were applied for cell nuclei and stroma detection. Firstly, the hematoxylin density map was normalized by employing min-max scaling and filtered using a lowpass Wiener filter in order to smooth high concentration values. Wiener filter applied to image processing is an adaptive noise-removal 2D filtering which uses pixel neighborhoods to estimate the local image mean and variance [24]. Subsequently, an improved version of Multiscale Adaptive Nuclei Analysis (MANA) algorithm [25] was developed for nuclei segmentation. Firstly, the probability mass function (PMF) was estimated by normalizing the whole image histogram and its weighted variance  $\sigma^2$  was computed. Then, several thresholds between 0 and 256 with a step of 2 (i.e. 129 overall values) were set and the following energy function was computed for each threshold  $T$ :

$$E(T) = -\sigma^2 \cdot \log(\sigma^2) + p_0^2(T) \cdot \sigma_0^2(T) \cdot \log(\sigma_0^2(T)) + p_1^2(T) \cdot \sigma_1^2(T) \cdot \log(\sigma_1^2(T)) \quad (\text{Eq. 5.5})$$

where  $p_0$  is the cumulative distribution function for each threshold value,  $p_1$  is the probability of complementary event. Hence,  $p_0$  represents the probability of having background with low absorption, while  $p_1$  is related to nuclei distribution. The weighted variances of PMF for each threshold are denoted as  $\sigma_0^2$  and  $\sigma_1^2$  respectively. Finally, Wiener filtered image was segmented using an optimal threshold which minimized the energy function in Eq. 5.5. Subsequently, the hematoxylin density map was subtracted from the eosin density map and the difference, after fixing any negative values to zero, is fed to a fast k-means clustering algorithm which classified grayscale image intensity into 2 classes ( $k=2$ ) and stroma mask was defined as the cluster with highest mean intensity. The algorithm was initialized with pixels which had the lowest and the highest intensity in the image obtained as difference. The use of histogram intensity values instead of raw image data during the clustering process reduced the computational time. The result of cellular structures segmentation is reported in Figure 5.3.



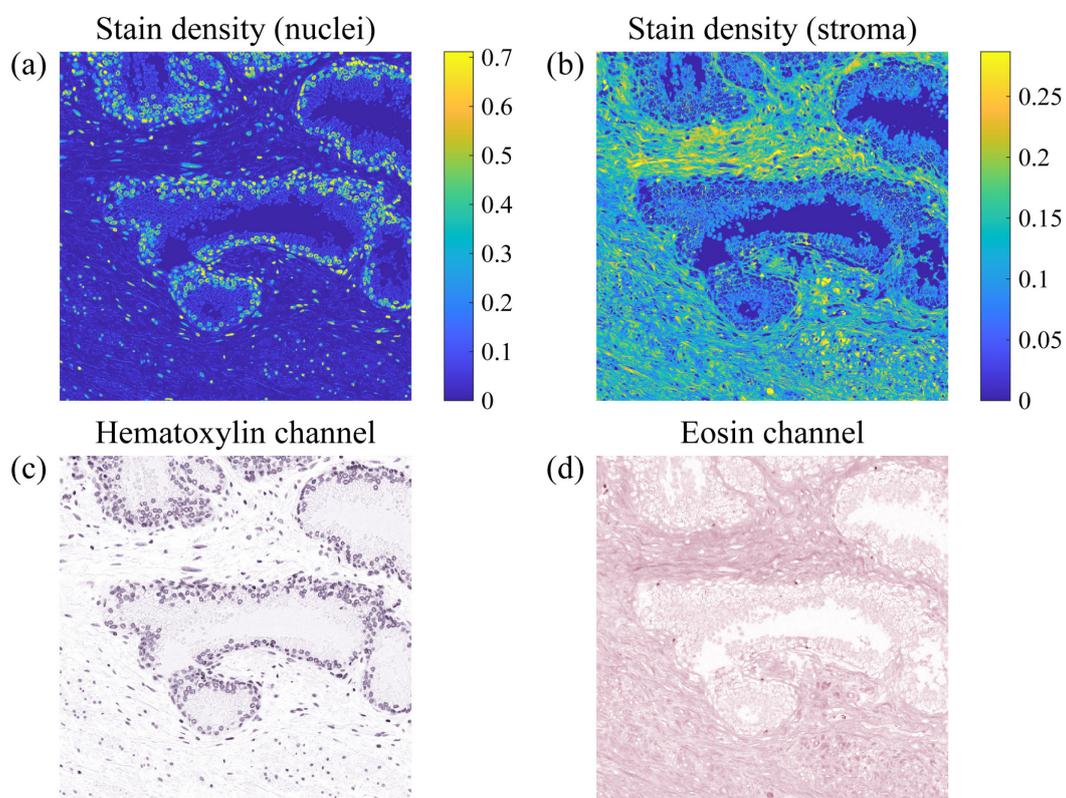
**Fig. 5.3** Cellular structures segmentation: hematoxylin nuclei detection (in green, on the left) and eosin stroma detection (in yellow, on the right).

### 5.2.3 Final stain separation

After having segmented nuclei and stroma regions, a median value of RGB color intensity was computed and converted into OD space, using Eq. 5.1. The final stain color appearance matrix  $W$  was estimated by employing an iterative algorithm which compared at each iteration the previous and the current values of stain vectors. When the Euclidean norm of the difference between two successive estimates of stain vectors was less than 0.10, then the iteration stopped. The new stain color appearance matrix  $W$  was normalized so that each row (i.e. stain vector) had unit Euclidean norm [9]. A new strategy for stain density map  $H$  was proposed in this thesis work. Several state-of-the-art methods [9], [11], [13] derived this matrix by solving the linear system with a Moore-Penrose pseudoinverse computation [26]:

$$H = ((W \cdot W^T)^{-1} \cdot W \cdot V^T)^T \quad (\text{Eq. 5.6})$$

The issue is that the linear system is unconstrained, hence  $H$ , using the numerical inversion, can assume negative values but this is impossible due to its physical meaning. This problem was solved by computing matrix  $H$  using its physical definition of concentration map in OD space. Therefore, the first column  $H(:,1)$  was computed as the median value of the ratio between hematoxylin triplets and first stain vector  $W(1, :)$ . The same process was carried out for the second column  $H(:,2)$  in order to obtain eosin concentration values [27].



**Fig. 5.4** Stain density maps using the new strategy for nuclei (a) and stroma (b) are shown in the first row. In the second row final stain separation is reported for 1<sup>st</sup> channel (c) and 2<sup>nd</sup> channel (d).

Finally, a color deconvolution was performed using new stain color appearance matrix  $W$  and stain density map  $H$ , as shown in Figure 5.4.

### 5.2.4 Stain normalization

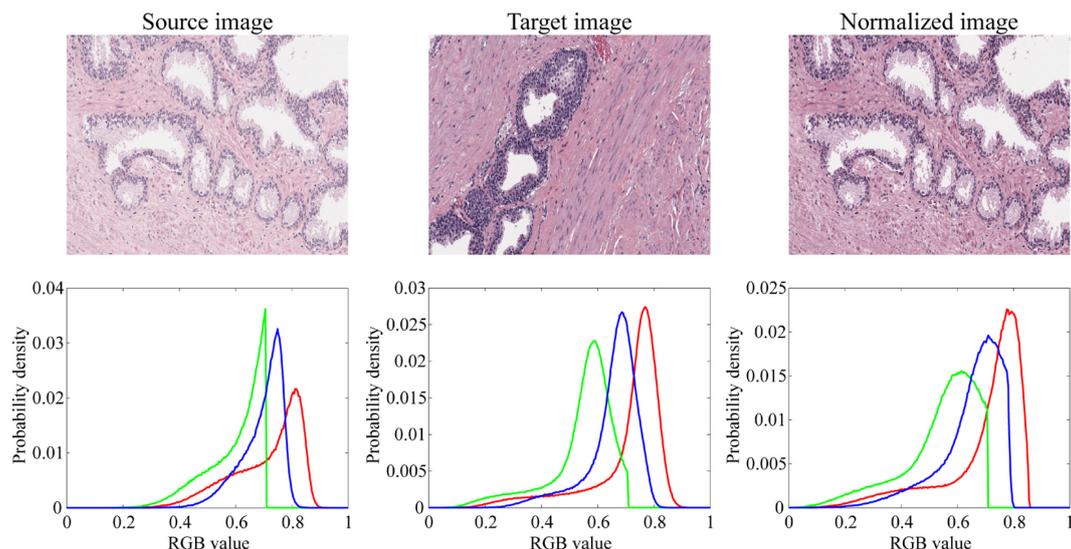
The last step of SCAN method is the stain normalization process. The computation of matrices  $W$  and  $H$ , as described in the previous sections, was repeated, in the same way, for an input image (also called source image) and a reference image (also called target image) chosen by an expert pathologist. The source image was normalized using the information extracted from the target image using the mathematical procedure described in the work of Vahadane et al. [21]. The stain density map of the source image  $H_s$  which contained the concentration values of each stain for all pixels, was normalized in order to obtain the same dynamic range of stain density map of the target image  $H_t$ . The dynamic range was estimated as the robust pseudo maximum (i.e. 99<sup>th</sup> percentile) of  $H_s$  and  $H_t$  distribution; the scaling procedure for the minimum is unnecessary since the lowest limit for stain density map is zero (with the new strategy proposed in this thesis work), which means no absorption. From a mathematical point of view, the definition of the scaled version of stain density map of the source image (defined as  $H_s^{norm}$ ) is:

$$H_s^{norm}(:,j) = H_s(:,j) \cdot \frac{H_t^{RM}(j)}{H_s^{RM}(j)}, \quad j = 1,2 \quad (\text{Eq. 5.7})$$

where the superscript  $RM$  is the robust maximum (i.e. 99<sup>th</sup> percentile) of each column (1<sup>st</sup> column: hematoxylin concentration; 2<sup>nd</sup> column: eosin concentration). Then,  $H_s^{norm}$  was combined with the stain color appearance matrix of the target image  $W_t$  instead of that of the source image, and by converting back in the original RGB color space, the final normalized image was obtained as:

$$I_s^{norm} = I_0 \cdot 10^{-(H_s^{norm} \cdot W_t)} \quad (\text{Eq. 5.8})$$

where the incident light intensity  $I_0$  was set to 255 for images with a dynamic range of 8 bits. The final result of stain normalization performed by SCAN algorithm is reported in Figure 5.5. After applying SCAN algorithm, the intensity distribution of the normalized image, for each of the 3 RGB channels, resembles more closely the color distribution of the target image; in addition, the cellular structures within the tissue of the source image are preserved and not distorted in the normalized image.



**Fig. 5.5** Stain normalization of a H&E stained tissue slide: source image (left), target image (center) and normalized image performed by SCAN algorithm (right). The digital RGB images are reported in the first row, while the normalized histogram intensity distributions (i.e. probability density functions) for each of the R, G, B color channels, are shown in the second row.

### 5.3 Experiments and Results

The SCAN algorithm was developed and implemented in MATLAB environment (MATLAB, The MathWorks, Inc., Natick, MA, USA). The proposed approach was validated by computing several performance measures and compared with several state-of-the-art methods [9], [11], [13], [15], [17], [21], [22] in the same research application, as described in Chapter 5.1. The ground truth stain vectors for the source and target image were generated with a manual procedure performed by an expert pathologist (with 20 years of experience). For each source and target image, the manual operator selected at least 100 points in tissue regions which belonged to nuclei and stroma portions, in order to find a robust ground truth stain color appearance estimation. The manual points were carefully annotated in the areas where presumably only one stain bound to cellular structures. For the validation procedure five optimal template histological slides, one for each tissue (adrenal gland, breast, colon, liver and prostate) were selected by the expert as target images. Then, the following relative Square Error (rSE) was employed:

$$rSE = \frac{\|(W - W_{GT})^T \cdot (W - W_{GT})\|}{\sqrt{\|W^T \cdot W\| \cdot \|W_{GT}^T \cdot W_{GT}\|}} \quad (\text{Eq. 5.9})$$

where  $\|\cdot\|$  is the matrix trace,  $W$  represents the stain color appearance matrix estimated by SCAN algorithm and  $W_{GT}$  is the ground truth stain color matrix computed on the tissue areas manually identified by the expert pathologist [21]. If the ground truth matrix and the estimate are equal, this error is zero, which means, of course, perfect estimation. According to the mathematical relation between matrix trace and Frobenius norm, this error can also be expressed in the following percentage form:

$$rSE(\%) = \frac{\|W - W_{GT}\|_F^2}{\|W\|_F \cdot \|W_{GT}\|_F} \cdot 100 \quad (\text{Eq. 5.10})$$

This error quantitatively assesses the capacity of an automated method to estimate correct stain vectors. If only the first/second row of  $W$  matrix was considered, the Frobenius norm reduced to vector Euclidean norm and the estimation of hematoxylin/eosin stain color could be evaluated individually. More specifically, if unit norm was applied to  $W$  matrix estimation, the denominator of rSE was equal to two. The performance measures were computed both for the validation of stain separation and stain normalization process. To the best of our knowledge, this was the first published work which attempted to quantitatively assess not only the normalization technique but also the previous stain estimation strategy, which can be considered a necessary step for a correct color deconvolution of histological channels. In addition, in this thesis work we tried to compare other performance measures in order to find the most robust one.

### 5.3.1 Stain separation performance

The first metric to validate stain separation performance was the rSE reported in Eq. 5.10. For this purpose,  $W$  was the stain color appearance matrix estimated by SCAN algorithm at the end of final stain separation and  $W_{GT}$  is the ground truth matrix for the source image (i.e. true stain colors manually selected by the expert). The rSE metric was computed for the whole matrix to evaluate global performances (*rSE glob*) and by considering the first/second column of matrix  $W$  to evaluate the error for hematoxylin (*rSE Hem*) and eosin (*rSE Eos*) individually. The rSE for stain separation is denoted in the following figures as  $rSE^{sep}(\%)$ . The results of rSE for the validation of stain separation methods were reported in Figure 5.6 for performance comparison at different magnifications and in Figure 5.7 for multi-tissue evaluation. The methods proposed by Ruifrok et al. [9] and Zheng et al. [22] were excluded from the bar graphs due to their worst stain separation performances with respect to the other methods, in terms of rSE.

Other measures to quantitatively assess the performances of stain separation are proposed in this thesis. The first metric is the Frobenius norm of the error between the original source image and the reconstructed image after the color deconvolution technique, defined as follows:

$$\|I - I_0 \cdot 10^{-(H \cdot W)}\|_F \quad (\text{Eq. 5.11})$$

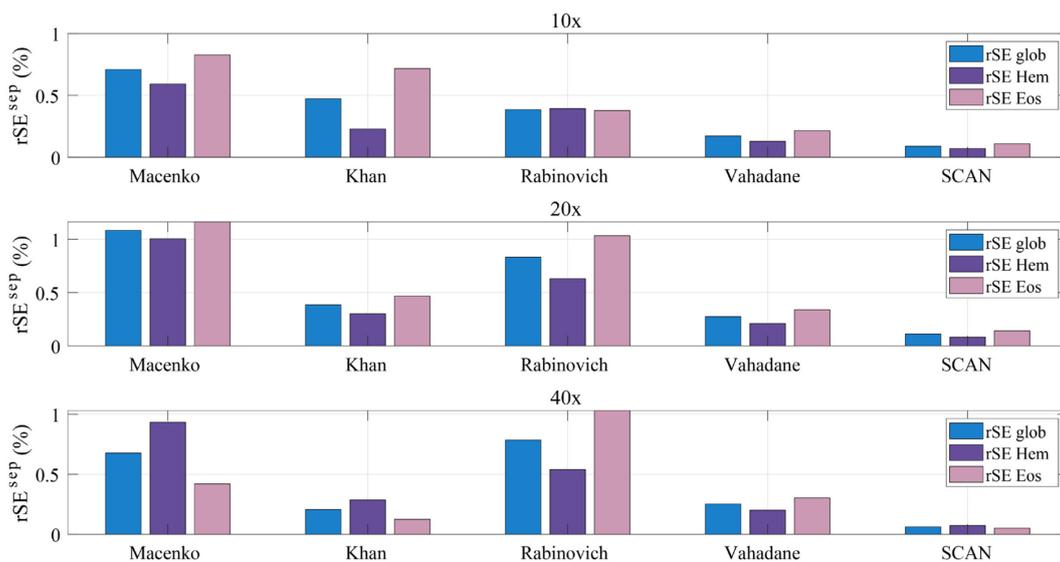
Moreover, the histogram-based errors between the color intensity distribution of the original RGB image and the deconvolved channels, after the color deconvolution were computed. To quantify this error two statistics were defined: mode (i.e. the value at which the histogram takes its maximum value) and skewness (i.e. the normalized third central moment). The Euclidean norm of the error between the mode of each R, G, B color distribution of the original and decomposed image was computed and the same was done for the skewness parameter. The perceptual

contrast in  $l\alpha\beta$  color space was computed for both channels (i.e. hematoxylin and eosin) by using the following definition:

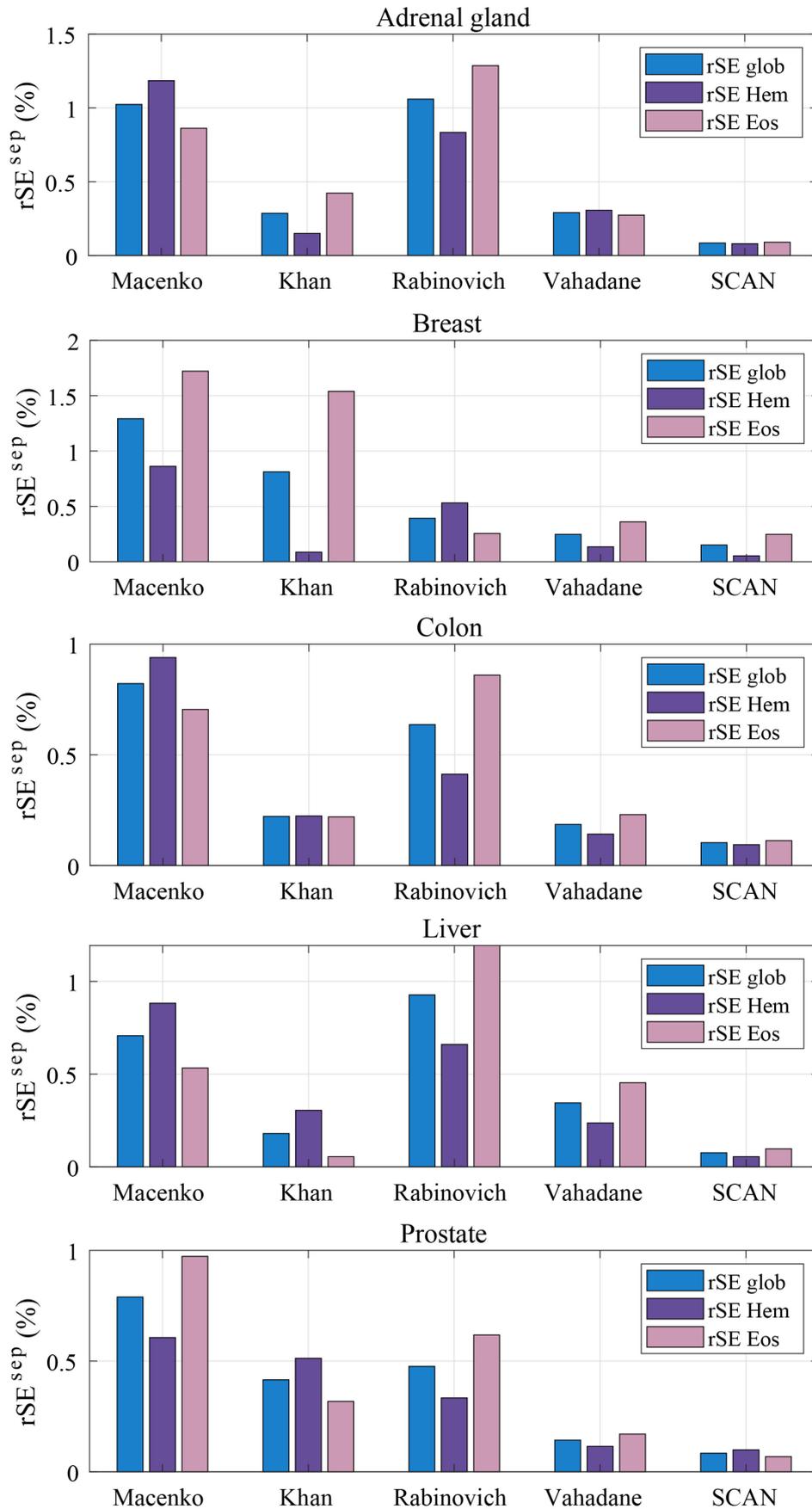
$$C_j = \sqrt{(\Delta l_j)^2 + (\Delta \alpha_j)^2 + (\Delta \beta_j)^2}, \quad j = 1, 2 \quad (\text{Eq. 5.12})$$

$$\Delta l_j = \langle l_j \rangle^{obj} - \langle l_j \rangle^{bg}; \quad \Delta \alpha_j = \langle \alpha_j \rangle^{obj} - \langle \alpha_j \rangle^{bg}; \quad \Delta \beta_j = \langle \beta_j \rangle^{obj} - \langle \beta_j \rangle^{bg}$$

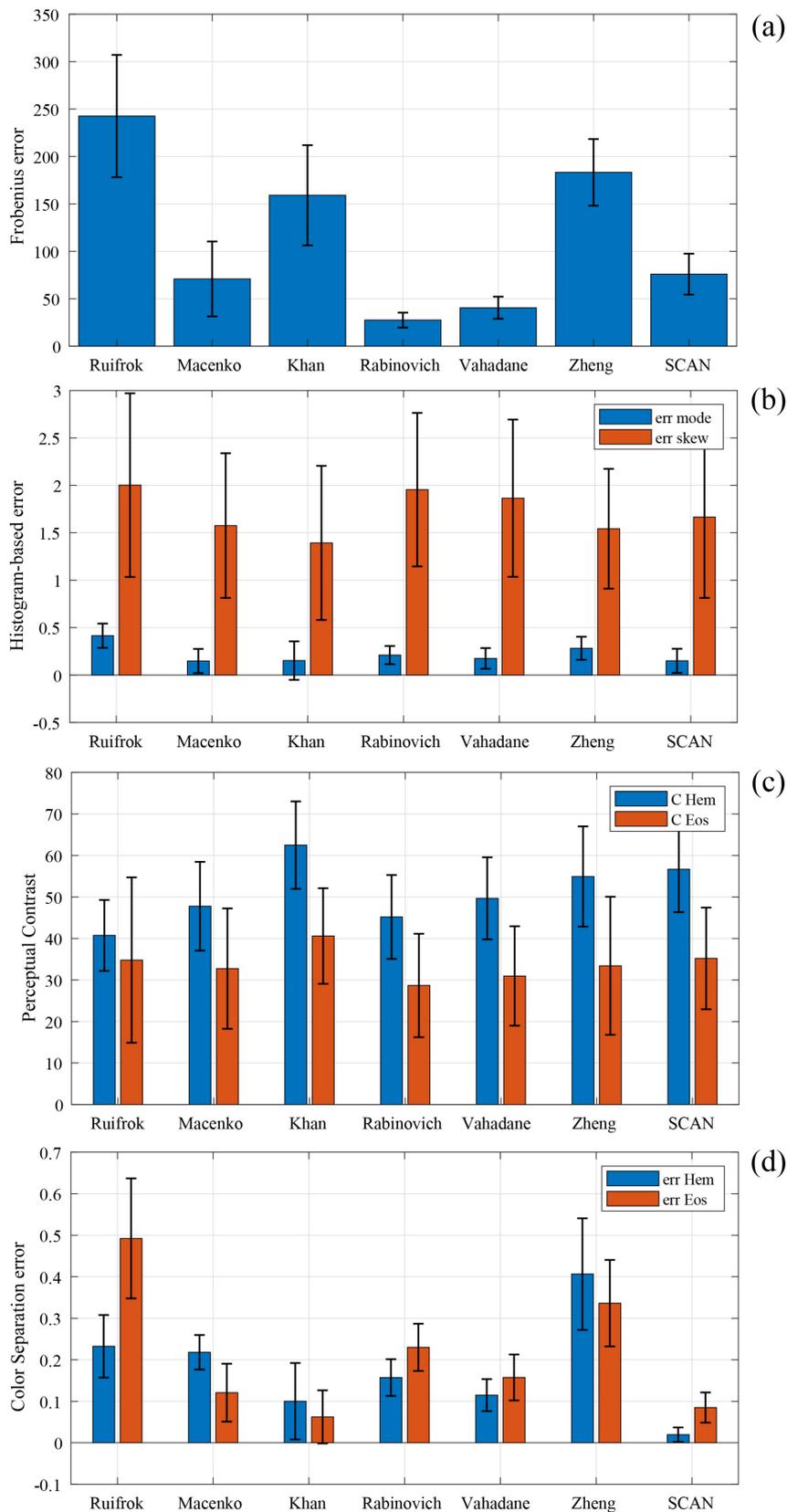
where  $\langle \cdot \rangle$  denotes the mean value and the superscripts *obj* and *bg* indicate the object (i.e. tissue) and background (i.e. unstained tissue) regions, respectively. The contrast was defined in  $l\alpha\beta$  color space since Euclidean distances between colors in this perceptually uniform space are related to the color difference perceived by a human observer [28]. Finally, the color separation error, defined as the mean difference of color intensity between the original image and the deconvolved channels (i.e. hematoxylin and eosin), was computed. In Figure 5.8, the results related to the previously described quantitative metrics are reported for the entire dataset. For all metrics reported in Figure 5.8, a lower value may be expected for correct stain separation process except for the perceptual contrast, which may be high in both channels for a better stain separation.



**Fig. 5.6** Quantitative comparison at different magnifications, between SCAN algorithm and other state-of-the-art methods in terms of global rSE (in blue), hematoxylin rSE (in purple) and eosin rSE (in pink) for stain separation.



**Fig. 5.7** Quantitative multi-tissue comparison between SCAN algorithm and other state-of-the-art methods in terms of global rSE (in blue), hematoxylin rSE (in purple) and eosin rSE (in pink) for stain separation.



**Fig. 5.8** Other quantitative measures of stain separation for the comparison between SCAN algorithm and other state-of-the-art methods. (a) Frobenius norm of the error between the original and reconstructed image. (b) Histogram-based errors: error for mode (in blue) and for skewness (in red). (c) Perceptual contrast in hematoxylin (in blue) and eosin (in red) channels. (d) Color separation error for hematoxylin (in blue) and eosin (in red). All bar graphs are reported with mean value and symmetrical error bars for standard deviation.

### 5.3.2 Stain normalization performance

The stain normalization process was evaluated by using the same rSE definition reported in Eq. 5.10, changing the roles of the matrices involved. In the rSE definition for stain normalization,  $W$  contained the stain colors computed on the manual points in the normalized image which corresponded to the manual points annotated by the expert pathologist in source image, since the structures were not distorted by SCAN method after the normalization. On the other hand,  $W_{GT}$  was the ground truth matrix for the target image (i.e. true stain colors manually selected by the expert on the target image). The rSE for stain normalization is denoted in the following figures as  $rSE^{norm}(\%)$ . As discussed in the previous section, the rSE can be computed on the entire stain color appearance matrix ( $rSE_{glob}$ ) or retaining only the first/second row, in order to compute the  $rSE_{Hem}/rSE_{Eos}$  respectively. The lower the error, the better the performance of the stain normalization method, since the normalized image assumes a stain color similar to that of target image (which is the aim of the standardization). The results for stain normalization are reported in Figure 5.9 for multiscale evaluation and in Figure 5.10 for multi-tissue comparison. The method proposed by Ruifrok et al. [9] was excluded from the bar graphs due to its worst stain separation performances.

In the last few years, several measures for the evaluation of stain normalization have been proposed. The normalized median intensity (NMI) was used to quantitatively evaluate to constancy of the stain normalization. Its mathematical definition, reported in the work of Zheng et al. [22] is:

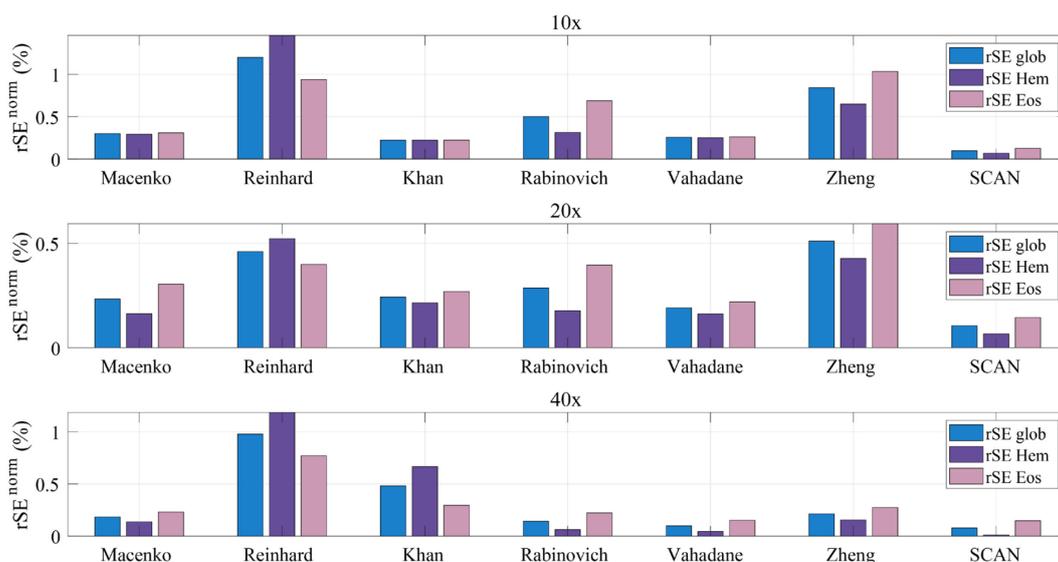
$$NMI = \frac{Med_{i \in I}(u_i)}{P_{95}(u_i)} \quad (\text{Eq. 5.13})$$

where  $I$  is the image after the normalization process and  $u_i$  denotes the mean value of the RGB triplet related to the  $i$ -th pixel. In the numerator there is a median value and, in the denominator, the 95<sup>th</sup> percentile. Typically, this value is not significant and the standard deviation (NMI SD) and the coefficient of variation (NMI CV) of the NMI values are computed. The lower the NMI SD and NMI CV, the more consistent the normalization process [22]. Structural similarity (SSIM) and quaternion SSIM (QSSIM) indices were proposed to evaluate the degradation of structural information. SSIM for a RGB image was computed as the mean of the SSIM of each grayscale R, G, B channel and its mathematical definition is derived in the work of Wang et al. [29]. QSSIM is a vectorial expansion of SSIM based on the quaternion theory, which tried to solve the problem of color cross correlation terms. QSSIM index measures the combination of image degradation in luminance and chrominance [30]. The histogram-based errors between the color intensity distribution of the normalized and target image were computed. To quantify this error two statistics were defined in the same way of stain separation: mode and skewness. The Euclidean norm of the error between the mode of each R, G, B color distribution of the normalized and target image was computed and the same was done for the skewness parameter. Finally, using the same definition reported in Eq. 5.12, the perceptual contrast between normalized tissue and background was

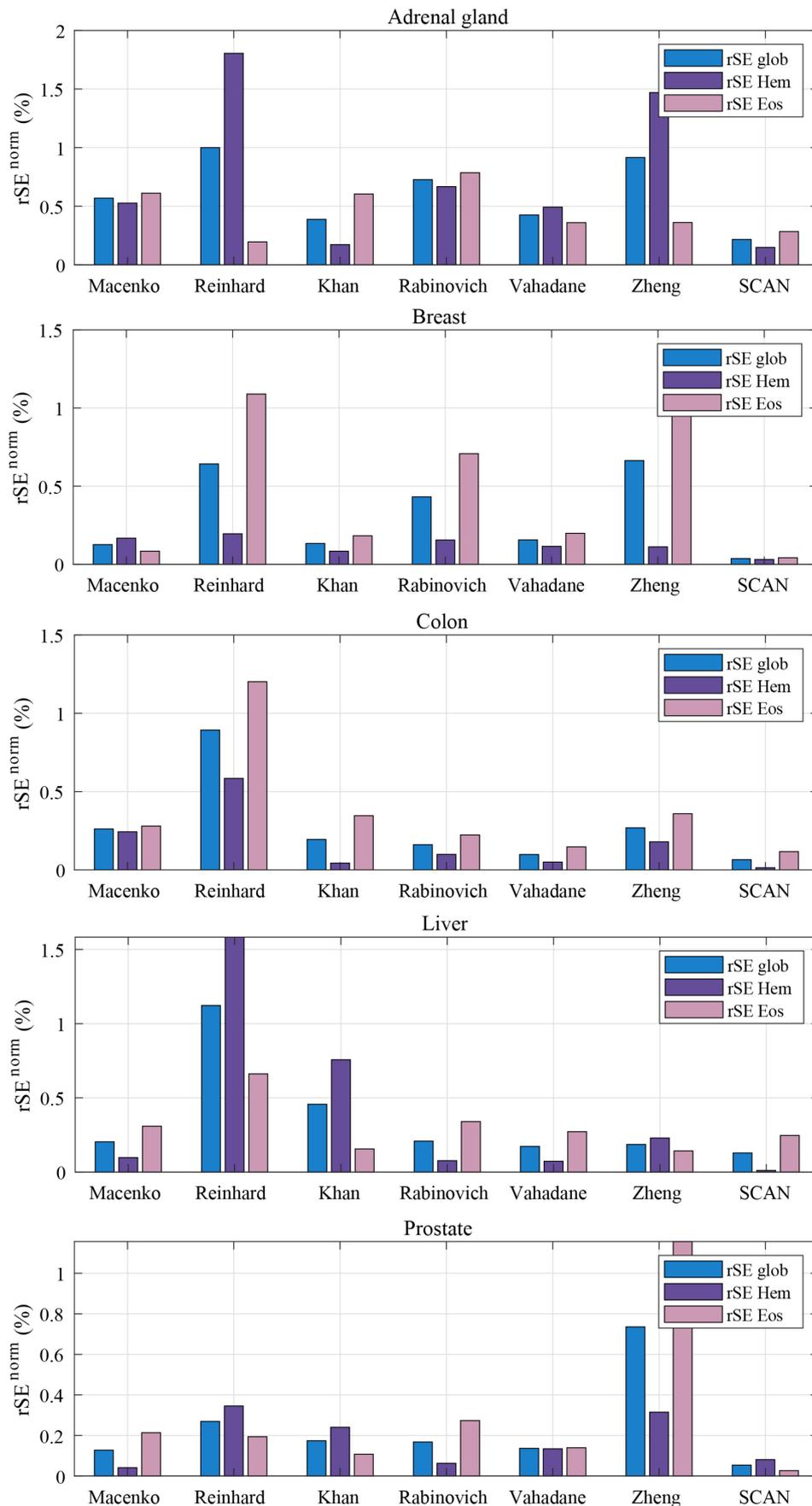
evaluated. The results are reported in Figure 5.11, where a low value of NMI SD, NMI CV and histogram-based errors may be expected for correct stain normalization process, while for the SSIM, QSSIM index and perceptual contrast, a high value corresponds to a better stain normalization.

In addition to quantitative performance measures, a qualitative visual comparison was reported in Figure 5.12, between three state-of-the-art methods about stain normalization and the SCAN algorithm, with respect to the same target image for each tissue. The tree methods that obtained the lowest  $rSE^{norm}(\%)$  were chosen for the comparison (i.e. Macenko et al. [11], Khan et al. [13] and Vahadane et al. [21]).

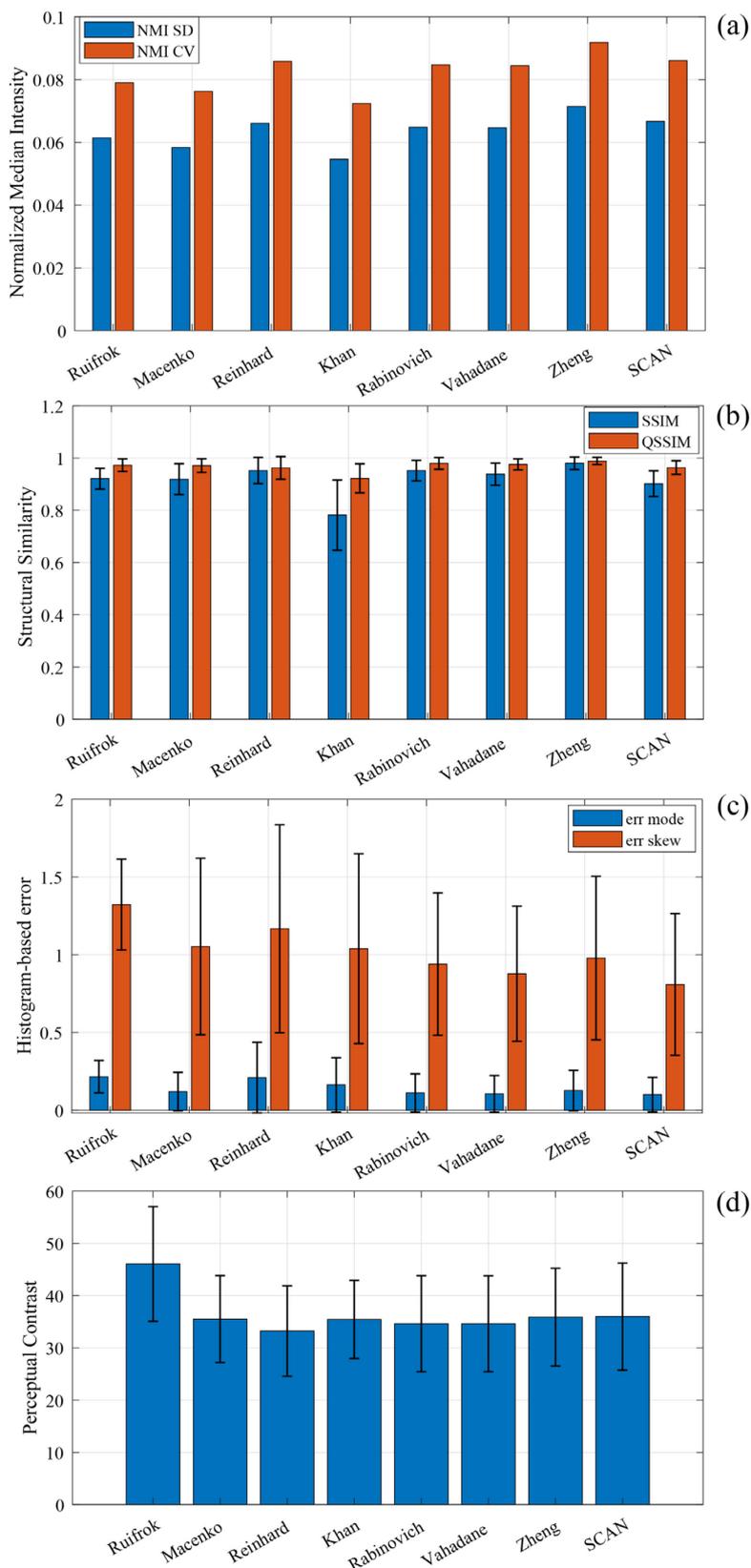
Another most important aspect for the evaluation of an automated algorithm for stain normalization is the computational time. In Figure 5.13, a joint plot of  $rSE^{norm}(\%)$  vs average running time to process a single histological image, expressed in seconds, is reported for several state-of-the-art methods. SCAN algorithm employed around 4.5 seconds to estimate source and target stain vectors and to perform stain normalization of the source image with respect to the target. The processing was performed on a workstation with 16 GB of RAM, 2.5 GHz quad-core CPU and 64-bit version of Windows.



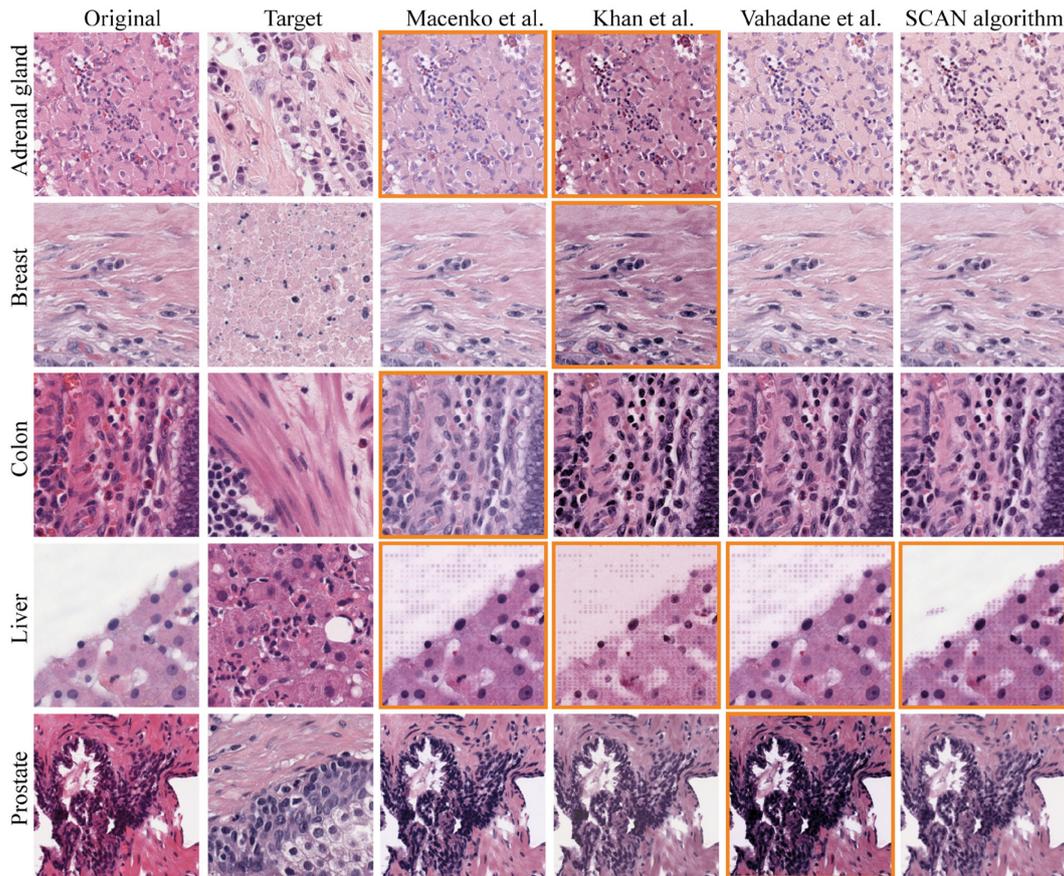
**Fig. 5.9** Quantitative comparison at different magnification, between SCAN algorithm and other state-of-the-art methods in terms of global rSE (in blue), hematoxylin rSE (in purple) and eosin rSE (in pink) for stain normalization.



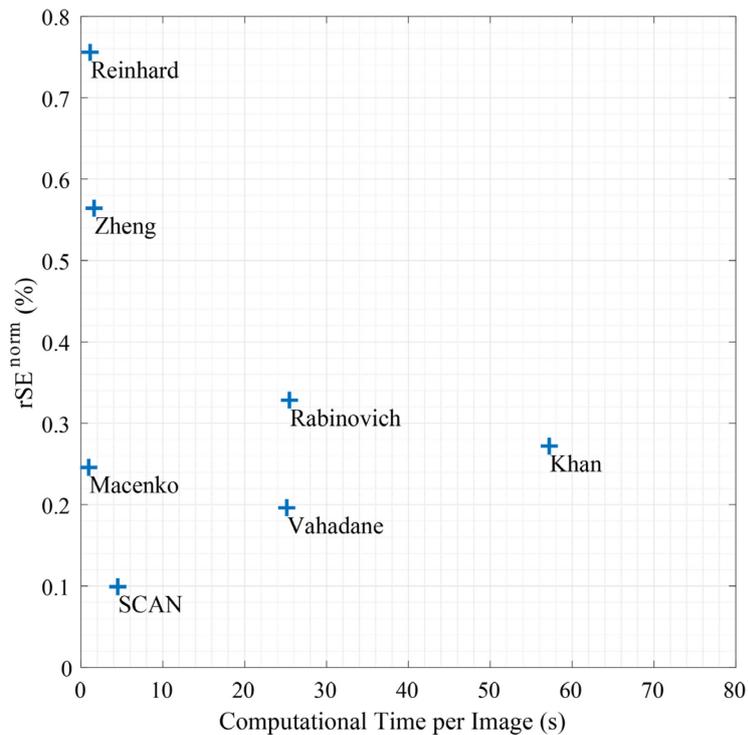
**Fig. 5.10** Quantitative multi-tissue comparison between SCAN algorithm and other state-of-the-art methods in terms of global rSE (in blue), hematoxylin rSE (in purple) and eosin rSE (in pink) for stain normalization.



**Fig. 5.11** Other quantitative measures of stain normalization for the comparison between SCAN algorithm and other state-of-the-art methods. (a) Standard deviation (in blue) and coefficient of variation (in red) of NMI. (b) SSIM (in blue) and QSSIM (in red) index. (c) Histogram-based errors: error for mode (in blue) and for skewness (in red). (d) Perceptual contrast between normalized tissue and unstained regions. All bar graphs, except for the NMI (a), are reported with mean value and symmetrical error bars for standard deviation.



**Fig. 5.12** Visual performance comparison between the proposed SCAN algorithm and other state-of-the-art methods. First column shows the original images while the second column displays the target images. Normalization methods are reported from the third to the sixth column and SCAN algorithm is displayed in the last one. Orange boxes denote images with visual artifacts.



**Fig. 5.13** Joint plot of rSE for stain normalization (y-axis) and average running time (x-axis) to normalize a single RGB histological image, for different automated methods.

## 5.4 Discussion and Conclusion

A novel fully automated stain separation and normalization approach for H&E stained histological slides was developed. The proposed algorithm, named SCAN, employed segmentation and clustering strategies for cellular structures detection. To the best of our knowledge, SCAN is the first method which employs cellular structures segmentation to estimate true stain colors of H&E for stain separation and successive stain normalization. In addition, a novel strategy for the computation of stain density matrix is proposed in this thesis work, with the aim of retaining the physical meaning of concentration map.

The most commonly used staining method in routine histology is the histochemical H&E, which is a bichrome staining and allows the constituents of a cell (i.e. nuclei and stroma) to be differently highlighted. By observing the histological slide, the pathologist provides the diagnosis for the specific patient. However, diagnosis based on histological images is often influenced by the staining quality of the histological preparation. In cases where the quality of the staining is not satisfactory for the pathologist, an additional portion of the tissue will be sectioned and stained with a consequent loss of time and financial resources. The reproducibility of histological specimens is affected by color intensity variation of stained images due to several factors such as the staining concentration and exposure time, the ability of histology technician, the non-uniform slide thickness, and the scanner specifications. The solution to this problem is the standardization of color appearance in digital pathology. In this context, emerges a stain normalization method which adapts the color intensity of histological images to a reference template, chosen by an expert pathologist, which presents optimal characteristics in terms of stain distribution, illuminant conditions and color intensity saturation. In fact, after applying SCAN algorithm, the intensity distribution of the normalized image, for each of the 3 RGB channels, resembles more closely the color distribution of the target image, as shown in Figure 5.5.

In this chapter, stain separation and normalization strategies are proposed with the aim of improving the performances with respect to other state-of-the-art approaches and preserving the unstained structures such as the lumen and the background. The proposed method was assessed on a multi-tissue (i.e. five different tissues were analyzed: adrenal gland, breast, colon, liver and prostate) and multiscale (i.e. at different magnifications) dataset and obtained better performance measures than other published methods. In literature, the stain estimation process is not validated but only the final stain normalization process was evaluated. To the best of our knowledge, this is the first work which attempted to quantitatively assess not only the normalization technique but also the previous stain estimation strategy, which can be considered a necessary step for a correct color deconvolution of histological channels.

The most robust performance metric employed in this work for the evaluation of stain separation is the relative square error (denoted as  $rSE^{sep}(\%)$ ) which represents the error distance in OD space between the stain color appearance matrix

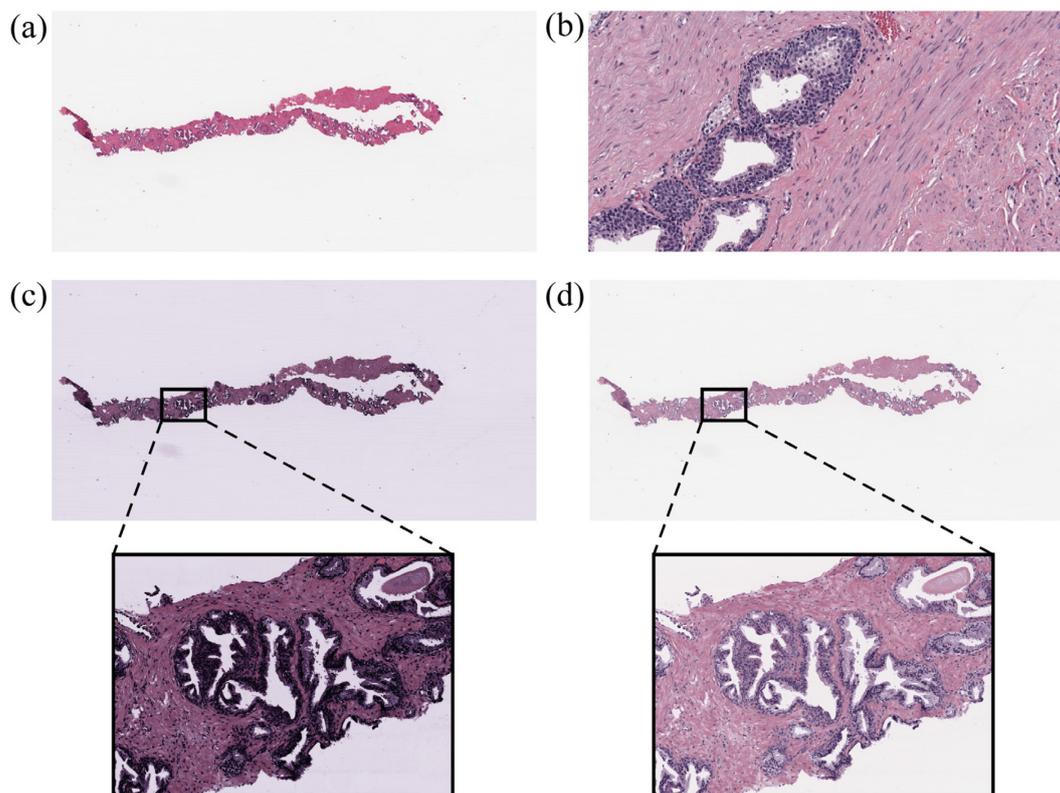
of original image after the stain estimation and the ground truth matrix computed on the manual coordinates annotated by an expert pathologist. SCAN method obtained a lower average value for all tissue and magnifications both for global and H&E channels except for the liver where the model of Khan et al. [13] showed a better value only for eosin, as reported in Figure 5.6 and 5.7. As reported in Figure 5.8, for the other metrics proposed for stain separation, the NMF [17] and SNMF [21] methods obtained better performance in terms of Frobenius norm of the deconvolution error since these mathematical methods tried to minimize this metric in their optimization algorithm, while SCAN method tried to obtain better stain vectors in terms of perceptual contrast and color separation error. Moreover, the minimum solved by the NMF or SNMF method may be a local minimum. Only the method developed by Khan et al. [13] reported a slightly better value of the perceptual contrast than SCAN. On the other hand, SCAN algorithm outperformed all methods in terms of color separation error (both mean value and standard deviation) in H&E deconvolved channels since it adopted segmentation strategies to estimate stain vectors.

For the evaluation of stain normalization performances, the most robust metric is the  $rSE^{norm}(\%)$  since it quantifies the similarity between normalized and target image, which is the aim of stain color standardization. SCAN method exhibited excellent results since it obtained a better global  $rSE^{norm}(\%)$  for all tissues and magnifications, as shown in Figure 5.9 and 5.10. The second best method was the SNMF strategy proposed by Vahadane et al. [21]. In Figure 5.11 other performance metrics proposed in literature were reported: the NMI measure, the SSIM and QSSIM index are intended only to measure the intensity degradation and no information of target color appearance is retained. Similarity measures are slightly lower for SCAN since we prefer preserving the white color of the unstained regions of the source image that were left unaltered in the normalized image rather than normalize them and create a blurring effect on the image. The histogram-based error metrics cannot be considered as robust measures due to their high variability. The perceptual contrast of the final normalized image is very similar to different methods except for Ruifrok et al. [9] which obtained a better value although it showed the worst values in the other performance metrics. Figure 5.13 shows a joint plot between the computational times and the relative square error, which is the most robust metric to quantitatively assess stain normalization, where SCAN hits the point at minimum distance from the origin. Hence, SCAN method is fast and accurate at the same time compared to other approaches.

A qualitative visual comparison is also reported in Figure 5.12; the results show that SCAN method reported less image artifacts with respect to other techniques, due to the correct stain vector estimation and computation of stain density map with the proposed novel strategy. The SNMF method [21] stained cell nuclei in the normalized image of prostate tissue with a darker color with respect to target image.

Finally, a visual comparison between SCAN method and SNMF was carried out on a whole-slide image (WSI), since in the original manuscript Vahadane et al. [21] also tested their method on a whole tissue section. The results reported in

Figure 5.14 are highly satisfactory: SCAN normalization on a WSI took 37 seconds vs 247 seconds of SNMF method. The big advantage of the proposed method is that all unstained regions were excluded from the analysis by using the white detection as a useful pre-processing step and this has a significant impact on computational times.



**Fig. 5.14** Comparison in WSI normalization: the original WSI image (a) was normalized with respect to target image (b) by SNMF strategy (c) and SCAN algorithm (d).

The limitation of the proposed method is related to the preliminary stain estimation: if one of two stains is missing or poorly represented in the image, the method could fail since it is based on the SVD-geodesic method proposed by Macenko et al. [11]. In addition, the subsequent cellular structures segmentation could struggle to detect nuclei regions if the input image size is too small, e.g. for scanned tissue slides acquired at  $2\times$  magnification, the cell nuclei are not visible. In fact, in routine histology the magnification factor is typically higher than  $10\times$  [31].

In conclusion, the aim of the stain normalization strategy is twofold: firstly the standardization of stain color appearance in digital pathology could improve the pathologist's work in the diagnosis of biological diseases, avoid manual re-staining process and reduce intra- and inter-operator variability; secondly this method could be used as a pre-processing step for subsequent CAD systems for accurate cellular structure segmentation [10] and classification [32] based on deep learning framework. In the last few years, deep learning has driven advances in the field of biomedical signal and medical image processing; in particular convolutional neural networks (CNNs) have obtained satisfactory results in image detection,

segmentation and classification tasks [33]. In this context, stain normalization strategy could standardize or adapt the WSI color intensity in order to improve the accuracy of a pre-trained CNN employed for cancer detection [34]. Our research group is currently working on the development of automated solutions based on deep learning approaches for image segmentation and classification and, for this purpose, the proposed SCAN algorithm can be involved in a future work.

## References

- [1] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological Image Analysis: A Review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009, doi: 10.1109/RBME.2009.2034865.
- [2] A. Mouelhi, H. Rmili, J. Ben Ali, M. Sayadi, R. Doghri, and K. Mrad, "Fast unsupervised nuclear segmentation and classification scheme for automatic allred cancer scoring in immunohistochemical breast tissue images," *Comput. Methods Programs Biomed.*, vol. 165, pp. 37–51, 2018.
- [3] J. I. Epstein, L. Egevad, M. B. Amin, B. Delahunt, J. R. Srigley, and P. A. Humphrey, "The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System.," *Am. J. Surg. Pathol.*, vol. 40, no. 2, pp. 244–252, Feb. 2016, doi: 10.1097/PAS.0000000000000530.
- [4] S. G. Veloso, M. F. Lima, P. G. Salles, C. K. Berenstein, J. D. Scalon, and E. A. Bambera, "Interobserver agreement of Gleason score and modified Gleason score in needle biopsy and in surgical specimen of prostate cancer.," *Int. Braz J Urol*, vol. 33, no. 5, pp. 639–46; discussion 647-51, 2007, doi: 10.1590/s1677-55382007000500005.
- [5] T. T. E. Yeo, S. H. Ong, Jayasooriah, and R. Sinniah, "Autofocusing for tissue microscopy," *Image Vis. Comput.*, vol. 11, no. 10, pp. 629–639, 1993, doi: [https://doi.org/10.1016/0262-8856\(93\)90059-P](https://doi.org/10.1016/0262-8856(93)90059-P).
- [6] J. D. Bancroft and M. Gamble, *Theory and practice of histological techniques*. Elsevier health sciences, 2008.
- [7] S. Hodgson, R. F. Harrison, and S. S. Cross, "An automated pattern recognition system for the quantification of inflammatory cells in hepatitis-C-infected liver biopsies," *Image Vis. Comput.*, vol. 24, no. 9, pp. 1025–1038, 2006, doi: <https://doi.org/10.1016/j.imavis.2006.02.019>.
- [8] A. H. Fischer, K. A. Jacobson, J. Rose, and R. Zeller, "Hematoxylin and eosin staining of tissue and cell sections.," *CSH Protoc.*, vol. 2008, no. 5, p. pdb.prot4986, May 2008, doi: 10.1101/pdb.prot4986.
- [9] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution.," *Anal. Quant. Cytol. Histol.*, vol. 23, no. 4, pp. 291–299, Aug. 2001.
- [10] J. Isaksson, I. Arvidsson, K. Åaström, and A. Heyden, "Semantic segmentation of microscopic images of H E stained prostatic tissue using CNN," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 1252–1256, doi: 10.1109/IJCNN.2017.7965996.
- [11] M. Macenko *et al.*, "A method for normalizing histology slides for quantitative analysis," in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009, pp. 1107–1110, doi: 10.1109/ISBI.2009.5193250.
- [12] I. Jolliffe, "Principal Component Analysis BT - International Encyclopedia of Statistical Science," M. Lovric, Ed. Springer Berlin Heidelberg, 2011, pp. 1094–1096.
- [13] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee, "A Nonlinear Mapping Approach to Stain Normalization in Digital Histopathology Images Using Image-Specific Color Deconvolution," *IEEE Trans. Biomed. Eng.*, vol. 61,

- no. 6, pp. 1729–1738, 2014, doi: 10.1109/TBME.2014.2303294.
- [14] M. E. Tipping, “The relevance vector machine,” in *Advances in Neural Information Processing Systems*, 2000, pp. 653–658.
- [15] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, “Color transfer between images,” *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34–41, 2001, doi: 10.1109/38.946629.
- [16] J. Schanda and I. Illumination, *Colorimetry: Understanding the CIE System*. 2007.
- [17] A. Rabinovich, S. Agarwal, C. Laris, J. H. Price, and S. J. Belongie, “Unsupervised Color Decomposition Of Histologically Stained Tissue Samples,” in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. K. Saul, and B. Schölkopf, Eds. MIT Press, 2004, pp. 667–674.
- [18] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, Inc., 2001.
- [19] D. D. Lee and H. S. Seung, “Algorithms for Non-Negative Matrix Factorization,” in *Proceedings of the 13th International Conference on Neural Information Processing Systems*, 2000, pp. 535–541, doi: 10.5555/3008751.3008829.
- [20] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, “Algorithms and applications for approximate nonnegative matrix factorization,” *Comput. Stat. Data Anal.*, vol. 52, no. 1, pp. 155–173, 2007, doi: <https://doi.org/10.1016/j.csda.2006.11.006>.
- [21] A. Vahadane *et al.*, “Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images,” *IEEE Trans. Med. Imaging*, vol. 35, no. 8, pp. 1962–1971, 2016, doi: 10.1109/TMI.2016.2529665.
- [22] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, J. Shi, and C. Xue, “Adaptive color deconvolution for histological WSI normalization,” *Comput. Methods Programs Biomed.*, vol. 170, pp. 107–120, Mar. 2019, doi: 10.1016/J.CMPB.2019.01.008.
- [23] S. E. Grigorescu, N. Petkov, and P. Kruizinga, “Comparison of texture features based on Gabor filters,” *IEEE Trans. Image Process.*, vol. 11, no. 10, pp. 1160–1167, 2002, doi: 10.1109/TIP.2002.804262.
- [24] J. Lim, *Two-dimensional signal and image processing*. Englewood Cliffs N.J.: Prentice Hall, 1990.
- [25] M. Salvi and F. Molinari, “Multi-tissue and multi-scale approach for nuclei segmentation in H&E stained images,” *Biomed. Eng. Online*, vol. 17, no. 1, p. 89, 2018, doi: 10.1186/s12938-018-0518-0.
- [26] R. Penrose, “A generalized inverse for matrices,” *Math. Proc. Cambridge Philos. Soc.*, vol. 51, no. 3, pp. 406–413, 1955, doi: DOI: 10.1017/S0305004100030401.
- [27] M. Salvi, N. Michielli, and F. Molinari, “Stain Color Adaptive Normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology,” *Comput. Methods Programs Biomed.*, vol. 193, p. 105506, 2020, doi: <https://doi.org/10.1016/j.cmpb.2020.105506>.
- [28] J. N. Kather, C. A. Weis, A. Marx, A. K. Schuster, L. R. Schad, and F. G. Zöllner, “New Colors for Histology: Optimized Bivariate Color Maps Increase Perceptual Contrast in Histological Images,” *PLoS One*, vol. 10, no. 12, pp. e0145572–e0145572, Dec. 2015, doi: 10.1371/journal.pone.0145572.

- [29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004, doi: 10.1109/TIP.2003.819861.
- [30] A. Kolaman and O. Yadid-Pecht, "Quaternion Structural Similarity: A New Quality Index for Color Images," *IEEE Trans. Image Process.*, vol. 21, pp. 1526–1536, Dec. 2011, doi: 10.1109/TIP.2011.2181522.
- [31] A. D. Belsare and M. M. Mushrif, "Histopathological image analysis using image processing techniques: An overview," *Signal Image Process.*, vol. 3, no. 4, p. 23, 2012.
- [32] H. Sharma, N. Zerbe, I. Klempert, O. Hellwich, and P. Hufnagl, "Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology," *Comput. Med. Imaging Graph.*, vol. 61, pp. 2–13, 2017, doi: <https://doi.org/10.1016/j.compmedimag.2017.06.001>.
- [33] M. Talo, "Automated classification of histopathology images using transfer learning," *Artif. Intell. Med.*, vol. 101, p. 101743, 2019, doi: <https://doi.org/10.1016/j.artmed.2019.101743>.
- [34] Y. Celik, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, "Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images," *Pattern Recognit. Lett.*, vol. 133, pp. 232–239, 2020, doi: <https://doi.org/10.1016/j.patrec.2020.03.011>.

## Chapter 6

# Inverse Color Deconvolution for Stain Normalization in Immunohistochemistry

*Part of this chapter has been patented as:*

F. Molinari, M. Salvi and N. Michielli, **System for processing an image relating to a histological tissue**. Italian patent application on 24/03/2020, number: 102020000006148. International patent application (PCT) on 15/03/2021, number: PCT/IB2021/052119.

*Part of the study described in this chapter has been carried out as part of the:*

PoC Instrument (cut-off 1) - LINKS Foundation initiative, with LIFTT support, based on funding from the Compagnia di San Paolo Foundation. F. Molinari, M. Salvi, O. Pennisi and N. Michielli, **STAINS - STANDARDIZATION & NORMALIZATION OF HISTOLOGICAL SLIDES**. Start date: 15/07/2020, duration of the project: 9 months.

## 6.1 Introduction

As discussed in Chapter 4, two major types of staining methods in histology and histopathology are employed to make cell components visible, under a light microscope, in an excised tissue section for diagnosis purposes: histochemical and immunohistochemical staining [1]. Histochemistry refers to histological stains which bind selectively to cell constituents and highlight chemical features and the most widely used histochemical staining in routine histology is hematoxylin and eosin (H&E) as discussed in Chapter 5.

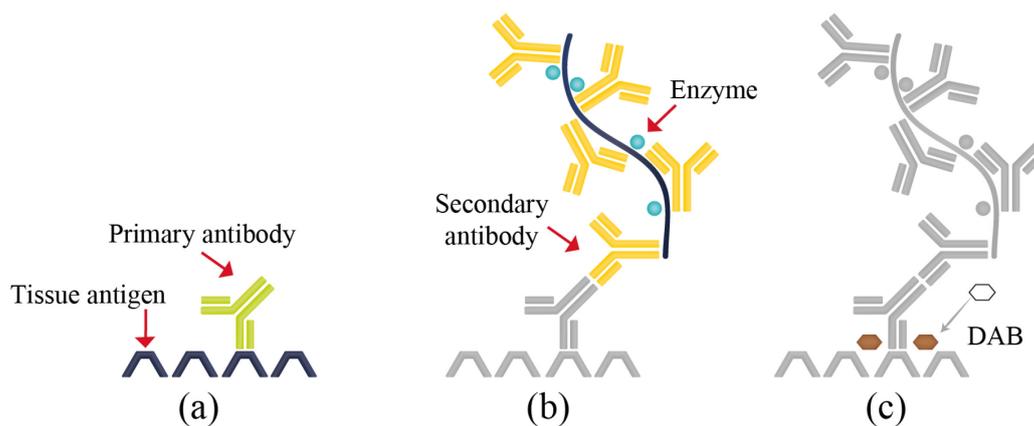
On the other hand, immunohistochemistry (IHC) is an immunology-based method which employs tissue-based biomarkers (i.e. antibodies marked with specific reagents) for the localization of specific tissue components (generally these are proteins) that act as antigens. Thus, IHC technique is based on antigen-antibody interaction combined with enzymatic or fluorescent detection systems that make the reaction visible under a light microscope. The presence or absence of a certain antigen can be easily detected with this staining technique which allows to visualize its protein expression by means of autoradiography (i.e. the specimen itself is the source of the radiation) if the antibody is radioactively labeled, with a fluorescence microscope, if the antibody is labelled with fluorescent material or with an optical microscope if the antibody is conjugated to enzymes that cause a chemical reaction capable of pigmenting the cell which contains the antigen [2]. For this reason, pathologists select specific antibodies in order to identify cellular structures or elements that are associated with certain human diseases for diagnostic, prognostic, and therapeutic purposes. The tissue stains differently in the presence of cancer, inflammatory response, infections or autoimmune diseases. IHC staining is widely used in the diagnosis of cancerous tumors, especially for breast cancer [3].

In IHC, monoclonal antibodies, i.e. antibodies with an absolute specificity for a single epitope on the antigen molecules, are used. Ki-67 is a monoclonal antibody specific for KI-67 antigen, also called MKI67 (marker of proliferation Ki-67), which is observed in all stages of the cell-division cycle (G1, S, G2 and mitosis), except the quiescent phase G0 due to its short half-life [4]. The KI-67 antigen is a nuclear protein involved in cell proliferation, therefore it can be considered as a useful maker of the growth fraction of a given population of cells [5], [6].

However, the antigen must not be altered in its tertiary structure for the immunoreaction to take place, otherwise it may lose its ability to react with the antigen. Many tissue fixatives such as formaldehyde, as described in Chapter 4, alter this structure, with the result of masking the antigenic sites. For this reason, in the past, IHC reactions were carried out only on cryostat sections of frozen tissues. The advent of various expedients employed to recovery masked antigens through the breaking of the aldehyde bond on proteins has allowed the application of IHC techniques on paraffin-embedded tissue samples [7].

The IHC staining process consists of several steps with the result of an optically scanned tissue slide where proliferating cells, i.e. cells which have been marked by the antibody are highlighted with respect to the rest of the tissue. Firstly, all the

preliminary processes such as chemical fixation, dehydration, embedding in paraffin wax and section cutting are applied by the histology technician as described in Chapter 4; then antigen retrieval is performed due to alterations during chemical fixation and the primary monoclonal antibody specifically binds to the antigen of interest, with the antigen-antibody reaction. Then, a secondary antibody is necessary for carrying the label enzyme and a chromogen is applied to visualize the antigen of interest within the tissue. The enzyme molecules are necessary for the chromogen to produce the colorimetric reaction [8]. In IHC, the chromogen employed to stain proliferating cells is the 3,3'-Diaminobenzidine (DAB), which is a highly thermochemically stable polybenzimidazole that provides brown-colored staining [9]. Thus, to summarize, the monoclonal antibody on one side binds to the KI-67 antigen (i.e. biomarker of proliferating cells) and on the other hand binds to the chromogen DAB, which highlights the nuclei of proliferating cells in brown color, as graphically shown in Figure 6.1. In addition, in IHC staining, hematoxylin acts as a counterstain (i.e. is used as a dye for contrast) in order to visualize cell nuclei, which are not marked by the antibody, and overall tissue architecture (i.e. cell stroma and cytoplasmic structures) [10].



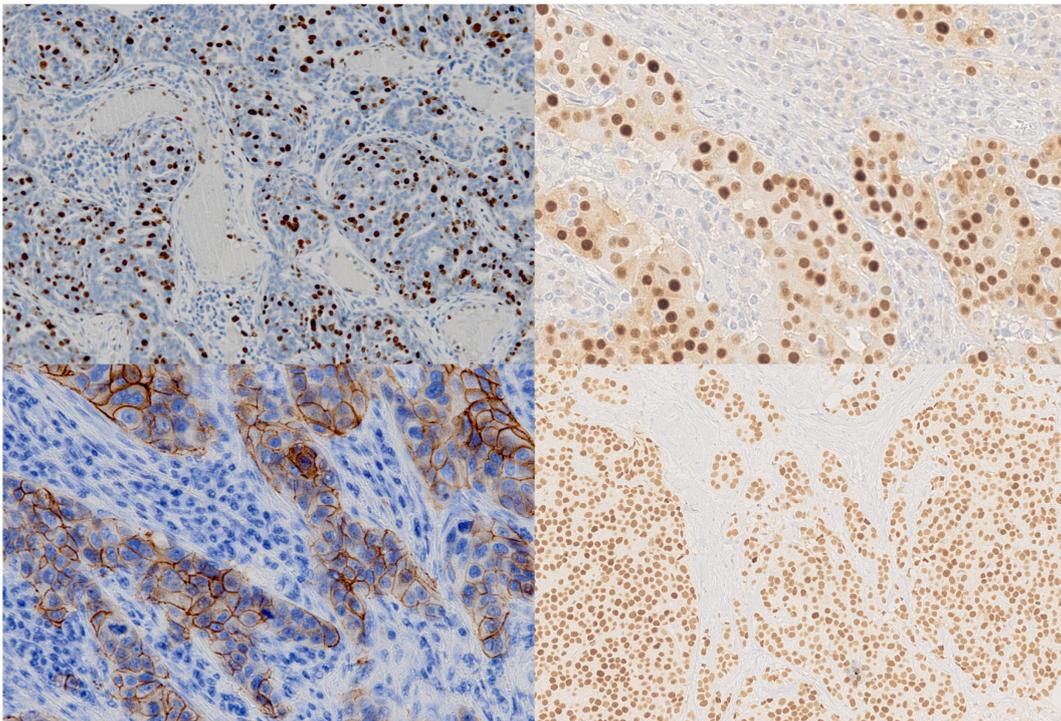
**Fig. 6.1** Analytical steps of the IHC staining process: a primary monoclonal antibody binds to the target antigen (a); the secondary antibody which binds to the primary one carries the label enzyme (b); the chromogen (DAB) is applied to highlight the antibody/antigen complex (c).

Subsequently, tissue sections are mounted on a glass slide and finally, the optically scanned digital images are analyzed by the pathologist which provides the diagnosis, and the oncologist prescribes therapy accordingly.

Pathologists select specific antibodies (e.g. Ki-67 biomarker) in order to identify cellular structures or elements that are associated with certain human diseases for diagnostic, prognostic, and therapeutic purposes. The selection of the primary antibody has a significant impact on the staining result, with negative consequences to diagnostic specificity and sensitivity. The tissue stains differently in the presence of cancer, inflammatory response, infections, or autoimmune diseases. IHC staining is widely used in the diagnosis of cancerous tumors,

especially for breast cancer which is the most common form which affects especially women [11].

As discussed for H&E staining in Chapter 5, the histological tissue can take on different staining intensities depending on the concentration, the pH and the degree of deterioration of the staining dye, the staining exposure time, the ability of histology technician, the illumination condition and the scanner performances. The reproducibility of color appearance recurs in this context with a major impact for this pathologist's diagnosis. An example of stain variability in digital IHC stained images is reported in Figure 6.2.



**Fig. 6.2** Staining variability of Ki-67 positive nuclei stained with DAB and hematoxylin used as counterstain.

In this context, the concept of stain standardization becomes relevant for medical support, diagnosis improvement, to reduce both time loss and financial resources for re-staining and finally this method could be used as a pre-processing step for subsequent computer-assisted diagnosis (CAD) systems for the estimation of proliferation index (PI). PI is an important parameter used in histopathology, defined as the ratio between the number of cell nuclei which are marked with chromogen (i.e. the number of tumor nuclei) over the total number of nuclei within tissue. The higher the PI, the higher the number of cells which are subjected to cell division, which means a higher tumor aggressiveness; thus it represents a prognostic marker in cancer detection [12].

As evidenced in Chapter 5, the first step in a stain normalization framework is the stain separation based on the mathematical technique of color deconvolution. The first step consists in the estimation of matrices  $W$  and  $H$  involved in the

decomposition. The physical law employed for the conversion of a RGB image into optical density (OD) space is the Beer-Lamber (BL) law, defined as follows:

$$V = -\log_{10}\left(\frac{I}{I_0}\right) = H \cdot W \quad (\text{Eq. 6.1})$$

where  $I$  and  $V$  are the source image in RGB color space and OD space respectively,  $I_0$  is the incident light intensity (set to 255 for 8-bit images), and  $W$  and  $H$  are unknown matrices involved in the factorization.  $W$  is the stain color appearance matrix which contains in each row a stain vector and  $H$  is the stain density map which contains in each column the concentration of a specific stain for all pixels in the image.

Several state-of-the-art methods have been proposed for the automated estimation of these two matrices especially for H&E stained slides, as discussed in Chapter 5. In histology, most of histochemical stains (e.g. hematoxylin and eosin) are light-absorbing, while this is not true in IHC. DAB and hematoxylin follow BL law at low concentration levels, while for high Ki-67 concentration, the term related to photonic light scattering becomes dominant [13]. Actually, there is always a combination of attenuation and photonic scattering inside the tissue. Absorbance only evaluates the ratio of transmitted over incident light intensity, as shown in Eq. 6.1, not the mechanism by which light intensity decreases (i.e. pure absorption or scattering) [14].

To the best of our knowledge, no robust or fully validated methods have been employed for the stain normalization in IHC digital images. The most robust methods employed in H&E failed for IHC stain separation since polymers, such as DAB, do not obey to BL law for their scattering behavior. In this case, the stain estimation in OD space, as done in the work of Ruifrok et. al. [15], Macenko et al. [16], Vahadane et al. [17] and Zheng et al. [18] is not effective in separating IHC stains with the consequence that concentration maps contain incorrect values. Other authors [19] proposed to remove DAB stained regions from the analysis prior to color deconvolution in order to consider only light absorbing stains, but this cannot be considered a solution.

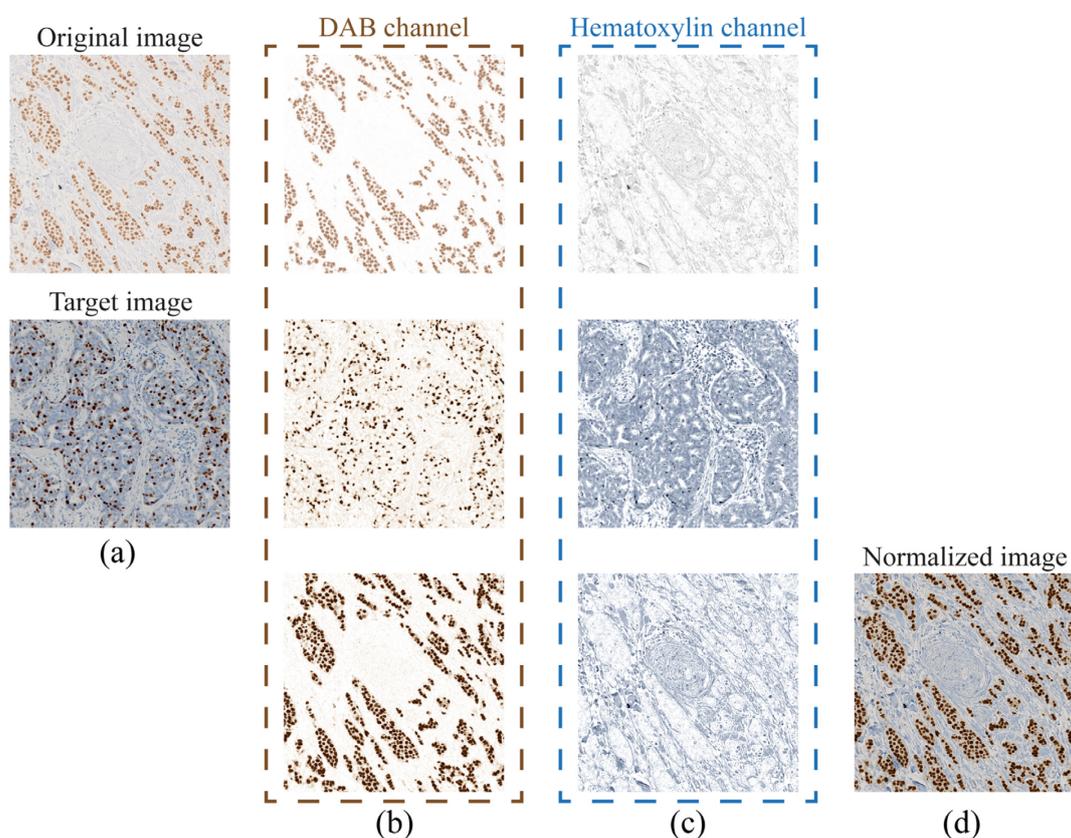
In this chapter, an automated strategy for the estimation of stain color appearance matrix for DAB and hematoxylin is proposed and a novel stain normalization approach, applied to IHC stained histopathological images, is developed.

## 6.2 Materials and Methods

The proposed method for stain normalization of IHC stained digital histopathological images consists of two main steps: *i*) stain estimation and separation of channels related to chromogenic dyes used in IHC, i.e. DAB for nuclei of proliferating cells and hematoxylin as counterstaining; *ii*) stain normalization of separate channels and reconstruction of final normalized image by using a novel

proposed strategy called Inverse Color Deconvolution (ICD). A graphical flowchart of the proposed method is reported in Figure 6.3.

The automated method was tested on a preliminary image dataset which consists of 94 optically scanned IHC stained histological images acquired with Aperio Scanscope (Leica Biosystems). Each digital slide is a RGB image with a size of  $1088 \times 1920$  or  $1128 \times 2000$  pixels; the image input dimension is not relevant for the processing. The histological sections belonged to breast cancer tissue, scanned at  $20\times$  and  $40\times$  magnification and stained with IHC staining method, using DAB for Ki-67 protein expression and hematoxylin as counterstain. The target or template image used for this study belongs to breast tissue, digitized with a  $20\times$  magnification and presents different DAB and hematoxylin stain hues and intensities with a better contrast for the quantification of PI; the choice was performed by an expert pathologist.



**Fig. 6.3** Flowchart of the proposed method: (a) original source (top row) and target image (central row); (b) DAB channel of the input (top row) and target image (central row) and normalized channel after stain normalization (bottom row); (c) Hematoxylin channel of the input (top row) and target image (central row) and normalized channel after stain normalization (bottom row); (d) final normalized image (bottom row) using Inverse Color Deconvolution.

### 6.2.1 Stain estimation and separation

The first step of the proposed method consists in the estimation of stain vectors for DAB stained nuclei and the rest of the tissue counterstained with hematoxylin. The SVD method in OD space, proposed by Macenko et al. [16], which was

employed in Chapter 5 for H&E stained histological images, cannot be used in this work, since IHC staining dyes not only absorb but also scatter light source, as previously described. Hence the separation of Ki-67 positive regions from hematoxylin stained tissue is the critical part of the pipeline.

Firstly, the unstained regions (e.g. gland lumen, background, etc.) which assume white colors in the original RGB image since the light has passed through without absorption, were detected using the same strategy described in Chapter 5, based on 2D Gabor filter. Subsequently, the hue saturation intensity model is employed for the segmentation of DAB stained cellular structures and the blue ratio space is used for the detection of hematoxylin regions. The hue saturation intensity (HSI) model has the advantage of decomposing red, green, and blue colors into intensity, which refers to the optical energy absorbed and the chromaticity, which is dependent on light color. The chromaticity, in turn, can be expressed as hue and saturation in a new coordinate system  $c_x c_y$  [20]. In this work, we employed the chromaticity layer  $c_x$ , from the HSI model, for the detection of DAB stained nuclei for their red component which is higher with respect to that of hematoxylin dye. The mathematical expression of  $c_x$  layer is reported here:

$$c_x = \frac{3 \cdot I_R}{I_R + I_G + I_B} - 1 \quad (\text{Eq. 6.2})$$

where  $I_R$ ,  $I_G$  and  $I_B$  are the original image intensities in red, green, and blue channel, respectively. Higher values of this layer refer to DAB cell nuclei. The segmentation was obtained by using the nonparametric Otsu's method which allows to select an automatic threshold that maximizes inter-class variance (or equivalently minimizes the intra-class variance) in a bimodal gray-level histogram [21].

For the detection of hematoxylin healthy nuclei and the rest of the tissue stained with hematoxylin, the original RGB image was converted into blue ratio (BR) image [22], using the following expression:

$$BR = \frac{100 \cdot I_B}{(1 + I_R + I_G) \cdot (1 + I_R + I_G + I_B)} \quad (\text{Eq. 6.3})$$

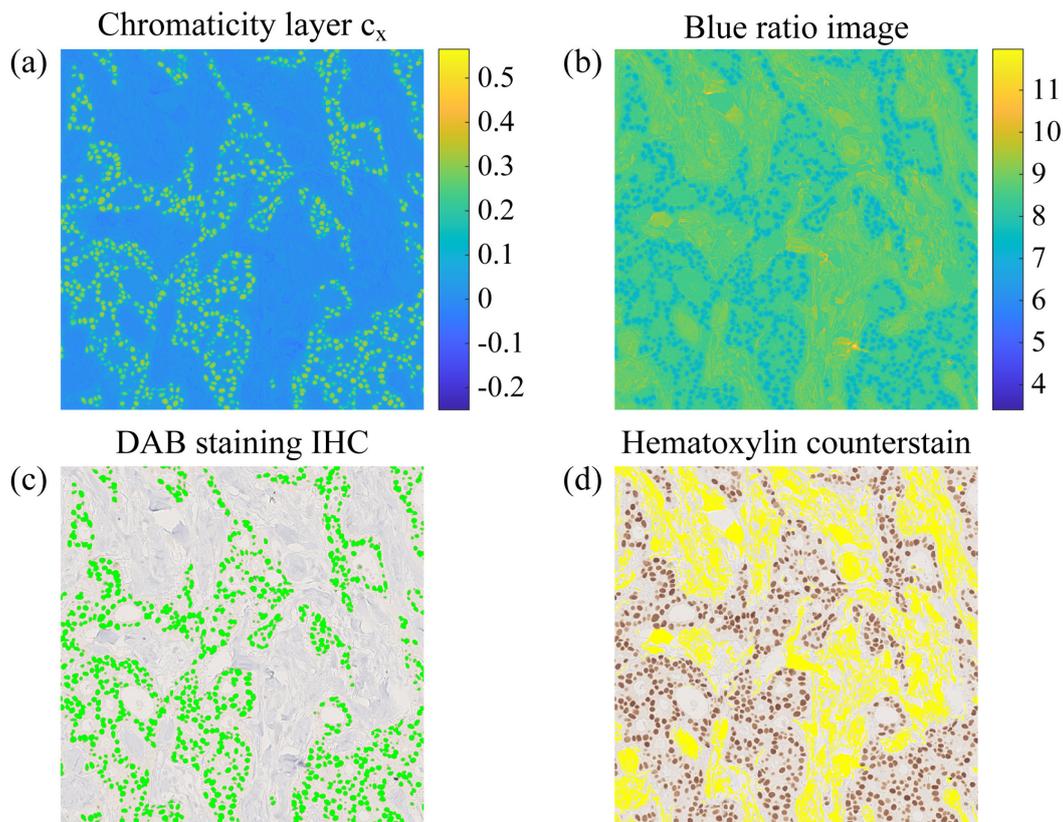
where  $I_R$ ,  $I_G$  and  $I_B$  are the original image intensities in red, green, and blue channel, respectively. Higher values of BR image refer to hematoxylin stained areas and the segmentation was performed using Otsu's method after removing white unstained regions. The results of the detection of DAB and hematoxylin areas is reported in Figure 6.4. In addition, a preliminary check was implemented to detect if cell nuclei stained with DAB were present in the image; in the negative case, the normalization was not performed.

After having segmented proliferating cell nuclear regions and counterstained tissue regions, a median value of color intensity was computed and the final stain color appearance matrix  $W$  was estimated. This matrix was normalized so that each row (i.e. stain vector) had unit Euclidean norm [15]. The stain density map  $H$  was estimated using the new strategy proposed in Chapter 5 with the aim to maintain physical meaning of concentration values. Therefore, the first column  $H(:,1)$  was

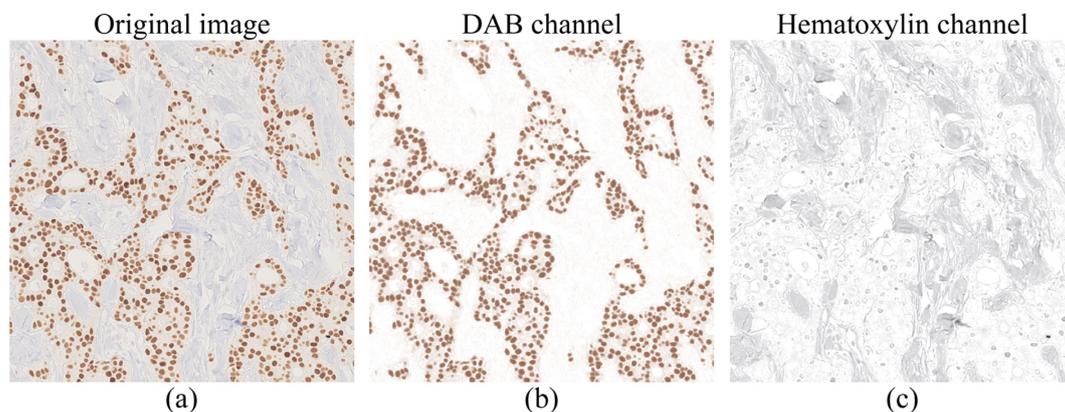
computed as the median value of the ratio between DAB triplets and first stain vector  $W(1, :)$ . The same process was performed for the second column  $H(:, 2)$  in order to obtain hematoxylin concentration values [23]. Finally, a color deconvolution was performed to separate DAB and hematoxylin channels, using the following expression, as described in Chapter 4:

$$I_j = I_0 \cdot 10^{-[H(:,j) \cdot W(j,:) ]}, \quad j = 1, 2 \quad (\text{Eq. 6.4})$$

where  $j=1$  refers to DAB deconvolved channel and  $j=2$  denotes the hematoxylin deconvolved channel. The deconvolved channels are reported in Figure 6.5.



**Fig. 6.4** Chromaticity layer (a) and blue ratio image (b) are shown in the first row. In the second row, segmentation masks of nuclei stained with DAB (c, in green) and hematoxylin regions (d, in yellow) are reported.



**Fig. 6.5** Original RGB image (a). Stain separation of DAB channel (b) and hematoxylin channel (c).

## 6.2.2 Inverse Color Deconvolution

In this thesis work, a novel mathematical method for stain normalization of separate channels and subsequent reconstruction of the final RGB normalized image is proposed. This strategy performs color deconvolution in the opposite direction and for this reason is called Inverse Color Deconvolution (ICD).

After having decomposed the original source image into two separate channels using Eq. 6.4, the same process was performed for target image, as reported in Figure 6.6, and the corresponding stain density maps were normalized as follows:

$$H_j^{norm} = H_j^{source} \cdot \frac{RM(H_j^{target})}{RM(H_j^{source})}, \quad j = 1,2 \quad (\text{Eq. 6.5})$$

where the term  $RM(\cdot)$  is the robust maximum (i.e. the 99<sup>th</sup> percentile). This expression is useful to adapt the dynamic range of stain density map of the source image with respect to that of target image for each channel individually.

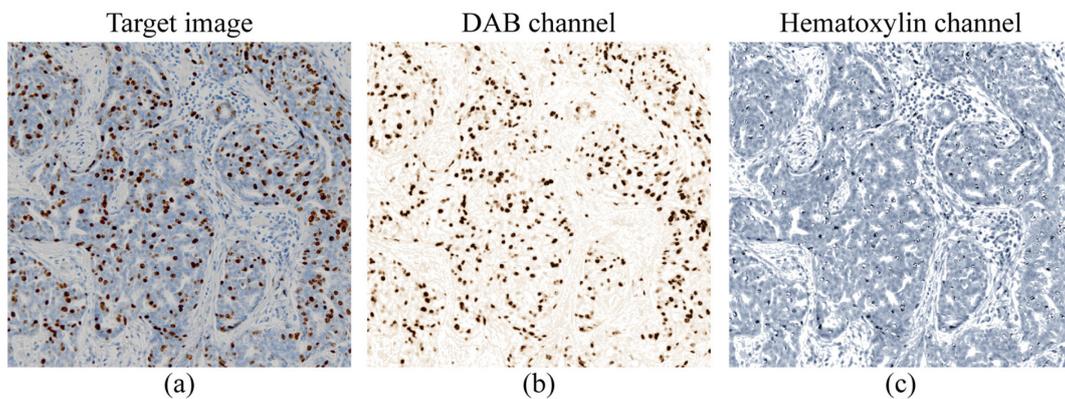


Fig. 6.6 Target RGB image (a). Stain separation of DAB channel (b) and hematoxylin channel (c).

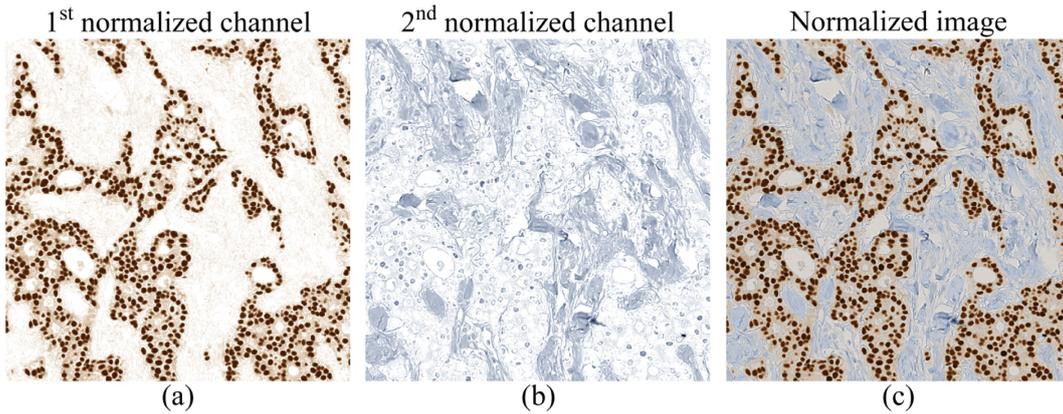
Subsequently, the normalized image for each channel was computed using the following definition:

$$I_j^{norm} = I_0 \cdot 10^{-(H_j^{norm} \cdot W_j^{target})}, \quad j = 1,2 \quad (\text{Eq. 6.6})$$

where  $W_1^{target}$  and  $W_2^{target}$  are the stain vectors of target image for DAB and hematoxylin dye, respectively. From these channels, new stain density maps were computed and merged in a new global matrix, denoted as  $H^{new}$ , which was used to reconstruct the final normalized image, in RGB color space, as follows:

$$I^{norm} = I_0 \cdot 10^{-(H^{new} \cdot W^{target})} \quad (\text{Eq. 6.7})$$

The IHC channels that were normalized separately and the final RGB normalized image, reconstructed using ICD strategy, are reported in Figure 6.7.



**Fig. 6.7** Normalized channel of positive cell nuclei stained with DAB (a) and areas stained with hematoxylin (b). Final normalized image, in RGB color space, performed by ICD (c).

### 6.3 Experiments and Results

The stain estimation and the ICD strategy were developed and implemented in MATLAB environment (MATLAB, The MathWorks, Inc., Natick, MA, USA). The proposed approach was both quantitatively and qualitatively validated and compared with two state-of-the-art methods. Most of the methods discussed in Chapter 5 were based on the BL law for the stain separation with the hypothesis of pure absorbance. The crucial issue in IHC is that such dyes as DAB not only absorb but also scatter light source and for this reason these methods failed the estimation in IHC stained histopathological slides [19]. The histogram transformation-based method proposed by Reinhard et al. [24] and the stain color descriptor proposed by Khan et al. [25] were compared to ICD strategy. The first method is a color transfer based on the computation of image statistics (i.e. mean and standard deviation) of the source image and target image in  $\lambda\beta$  color space, while the second one introduces a new color descriptor to estimate stain concentrations and relevance vector machine is used to classify stain color. The normalization of the source image with respect to target image is performed using spline-based nonlinear functions.

The ground truth stain vectors for the source and target image were generated with a manual procedure performed by an expert pathologist (with 20 years of experience). For each source and target image, the manual operator selected at least 100 points in tissue regions which belonged to DAB stained nuclei and hematoxylin tissue portions, in order to find a robust ground truth stain color appearance estimation. The manual points were carefully annotated in the areas where presumably only one stain bound to cellular structures. Then, the following relative Square Error (rSE) was employed:

$$rSE^{norm}(\%) = \frac{\|W - W_{GT}\|_F^2}{\|W\|_F \cdot \|W_{GT}\|_F} \cdot 100 \quad (\text{Eq. 6.8})$$

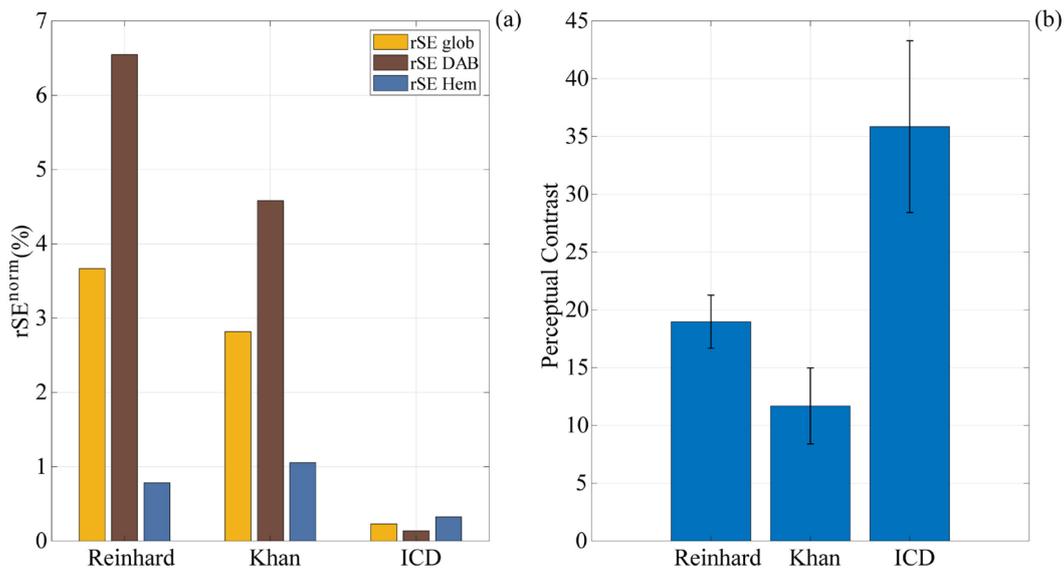
where the subscript  $F$  denotes the Frobenius norm. In Eq. 6.8, the matrix  $W$  contained the stain colors computed on the manual points in the normalized image corresponding to the manual points annotated by the expert pathologist in source image, since the structures were not distorted by ICD method after the

normalization. On the other hand, the matrix  $W_{GT}$  was the ground truth matrix for the target image (i.e. true stain colors manually selected by the expert on the target image). In addition, the rSE can be computed on the entire stain color appearance matrix ( $rSE_{glob}$ ) or retaining only the first/second row, in order to compute the  $rSE_{DAB}/rSE_{Hem}$  respectively. The lower the error, the better the performance of the stain normalization method, since the normalized image assumes a stain color similar to that of target image (which is the aim of the standardization). Another quantitative metric used for validation was the perceptual contrast between normalized tissue and background in  $l\alpha\beta$  color space, defined as:

$$C = \sqrt{(\Delta l)^2 + (\Delta\alpha)^2 + (\Delta\beta)^2} \quad (\text{Eq. 6.9})$$

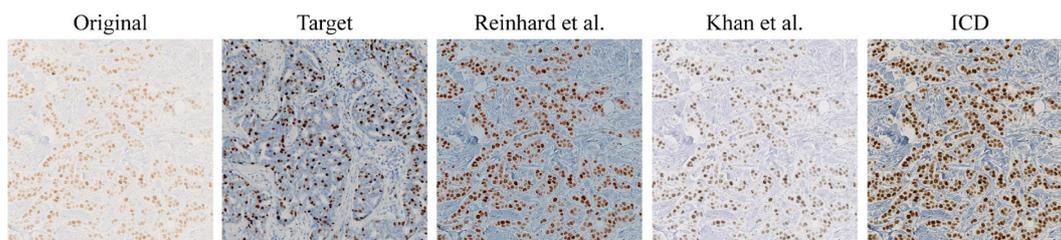
$$\Delta l = \langle l \rangle^{obj} - \langle l \rangle^{bg}; \quad \Delta\alpha = \langle \alpha \rangle^{obj} - \langle \alpha \rangle^{bg}; \quad \Delta\beta = \langle \beta \rangle^{obj} - \langle \beta \rangle^{bg}$$

where  $\langle \cdot \rangle$  denotes the mean value and the superscripts  $obj$  and  $bg$  indicate the object (i.e. tissue) and background (i.e. unstained tissue) regions, respectively. The contrast was defined in  $l\alpha\beta$  color space since Euclidean distances between colors in this perceptually uniform space are related to the color difference perceived by a human observer [26]. The quantitative performance measures for stain normalization are reported in Figure 6.8.



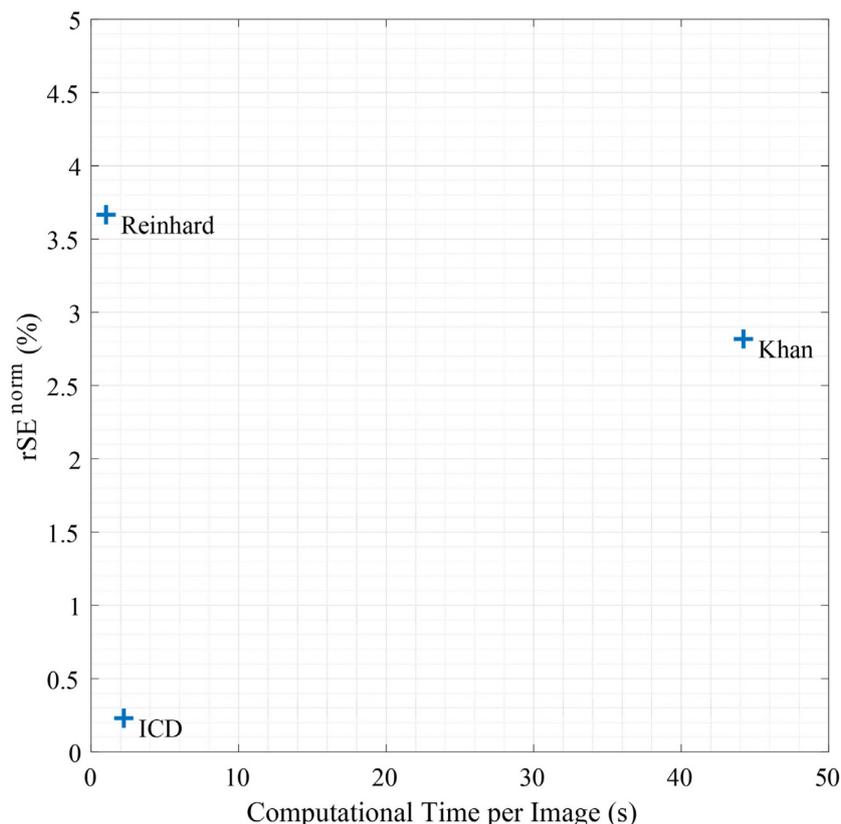
**Fig. 6.8** (a) Quantitative comparison between ICD algorithm and other state-of-the-art methods in terms of global rSE (in orange), DAB rSE (in brown) and hematoxylin rSE (in blue) for stain normalization. (b) Perceptual contrast between normalized tissue and unstained regions reported with mean value and symmetrical error bars for standard deviation.

In addition to quantitative performance measures, a qualitative visual comparison was reported in Figure 6.9, between the two published methods about stain normalization and the ICD algorithm, with respect to the same target image.



**Fig. 6.9** Visual performance comparison between the proposed ICD algorithm and other state-of-the-art methods. First column displays the original IHC stain histopathological image while the second column shows the target image. Normalization methods are displayed in the third and fourth column while ICD algorithm is reported in the last one.

Another most important aspect for the evaluation of an automated algorithm for stain normalization is the computational time. In Figure 6.10, a joint plot of  $rSE^{norm}(\%)$  vs average running time to process a single histopathological image, expressed in seconds, is reported for different state-of-the-art methods. ICD algorithm employed around 2.2 seconds to estimate source and target stain vectors and perform stain normalization of the source image with respect to the target. The processing was performed on a workstation with 16 GB of RAM, 2.5 GHz quad-core CPU and 64-bit version of Windows.



**Fig. 6.10** Joint plot of  $rSE$  for stain normalization (y-axis) and average running time (x-axis) to normalize a single RGB histopathological image, for different automated methods.

## 6.4 Discussion and Conclusion

A novel fully automated stain separation and normalization approach for immunohistochemical stained tissue slides is presented. The algorithm, named ICD (Inverse Color Deconvolution), is based on stain estimation using chromaticity layer and blue ratio image and the normalization is performed on stain channels, separately. Finally, the normalized image is reconstructed using a novel proposed mathematical technique. To the best of our knowledge, this is the first method which employs a novel strategy for stain normalization in immunohistochemistry.

IHC is an immunology-based method which employs tissue-based biomarkers (i.e. antibodies marked with specific reagents) for the localization of specific tissue components (generally these are proteins) that act as antigens. The tissue stains differently in the presence of cancer, inflammatory response, infections, or autoimmune diseases. IHC staining is widely used in the diagnosis of cancerous tumors, especially for breast cancer [3]. In IHC, the chromogen employed to stain proliferating cells is DAB and hematoxylin is used for counterstaining.

The reproducibility of histological specimens is affected by color intensity variation of stained images due to several factors such as the staining concentration and exposure time, the ability of histology technician, the non-uniform slide thickness, and the scanner specifications. The solution to this problem is the standardization of color appearance in digital pathology. In this context emerges the stain normalization method which adapts the color intensity of histological images to a reference template, chosen by an expert pathologist, which presents optimal characteristics in terms of stain distribution, perceptual contrast and color intensity saturation.

In histology, most of histochemical stains (e.g. hematoxylin and eosin) are light-absorbing, while this is not true in IHC. DAB and hematoxylin follow BL law at low concentration levels, while for high Ki-67 concentration, the term related to photonic light scattering becomes dominant [13]. Actually, there is always a combination of attenuation and photonic scattering inside the tissue. To the best of our knowledge, no robust or fully validated methods have been employed for the stain normalization in IHC digital images. The most robust methods employed in H&E failed for IHC stain separation since polymers, such as DAB, do not obey to BL law for their scattering behavior. In this case, the stain estimation in OD space, as done in Chapter 5 for H&E, is not effective in separating IHC stains with the consequence that concentration maps contain incorrect values. Other authors [19] proposed to remove DAB stained regions from the analysis prior to color deconvolution in order to consider only light absorbing stains, but this cannot be considered a solution.

In this chapter, an automated strategy for the estimation of stain color appearance matrix for DAB and hematoxylin is proposed and a novel stain normalization approach, applied to IHC stained histopathological images, is developed with the aim of preserving the unstained structures such as the lumen and the background.

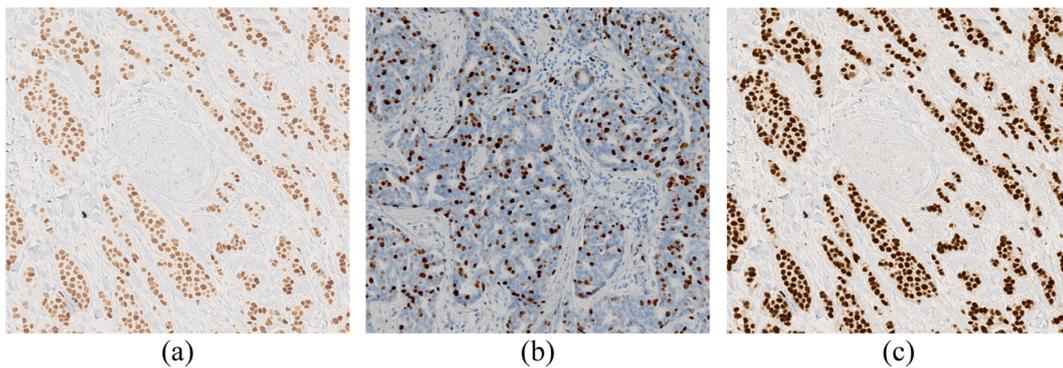
The proposed method was validated on a dataset composed of 94 breast cancer images and obtained better performance measures than other published methods. The most robust performance metric employed in this work for the evaluation of stain normalization is the relative square error (denoted as  $rSE^{norm}(\%)$ ) which quantifies the similarity between normalized and target image, which is the aim of stain color standardization. ICD method obtained a lower average value both for global and separate channels (i.e. DAB and hematoxylin), as shown in Figure 6.8 (a). Another metric for the evaluation of stain normalization performance is the perceptual contrast between the normalized tissue and the unstained regions (e.g. background, gland lumen, etc.). ICD outperformed the other methods, as shown in Figure 6.8 (b), but the standard deviation is very high, and this represents a problem for the contrast uniformity of this procedure. A qualitative visual comparison was also reported in Figure 6.9; the results show that the method proposed by Khan et al. [25] failed in the estimation of stain colors, while the method of Reinhard et al. [24] normalized also the unstained regions with the color of hematoxylin dye. On the other hand, ICD method correctly normalized DAB and hematoxylin stained areas and preserved white regions due to the correct stain vector estimation and the preliminary white detection. Figure 6.10 shows a joint plot between the computational times and the relative square error, where ICD hits the point at minimum distance from the origin. Hence, ICD method is fast and accurate at the same time compared to other approaches.

The stain normalization strategy becomes relevant for medical support, diagnosis improvement, to reduce both time loss and financial resources for re-staining and this method could be used as a pre-processing step for subsequent computer-assisted diagnosis (CAD) systems for the estimation of proliferation index (PI). PI is an important parameter used in histopathology, defined as the ratio between the number of cell nuclei which are marked with chromogen (i.e. the number of tumor nuclei) and the total number of nuclei within tissue. The higher the PI, the higher the number of cells which are subjected to cell division, which means a higher tumor aggressiveness; thus it represents a prognostic marker in cancer detection [12].

In IHC, it is sometimes not desired to normalize all stains because the stain intensity may indicate the expression or absence of a particular antigen, hence the ICD became useful for this purpose. The ICD strategy allows to normalize a single stain selected by the clinician, i.e. to normalize DAB stain without affecting hematoxylin dye and viceversa; the result is reported in Figure. 6.11. In addition, if one of two stains is poorly represented in the image, the ICD may solve this problem.

The limitation of this method is that the strategy for image segmentation of bio-marked cell nuclei and healthy cellular structures may be too selective for images with a heterogeneous color distribution, although the color estimation may be equally correct since an average value is enough for the computation of stain vectors. In the near future, a more robust classification method will be applied for the detection of DAB and hematoxylin regions in chromaticity plane and blue ratio

image, respectively. Moreover, the method was validated on a small dataset, belonging to a single tissue and for this reason it is considered a preliminary study. Analysis of a larger number of images is mandatory to further validate the method. The big advantage is the low computational time to perform stain normalization. Another shortcoming of the proposed method may be found if red blood cells and blood vessels were present in the image; the stain estimation based on chromaticity layer could fail due to the higher red color component of this structures. Considering this, in future works, a preliminary detection algorithm to identify these regions will be implemented for minimizing the influence of bleeding areas on the estimation of stain color appearance matrix estimation.



**Fig. 6.11** Example of ICD algorithm capability to normalize only DAB stained nuclei (c) in a histopathological image (a) with respect to a given template (b).

In conclusion, the aim of the stain normalization strategy is twofold. Firstly, the standardization of stain color appearance in digital pathology could improve the pathologist's work in the diagnosis of biological diseases, avoid manual re-staining process and reduce intra- and inter-operator variability. Secondly, this method could be used as a pre-processing step for improving the performances of CAD systems employed for automated Ki-67 quantification in IHC stained images [27], automatic computation of PI in breast cancer histopathological images [12] and carcinoma classification based on deep learning framework [28]. In addition, a potential aspect of the proposed ICD strategy is its capability to normalize only a single channel which may be useful especially in IHC where pathologists select specific antibodies in order to identify biomolecules that are associated with certain human diseases for diagnostic, prognostic and therapeutic purposes.

Our research group is currently working on the development on a more general multi-stain separation and normalization approach both for histochemical and immunohistochemical staining and, for this purpose, the proposed ICD algorithm can be involved in a future work.

## References

- [1] V. Varone, C. Bellevicine, and G. Troncone, "Biopsic sampling (cancer)," in *Comprehensive Sampling and Sample Preparation*, vol. 1, Elsevier Inc., 2012, pp. 413–439.
- [2] J. A. Ramos-Vara and M. A. Miller, "When Tissue Antigens and Antibodies Get Along: Revisiting the Technical Aspects of Immunohistochemistry—The Red, Brown, and Blue Technique," *Vet. Pathol.*, vol. 51, no. 1, pp. 42–87, Oct. 2013, doi: 10.1177/0300985813505879.
- [3] E. Gudlaugsson *et al.*, "Comparison of the effect of different techniques for measurement of Ki67 proliferation on reproducibility and prognosis prediction accuracy in breast cancer," *Histopathology*, vol. 61, no. 6, pp. 1134–1144, Dec. 2012, doi: 10.1111/j.1365-2559.2012.04329.x.
- [4] R. Schwarting, J. Gerdes, J. Niehus, L. Jaeschke, and H. Stein, "Determination of the growth fraction in cell suspensions by flow cytometry using the monoclonal antibody Ki-67," *J. Immunol. Methods*, vol. 90, no. 1, pp. 65–70, Jun. 1986, doi: 10.1016/0022-1759(86)90384-4.
- [5] A. Zacchetti, E. van Garderen, E. Teske, H. Nederbragt, J. H. Dierendonck, and G. R. Rutteman, "Validation of the use of proliferation markers in canine neoplastic and non-neoplastic tissues: comparison of KI-67 and proliferating cell nuclear antigen (PCNA) expression versus in vivo bromodeoxyuridine labelling by immunohistochemistry.," *APMIS*, vol. 111, no. 3, pp. 430–438, Mar. 2003, doi: 10.1034/j.1600-0463.2003.t01-1-1110208.x.
- [6] L. T. Li, G. Jiang, Q. Chen, and J. N. Zheng, "Ki67 is a promising molecular target in the diagnosis of cancer (Review)," *Mol. Med. Rep.*, vol. 11, no. 3, pp. 1566–1572, 2015, doi: 10.3892/mmr.2014.2914.
- [7] A. J. Norton, S. Jordan, and P. Yeomans, "Brief, high-temperature heat denaturation (pressure cooking): a simple and effective method of antigen retrieval for routinely processed tissues.," *J. Pathol.*, vol. 173, no. 4, pp. 371–379, 1994, doi: 10.1002/path.1711730413.
- [8] G. Coggi, P. Dell'Orto, and G. Viale, "Avidin-biotin methods," *Immunocytochemistry. Mod. methods Appl.*, pp. 54–70, 1986.
- [9] R. C. GRAHAM and M. J. KARNOVSKY, "THE EARLY STAGES OF ABSORPTION OF INJECTED HORSERADISH PEROXIDASE IN THE PROXIMAL TUBULES OF MOUSE KIDNEY: ULTRASTRUCTURAL CYTOCHEMISTRY BY A NEW TECHNIQUE," *J. Histochem. Cytochem.*, vol. 14, no. 4, pp. 291–302, Apr. 1966, doi: 10.1177/14.4.291.
- [10] C. R. Taylor and L. Rudbeck, *Educational IHC guidebook: immunohistochemical staining methods*. Dako, 2013.
- [11] C. E. DeSantis *et al.*, "Breast cancer statistics, 2019.," *CA. Cancer J. Clin.*, vol. 69, no. 6, pp. 438–451, Nov. 2019, doi: 10.3322/caac.21583.
- [12] S. M. Veronese, M. Gambacorta, O. Gottardi, F. Scanzi, M. Ferrari, and P. Lampertico, "Proliferation index as a prognostic marker in breast cancer.," *Cancer*, vol. 71, no. 12, pp. 3926–3931, 1993, doi: 10.1002/1097-0142(19930615)71:12<3926::aid-cnrc2820711221>3.0.co;2-2.
- [13] C. M. van der Loos, "Multiple Immunoenzyme Staining: Methods and Visualizations for the Observation With Spectral Imaging," *J. Histochem. Cytochem.*, vol. 56, no. 4, pp. 313–328, Dec. 2007, doi: 10.1369/jhc.2007.950170.
- [14] L. Wind and W. W. Szymanski, "Quantification of scattering corrections to

- the Beer-Lambert law for transmittance measurements in turbid media,” *Meas. Sci. Technol.*, vol. 13, no. 3, pp. 270–275, 2002, doi: 10.1088/0957-0233/13/3/306.
- [15] A. C. Ruifrok and D. A. Johnston, “Quantification of histochemical staining by color deconvolution,” *Anal. Quant. Cytol. Histol.*, vol. 23, no. 4, pp. 291–299, Aug. 2001.
- [16] M. Macenko *et al.*, “A method for normalizing histology slides for quantitative analysis,” in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009, pp. 1107–1110, doi: 10.1109/ISBI.2009.5193250.
- [17] A. Vahadane *et al.*, “Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images,” *IEEE Trans. Med. Imaging*, vol. 35, no. 8, pp. 1962–1971, 2016, doi: 10.1109/TMI.2016.2529665.
- [18] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, J. Shi, and C. Xue, “Adaptive color deconvolution for histological WSI normalization,” *Comput. Methods Programs Biomed.*, vol. 170, pp. 107–120, Mar. 2019, doi: 10.1016/J.CMPB.2019.01.008.
- [19] M. Gavrilovic *et al.*, “Blind Color Decomposition of Histological Images,” *IEEE Trans. Med. Imaging*, vol. 32, no. 6, pp. 983–994, 2013, doi: 10.1109/TMI.2013.2239655.
- [20] J. A. van Der Laak, M. M. Pahlplatz, A. G. Hanselaar, and P. C. de Wilde, “Hue-saturation-density (HSD) model for stain recognition in digital images from transmitted light microscopy,” *Cytometry*, vol. 39, no. 4, pp. 275–284, Apr. 2000.
- [21] N. Otsu, “A Threshold Selection Method from Gray-Level Histograms,” *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66, 1979, doi: 10.1109/TSMC.1979.4310076.
- [22] M. Saha, C. Chakraborty, and D. Racoceanu, “Efficient deep learning model for mitosis detection using breast histopathology images,” *Comput. Med. Imaging Graph.*, vol. 64, pp. 29–40, 2018, doi: <https://doi.org/10.1016/j.compmedimag.2017.12.001>.
- [23] M. Salvi, N. Michielli, and F. Molinari, “Stain Color Adaptive Normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology,” *Comput. Methods Programs Biomed.*, vol. 193, p. 105506, 2020, doi: <https://doi.org/10.1016/j.cmpb.2020.105506>.
- [24] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, “Color transfer between images,” *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34–41, 2001, doi: 10.1109/38.946629.
- [25] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee, “A Nonlinear Mapping Approach to Stain Normalization in Digital Histopathology Images Using Image-Specific Color Deconvolution,” *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1729–1738, 2014, doi: 10.1109/TBME.2014.2303294.
- [26] J. N. Kather, C. A. Weis, A. Marx, A. K. Schuster, L. R. Schad, and F. G. Zöllner, “New Colors for Histology: Optimized Bivariate Color Maps Increase Perceptual Contrast in Histological Images,” *PLoS One*, vol. 10, no. 12, pp. e0145572–e0145572, Dec. 2015, doi: 10.1371/journal.pone.0145572.
- [27] P. Shi, J. Zhong, J. Hong, R. Huang, K. Wang, and Y. Chen, “Automated Ki-67 Quantification of Immunohistochemical Staining Image of Human

- Nasopharyngeal Carcinoma Xenografts,” *Sci. Rep.*, vol. 6, no. 1, p. 32127, 2016, doi: 10.1038/srep32127.
- [28] H. Sharma, N. Zerbe, I. Klempert, O. Hellwich, and P. Hufnagl, “Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology,” *Comput. Med. Imaging Graph.*, vol. 61, pp. 2–13, 2017, doi: <https://doi.org/10.1016/j.compmedimag.2017.06.001>.

# Conclusions and Final Remarks

The aim of this thesis is the development and validation of signal and image decomposition techniques for biomedical applications. The proposed approaches aim to overcome the limitation of time-consuming manual processes which are also subjected to intra- and inter-operator variability, through the development of CAD systems which become the bridge technology that enables the effective extraction of quantitative data from biomedical signals and medical images.

In the first part, the process of falling asleep which involves gradual changes between the relaxed waking state and deep sleep, was addressed. The proposed approach based on RNN architecture obtained the highest value in the detection of N1 and N3 sleep with respect to other single-channel based approaches and at the same time satisfactory performances in the other sleep stages. In the near future data augmentation techniques will be employed to address class imbalance. The following unresolved issue emerged in this context: single-channel vs multichannel approach. From one side, with the use of single-channel EEG signals remains very difficult to improve the performances of sleep stages due to their heterogeneity, but the acquisition setup is simpler with less noise interference and it may be used for NCP assessment in real-time applications (e.g. the detection of driver drowsiness). On the other side, multichannel recording can be used only as a support for clinical diagnosis but the performances in sleep scoring may be improved. In a future study the proposed workflow will be used by taking into account multiple EEG channels, EOG and EMG signals in order to understand their robustness. In addition, signal decomposition techniques can also be evaluated for the analysis and interpretation of other biomedical signals (e.g. ECG, heart rate variability or evoked potentials) in a wide range of applications.

In the second part, the problem of stain color variability in histological and histopathological tissue slides was addressed. The proposed method of stain separation and normalization was validated on a multi-tissue dataset of H&E stained images and achieved better performances than other published methods. The same problem was addressed in IHC with greater complexity due to the scattering nature of the staining dyes. In IHC, the algorithm was validated on a small dataset, belonging to a single tissue and for this reason it is considered a preliminary study. The aim of the stain normalization strategy is twofold: firstly the standardization of stain color appearance in digital pathology could improve the pathologist's work in the diagnosis of biological diseases, avoid manual re-staining process and reduce intra- and inter-operator variability; secondly this method will be used, in the near future, as a pre-processing step for subsequent CAD systems for accurate cellular structure segmentation and classification based on deep learning framework.

# List of Contributions

## Publications

- N. Michielli, U. R. Acharya and F. Molinari. "**Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals.**" *Computers in Biology and Medicine* 106 (2019): 71-81.
- M. Salvi, N. Michielli and F. Molinari. "**Stain Color Adaptive Normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology.**" *Computer Methods and Programs in Biomedicine* 193 (2020): 105506.
- K. M. Meiburger, A. Naldi, N. Michielli, L. Coppo, K. Fassbender, F. Molinari and P. Lochner. "**Automatic Optic Nerve Measurement: A New Tool to Standardize Optic Nerve Assessment in Ultrasound B-Mode Images.**" *Ultrasound in Medicine & Biology* 46.6 (2020): 1533-1544.
- A. Mazza, S. Colombo, N. Michielli, P. Sarasso, F. Montagna, O. Dal Monte and R. Ricci. "**EEG and eye-tracking correlates of divergent thinking**", paper in preparation.

## Patent and Proof of Concept (PoC)

- F. Molinari, M. Salvi and N. Michielli, **System for processing an image relating to a histological tissue.** Italian patent application on 24/03/2020, number: 102020000006148. International patent application (PCT) on 15/03/2021, number: PCT/IB2021/052119.
- F. Molinari, M. Salvi, O. Pennisi and N. Michielli, **STAINS - STANDARDIZATION & NORMALIZATION OF HISTOLOGICAL SLIDES.** The project is part of the PoC Instrument (cut-off 1) - LINKS Foundation initiative, with LIFTT support, based on funding from the Compagnia di San Paolo Foundation. Start date: 15/07/2020, duration of the project: 9 months.
- F. Molinari, M. Salvi, O. Pennisi and N. Michielli, **ASSIST - Automated Structure-based Segmentation method for the Identification and Staging of Tumor aggressiveness.** The project is part of the "PoC-off" program, based on funding from the Italian Ministry of Economic Development (MISE), patent and trademark office. Start date: 15/02/2021, duration of the project: 15 months.

## Conference Papers

- N. Michielli, S. Seoni, B. De Santi, M. Salvi, K. M. Meiburger, A. Iadarola, A. Cicolin and F. Molinari. "**Automatic sleep stage classification: a step**

- forward to automated assessment of neurocognitive performance."** *6<sup>th</sup> National Congress of Bioengineering*, June 25<sup>th</sup>-27<sup>th</sup>, 2018, Politecnico di Milano, Milan, Italy.
- M. Salvi, K. M. Meiburger, B. De Santi, N. Michielli, L. Molinaro and F. Molinari. "**Automated Gleason grading in prostate cancer histopathology images.**" *6<sup>th</sup> National Congress of Bioengineering*, June 25<sup>th</sup>-27<sup>th</sup>, 2018, Politecnico di Milano, Milan, Italy.
  - B. De Santi, M. Salvi, N. Michielli, K. M. Meiburger, N. Bonelli, R. Rossetto, R. Garberoglio and F. Molinari. "**Ultrasound thyroid nodule texture-based classification with artificial neural networks.**" *6<sup>th</sup> National Congress of Bioengineering*, June 25<sup>th</sup>-27<sup>th</sup>, 2018, Politecnico di Milano, Milan, Italy.
  - K. M. Meiburger, M. Salvi, N. Michielli, B. De Santi, M. A. Minetto and F. Molinari. "**The role of beamforming in quantitative muscle ultrasonography: a texture analysis approach.**" *6<sup>th</sup> National Congress of Bioengineering*, June 25<sup>th</sup>-27<sup>th</sup>, 2018, Politecnico di Milano, Milan, Italy.
  - B. De Santi, M. Salvi, V. Giannini, K. M. Meiburger, N. Michielli, S. Seoni, D. Regge and F. Molinari. "**Multimodal T2w and DWI Prostate Gland Automated Registration.**" *2019 41<sup>st</sup> Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019.
  - K. M. Meiburger, F. Veronese, V. Tarantino, M. Salvi, M. Fadda, S. Seoni, E. Zavattaro, B. De Santi, N. Michielli, P. Savoia and F. Molinari. "**Automatic Extraction of Dermatological Parameters from Nevi Using an Inexpensive Smartphone Microscope: A Proof of Concept.**" *2019 41<sup>st</sup> Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019.
  - S. Seoni, F. Veronese, V. Tarantino, E. Zavattaro, M. Salvi, N. Michielli, B. De Santi, F. Molinari, P. Savoia and K. M. Meiburger. "**Non-invasive analysis of actinic keratosis using a cold stimulation and near-infrared spectroscopy.**" *2019 41<sup>st</sup> Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019.
  - M. Fusella, M. Vagni, C. Fiandra, N. Michielli, A. Scaggion, C. Vecchi, S. Zara, F. Molinari and G. Loi. "**Inverse Consistency Error as a validation metric for Deformable Image Registration: preliminary implementation research**", submitted to *3<sup>rd</sup> European Congress of Medical Physics (ECMP)*, 2021.
  - M. Fusella, M. Vagni, C. Fiandra, N. Michielli, A. Scaggion, C. Vecchi, S. Zara, F. Molinari and G. Loi. "**Inverse Consistency Error for quantifying uncertainty in Deformable Image Registration: validation on three different sites**", submitted to *European Society for Radiotherapy and Oncology (ESTRO) Congress*, 2021.