Intelligent Frameworks for Diagnosis in the Precision Medicine Era

Department of Electrical and Information Engineering
ELECTRICAL AND INFORMATION ENGINEERING
Ph.D. Program

SSD: ING-INF/06 - ELECTRONIC AND INFORMATION BIOENGINEERING

Final Dissertation

Intelligent Frameworks for Diagnosis in the Precision Medicine Era

---

by

Brunetti Antonio

Supervisor:
Prof. Vitoantonio Bevilacqua, Ph.D.

Coordinator of Ph.D. Program:
Prof. Luigi Alfredo Grieco, Ph.D.

Course n°32, 01/11/2016 - 31/10/2019

Department of Electrical and Information Engineering
ELECTRICAL AND INFORMATION ENGINEERING
Ph.D. Program

SSD: ING-INF/06 - ELECTRONIC AND INFORMATION BIOENGINEERING

Final Dissertation

Intelligent Frameworks for Diagnosis in the Precision Medicine Era

---

by

Brunetti Antonio

Referees:
Prof. Stefano Cagnoni, Ph.D.
Prof. Mario Cesarelli

Supervisor:
Prof. Vitoantonio Bevilacqua, Ph.D.

Coordinator of Ph.D. Program:
Prof. Luigi Alfredo Grieco, Ph.D.

Course n°32, 01/11/2016 - 31/10/2019

To my wife
To my family

Acknowledgements

Undertaking my Ph.D. was an outstanding experience that really changed my life, both from professional and personal point of views. The study and the activities done in these three years have made this period of my life really intense, and it would not have been possible to achieve the results obtained thanks to the support and guidance I received from many people.

I want to thank my Ph.D. supervisor, Professor Vitoantonio Bevilacqua, for continually supporting me in these years. He is one of the smartest and brilliant people I know. Since my master thesis, I have had the opportunity to get to know this fantastic person. He has been a guide always present and available in times of need, always ready to give you new stimuli to do better and better, capable of making express yourself as best as possible. He is a person who is not forgotten once he is known. I hope to be as lively, enthusiastic, energetic and resourceful as he is.

A grateful thank to my Ph.D. referees, who have been able to appreciate the work done over the years and gave me important suggestions to improve the clarity and presentation of my thesis.

A huge thank to my colleagues (friends) of the Industrial Informatics Lab, both those currently present and those now geographically distant. Daily and constant presences. A cohesive group, always ready to help each other in times of difficulty, always available to confront constructively. Without their constant support and work, it would not have been possible to achieve certain objectives, both scientific and professional. I am sure that together we are doing, and will continue to do, important things.

Many thanks also to the people who collaborated with me and my research group during these years, especially from Polyclinic of Bari; correct and helpful people, respectable researchers.

I especially thank my family. My hard-working parents have sacrificed their lives for me and my brother. Despite being a bit absent these years, they have always been there for me, always available to help me and give me their support for whatever I need.

Special thanks to my wife's wonderful family who all have been supportive and caring.

Finally, the biggest thanks to my wife Amalia, my lifelong companion. She made many sacrifices to allow me to achieve this goal. Her love and unconditionally support let me go forward these years, making me recover in bad moments and rejoice in good periods. She has always shared my successes and failures, always by my side, supporting me in all my choices and standing by me every day of my life. I couldn't have wished for a better person by my side in life.

Abstract

This Ph.D. thesis aims to describe all the research works conducted for the design, development and evaluation of intelligent frameworks for supporting the diagnosis in several clinical units of the healthcare system with respect to the precision medicine approach.

The presented work focuses on the innovative aspects of Computer-Aided Diagnosis Systems employed in the radiological area and pathological field, and on Decision Support Systems in the neurological and psychological area. Systems for supporting the physiatrics activities for performing and supporting the rehabilitation of people affected by neurodegenerative diseases are also investigated and discussed.

The first chapter introduces the importance of having an objective assessment of pathologies by the support of automatic systems in order to reduce both diagnostic errors and the time to obtain an accurate diagnosis. The importance of Decision Support Systems for precision medicine is also presented by describing the advantages of using their clinical outcomes, integrated with radiogenomics data, or genetic information in general, for performing targeted therapies.

In the second chapter, there is a description of Computer-Aided Diagnosis systems by introducing the traditional workflow historically implemented by such systems; subsequently, a classification of these systems is described from the Machine Learning perspective. Afterwards, the novelty from using Deep Learning methodologies for supporting diagnosis is described, also detailing the way of working of these new classifiers. Finally, the last section describes the contribution of using Virtual Reality for supporting diagnosis and rehabilitation, including cognitive and physiological assessment and implementation of protocols allowing the de-hospitalization of rehabilitation treatments.

The third chapter introduces all the research contributions for improving the state of the art about Decision Support Systems in the clinical areas of radiology and pathology. Innovative and intelligent frameworks for the radiological assessment are described, including both classification and segmentation approaches. Specifically, chapter three reports the research contribution about the design and development of Computer-Aided Diagnosis systems for breast cancer diagnosis, liver tumour staging and segmentation of kidneys affected by

Autosomal Dominant Polycystic Disease. The pathological support, instead, is detailed in nephrological and haematological areas, describing both Machine Learning and Deep Learning approaches. Specifically, novel approaches for the segmentation of vessels and tubules in kidney biopsies and for automatically counting white blood cells are investigated.

The fourth chapter describes the works for designing and developing intelligent frameworks for assessing different neurophysiological disorders, including Alzheimer's Disease and Fibromyalgia, or supporting elderly people in their living environments for improving their living conditions. Finally, a research work about Machine Learning methodologies for support staging the Parkinson's Disease progression is also described.

Chapter five reports the conclusions, highlighting the great importance of integrating these intelligent frameworks in the clinical practice, combining information coming from different sources, in order to proceed toward the real new era of precision medicine.

Table of contents

List of figures

List of tables

Chapter 1

On the Importance of an Accurate Clinical Diagnosis

Health care delivery has been going on with a blind spot for decades: diagnosis errors and inaccurate, or delayed, results persist in all care settings and affect an excessive number of patients. Getting the right diagnosis, also on time, is a crucial aspect of health care, as it allows to describe the health status of a person and influences the future clinical decisions about the treatment of the pathology affecting a subject [1].

Errors in diagnosis can result in negative health outcomes, psychological distress, and financial costs. If a medical error occurs, a patient may receive inadequate or unnecessary cares, or potentially life-saving treatments may be postponed or delayed. To date, however, efforts have been mostly limited to systematically recognising and minimising diagnostic errors. In the absence of a spotlight to highlight this critical problem, the reliability and patient safety campaigns have generally failed to appreciate diagnostic errors. The effect of this inattention is important: many people are likely to experience at least one diagnostic error in their lifetime, with probably devastating consequences for all the actors of the healthcare system.

The topic of clinical diagnosis raises a number of clinical, personal, cultural, ethical, and even political issues that commonly capture the public interest. Members of the general public are concerned about diagnosis, and many have reported experiencing diagnostic errors. Recent data analysis and literary reviews found that medical errors could be identified as the third leading cause of death in the US, and approximately half of these errors were diagnostic errors [2, 3].

Despite a large number of deaths due to diagnostic errors, there is still limited awareness of the possibility of reducing mortality and morbidity by investing in improving diagnostic accuracy and speed [4].

1.1 Precision Medicine

The advances in Information Technology (IT) and healthcare performed in the last decades allowed to create more Electronic Medical Records (EMR); in fact, seventy to eighty per cent of the digitalised and structured clinical information includes laboratory data and biomedical signals [5]. The likelihood of performing a diagnosis using information from one or more of the associated diagnostic areas of anatomic pathology and radiology is extremely high in industrialised countries. In fact, recent advances made information readily available thanks to more efficient devices and, in some cases, low costs for acquiring the biomedical signals [4].

Also, recent biotechnological advances have led to an explosion of disease-relevant molecular information, with the potential for greatly advancing patient care. Coupling established clinical-pathological indexes with state-of-the-art molecular profiling to create diagnostic, prognostic, and therapeutic strategies precisely tailored to each patient's requirements lead to the so-called *Precision Medicine* (PM) [6].

According to the Precision Medicine Initiative (PMI), a long-term research endeavour, involving the National Institutes of Health (NIH) and multiple other research centres, *Precision Medicine* is "an emerging approach for disease treatment and prevention that takes into account each person's variability in genes, environment, and lifestyle" [7].

This modern approach will allow doctors, researchers, and clinical staff, to predict more accurately which treatment and prevention strategies for a particular disease will work better, and in which group of people. This way to consider the patient as a unique individual, with its own specificity, is in contrast to the currently used one-size-fits-all approach. In fact, traditionally, the disease treatments, or disease prevention strategies, are studied and developed for the average person, with less consideration for the specific differences between individuals.

Precision medicine is a young and growing field, even though it is not a new concept. There are some precursor approaches, or rather primitive forms, of precision medicine, e.g. blood typing has been using to guide blood transfusions for more than a century. However, many of the technologies needed to meet the goals of the PMI currently are in the early stages of development, or, unfortunately, have not yet been developed.

The Precision Medicine Initiative also raises ethical, social, and legal issues. It will be critical to find ways to protect participants' privacy and the confidentiality of their health information. Participants will need to understand the risks and benefits of participating in such research, which means researchers will have to develop a rigorous process of informed consent.

Cost is also an issue with precision medicine, including the management, technological and development costs. In fact, drugs developed to target a person's genetic or molecular characteristics are likely to be expensive. Reimbursement from third-party payers (such as private insurance companies) or National Healthcare System for these targeted drugs is also likely to become an issue to be handled by specific laws. In addition, the design and development of strategies for implementing targeted diagnosis and therapies will require incredible efforts by the scientific community in doing research in this sense, also requiring the collaboration with industrial partners for developing performing processing hardware.

Finally, if precision medicine approaches are to become part of routine healthcare, doctors and other healthcare providers will need to know more about molecular genetics and biochemistry, thus they need to study and train in this sense. They will increasingly find themselves needing to interpret the results of genetic tests, understand how that information is relevant to treatment or prevention approaches, and convey this knowledge to patients.

1.2 Improving Diagnosis in Specific Clinical Settings

Although it has a noble goal aimed at improving the living conditions of people with different diseases, to date precision medicine has a series of limitations that, in practice, limit its usage in the current clinical practice. Its current applications, in fact, are very experimental and used only in cases where traditional clinical treatment has failed.

Much research is in progress, but even more must still be done in this regard, especially in the field of bioinformatics for the identification of specific and accurate tumour markers, also allowing to reduce the invasiveness of such diagnosing. In fact, the key concepts of *radiomics* and *molecular imaging* are becoming more and more widespread in research contexts, thanks to the efforts of several research groups [8, 9].

Radiomics refers to the process of extracting mineable, high-dimensional data from the routine, standard-of-care, Computed Tomography (CT), Magnetic Resonance (MR), Positron Emission Tomography (PET) images, or from other imaging techniques, using automatic or semi-automatic algorithms for computing and extracting characterising features [10, 11]. Along with genomics, also radiomics shows great promise in the era of precision medicine.

Quantitative imaging data, e.g. imaging features extracted after the tumours segmentation, such as shape, size, volume, signal intensity, CT attenuation, maximum standardised uptake value, together with CT and MR textural analysis, are valuable information for supporting clinical diagnoses. Recent studies, in fact, reported that such information might constitute innovative biomarkers in tumour prognosis, thus predicting the response to therapy and the patient outcome [12–18].

Molecular imaging, defined as visualisation, characterisation, and measurement of biological processes at the molecular and cellular levels in humans and other living systems, is elegantly poised to become an invaluable tool in the era of precision medicine [19]. By enabling disease imaging at the cellular level, molecular imaging may help to identify disease in preclinical states, classify which group of patients might or not benefit from a particular targeted therapy, and accurately evaluate response to therapy. Numerous endogenous molecules and exogenous molecular imaging agents are currently available, including radio-labelled, fluorescently labelled, and nanoparticle-based molecular imaging probes.

Image-guided biopsies will play an increasing role in precision medicine, not only for the initial diagnosis of a disease but also in the evaluation of treatment resistance. Tissue from selectively targeted biopsies will provide substrates for genetic and molecular characterisation. Obtaining such up to now unavailable genetic information might have a positive impact on the pursuit of individualised therapy.

The following sections will investigate the importance of a quantitative assessment of diseases in different clinical fields, namely radiology, pathology, neurology and physiatry. In will be analysed how quantitative and objective measurements may be fundamental for the new era of precision medicine, in order to extract fundamental information for the genetic characterisation of diseases, thus leading to even more patient-targeted diagnosis and therapeutic interventions.

1.2.1 Radiology

Imaging will play a pivotal role in precision medicine for different aspects, including screening, early diagnosis, guiding treatment, evaluating response to therapy, and assessing the likelihood of disease recurrence [20]. For precision medicine to succeed, it is critically essential that imaging be able to help identify and classify patients in different subgroups that have identical disease characteristics and share similar responses to treatments and prognosis. This new precision will require an ever better characterisation of patients and their diseases through new applications of genomics and improved methods of phenotyping, to define

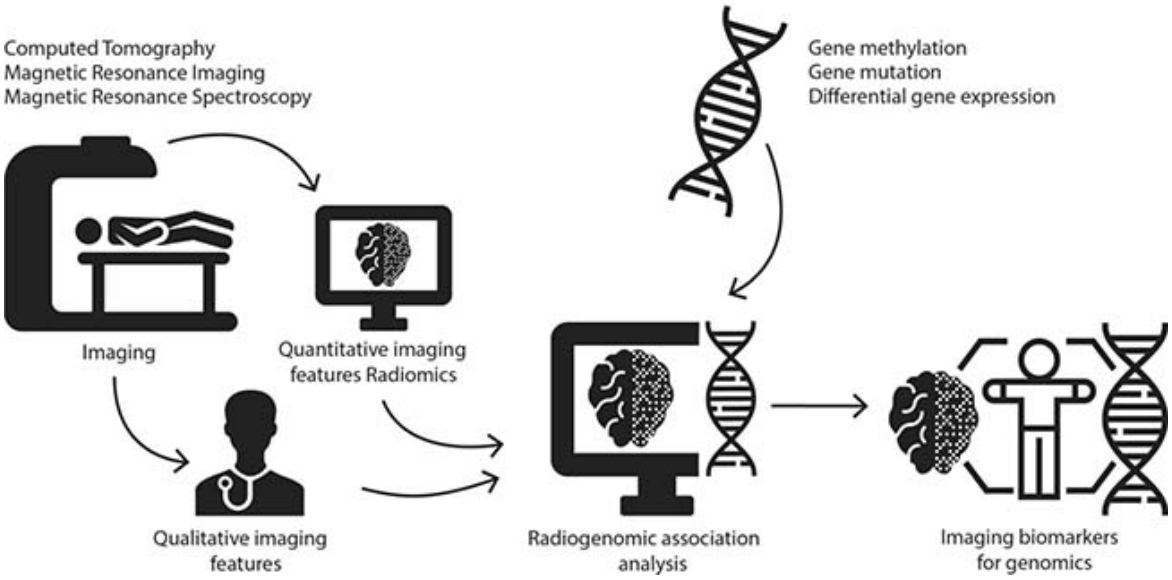


Fig. 1.1 A representation of the radiogenomics approach in the precision medicine context. Image from [23].

progressively smaller sub-populations of patients with similar disease characteristics who would benefit from similar therapies [21].

To realise this vision, all disciplines of medicine must work together to develop a more precise taxonomy of disease rooted in molecular biology. Biomedical imaging speaks directly to this requirement by virtue of its ability to provide a multi-parametric characterisation of disease and represents a key component in the infrastructure supporting the concept of precision medicine.

In this sense, *radiogenomics* assumes a key position in the context of precision medicine. Although the term radiogenomics is perceived by radiation oncologists to refer to the study of the correlation of genetic variation with the response to radiation therapy, it has a different meaning in the radiology community. In radiology, the term radiogenomics (also called *imaging genomics*) refers to the correlation of imaging phenotypes with genotypic expressions [22]. Radiogenomic studies that help determine statistically significant linkages between imaging features and gene expressions may help create models that predict patient outcomes based on imaging features (Fig. 1.1). Radiogenomics has already attracted major interest in the radiology community, with research undertaken in various cancers such as glioblastoma, breast carcinoma, and renal cell carcinoma [23].

Historically, imaging has provided morphologic information about diseases by localising, measuring, and characterising abnormalities. Currently, the increasingly diverse arsenal of

modern imaging tools and imaging-based biomarkers are valuable instruments needed to define the imaging phenotype of the disease. The integration of functional, molecular and metabolic information provided by radiologic techniques has greatly increased the ability to segment patient populations into phenotypic subsets that share similar prognoses and are likely to respond similarly to therapies. Classification and scoring in imaging studies will likely provide the phenotypic information necessary to improve clinical trial design and assess therapy response.

An impressive example of the potential value of imaging phenotyping for patient care consists of using a scoring system for ranking the severity of intracranial haemorrhage. Romero *et al.* graded intracranial haemorrhage based on the appearance of contrast material within the site of active haemorrhage [24]. The authors also revealed that the grade generated by their "spot sign" scoring system, which can also be thought of as a phenotyping system, was associated with a risk for haematoma expansion and poor clinical outcome [24]. With this system, they were able to define a subpopulation that was most likely to benefit from engagement in a specific clinical trial, allowing non-members of this subpopulation to be spared the adverse effects of therapy [25]. Thus, imaging phenotyping provided the filtering mechanism to enhance the design of the clinical trial and improve patients care.

Precision medicine and precision imaging necessitate the development of interactive clinical Decision Support Systems (DSS), or Computer-Aided Diagnosis (CAD) systems, that assist physicians with decision making at the point of care. The increasing utilisation of these systems will be catalysed by the implementation and standardisation of Electronic Health Records (EHR), including both advantages and disadvantages. In fact, integrating data and outcomes from intelligent systems into the EHR could improve the quality of care, reduce medical errors, and lead to other improvements in patient - level measures that describe the appropriateness of care. However, some authors have identified potential disadvantages associated with this technology. These include financial issues, changes in workflow, temporary loss of productivity associated with EHR adoption, privacy and security concerns, and several unintended consequences [26].

Given the volume and complexity of data that are now available, and the expectation of exponential increases in available data in the future, decision support tools designed to intelligently filter patient data are already essential for optimising the entire clinical workflow [6].

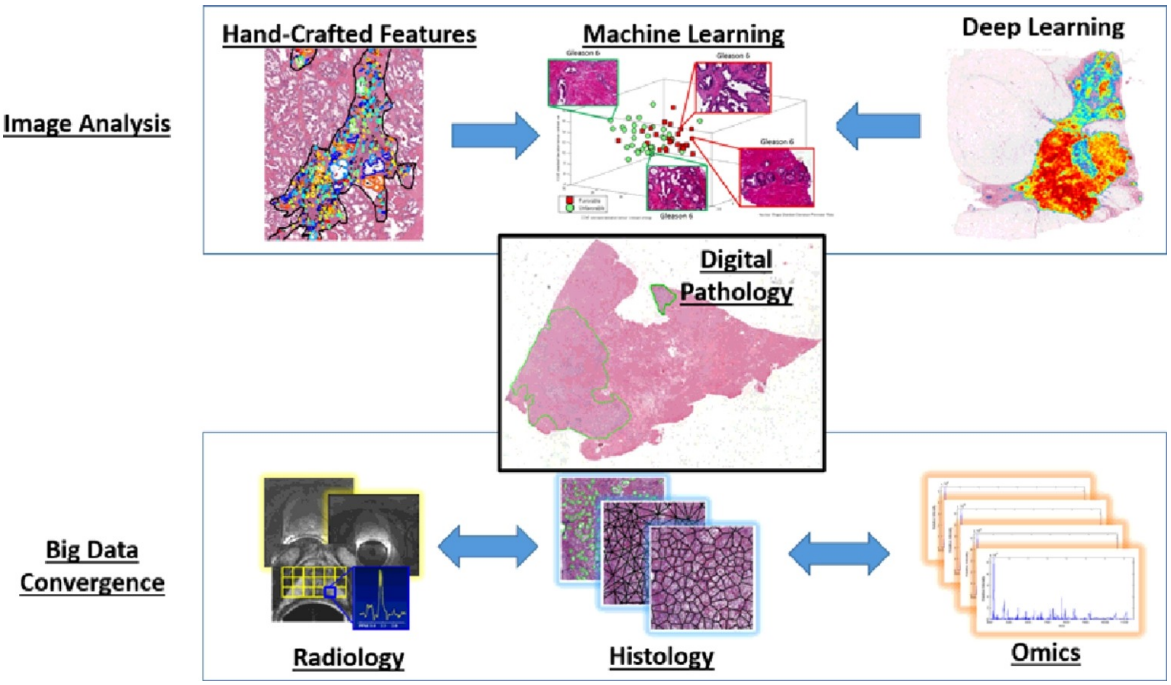


Fig. 1.2 Integration between radiological and pathological outcomes to correlate genotypes and phenotypes.

1.2.2 Pathology

Like radiology, pathology is a crucial area for precision medicine. Oncologists need the correct histological diagnosis and information on underlying gene mutations to offer the right personalised medication. Pathologists play a vital role in this respect since they make the diagnosis and also offer information about the mutations present in a tumour. Also, pathology is strictly linked to radiology, as they contribute to inferring more accurate and significant relationships among genotypes and phenotypes.

Over the last decade, the advent and subsequent proliferation of whole slide digital slide scanners have resulted in a substantial amount of clinical and research interest in digital pathology, the process of digitisation of tissue slides (Fig. 1.2) [27].

Digitisation of tissue glass slides facilitates telepathology, i.e. using the telecommunication technology to facilitate the transfer of image-rich pathology data between distant locations for research, diagnosis, and education purposes. Telepathology can now be enabled more easily and seamlessly between multiple different remote sites and more critically connecting top academic centres with pathology labs in low-resource settings. Telepathology is particularly convenient for solicitation of "second opinions" on challenging cases or the ability to do remote pathology consults without the need to ship the slides around physically.

Digital pathology also has the potential to help improve clinical workflows, reducing the need for storing glass slides on-site and reducing the risk of physical slides getting broken or lost over time. Additionally, access to large digital repositories of tissue slides is a huge-potential educational resource for medical students and pathology residents. Further, in conjunction with smart analytics like content-based image retrieval algorithms [28], students could be trained to identify and recognise pathology slides in a dynamic fashion.

Apart from the education of medical students and residents, and clinical adoption, of course, digital pathology has been transformative for computational imaging research. With digital slide archives, such as *The Cancer Genome Atlas* (TCGA, https://www.cancer.gov/tcga), several thousands of slides are now freely available online; this resulted in a very large number of digital pathology image analysis related publications, confirming the growing interest in this research area. Also, computer-based image analysis is already available in commercial diagnostic systems, but further advances in image analysis algorithms are required in order to fully realise the benefits of digital pathology in medical discovery and patient care [29].

The abilities to quantitatively characterise disease classification and to process from multiple biological scales and dimensions have the potential to enable the development of preventive strategies and medical treatments precisely targeted to each class of patients, implementing the hoped for precision medicine. The advances in pathology imaging technologies, along with comparably dramatic advances in the "-omics" and radiology domains, are revolutionising the medical professionals' ability to rapidly capture and exploit vast amounts of multi-scale, multi-dimensional data on each patient's genetic background, biological function, and structure.

In the past, pathologists classified tissues by manually-recognised patterns. In recent years, many researchers demonstrated that, in some cases, image analysis algorithms could reproduce the pathologist-rendered diagnoses [30]. High-resolution and high-throughput instruments are being employed routinely at an accelerating rate, not only in medical science but also in healthcare delivery settings. As this decade progresses, significant advances in medical information technologies will transform very large volumes of multi-scale and multi-dimensional data into actionable information to drive the discovery, development, and delivery of new mechanisms of preventing, diagnosing and healing complex diseases. In the current practice, molecular information and human-generated pathology interpretation guide the treatments' administration for handling several diseases. Tumour subtypes have crucial treatment implications and play increasingly crucial roles in the development of successful

targeted therapeutic regimens. For instance, in current neuro-oncologic practice, pathology classification and molecular subtyping are used together to guide the choice of treatment.

1.2.3 Neurology and Psychology

The application of the precision medicine paradigm to the treatment and prevention of neurodegenerative disorders appears to be highly promising in contrast to the traditional "one-drug-fits-all" approach. In fact, neurodegenerative pathologies can present variable clinical features even in patients with the same disease who therefore are very unlikely to benefit from assuming single drug. In this context, the development of a precision medicine approach could represent an excellent possibility to identify preclinical stages of the disease, allowing making an adequate differential diagnosis and providing timely and optimal treatments instead of the traditional ones, normally utilised at a later stage of the disease [31–33].

Genes, neuroepigenetic modifiers, non-genetic factors (dietary, smoking habits, physical/mental exercise, microbiome) and drugs impact the susceptibility to neurodegenerative disorders. Prominent attention should be given to the dynamics of neuroepigenetic changes occurring at the inter- and intra-individual levels. In fact, neuroepigenetic features are characterised by a high plasticity degree during the lifespan which ultimately modulates the function of specific genes in response to ageing and specific environmental pressure, influencing thereby the likelihood to develop neurodegenerative disorders [34]. The combination of the overall data can be utilised to generate omic profiles and provide a complete overview of patients. The resulting omic profiles can be exploited by precision medicine to create a stratified medicine able to assign patients to specific treatment classes, such as high-, intermediate- or low-risk individuals, good, intermediate or poor responders, high-, intermediate- or low-dosage receivers. The translation of the theoretical concept of precision/stratified medicine into the clinical practice can be achieved by developing, or even accommodating, the computational facilities already in use to integrate the omic informations into a single algorithm, able to predict the patient's disease course and to support the clinical and therapeutic decisions toward a more participatory and preventive medicine [34, 35].

The availability of social networks allowing the simultaneous sharing of huge amounts of data across the world closed the gap due to geographical distance and difficulties in getting access to information about the psychological and neurological spheres. In this context, the realisation of a web-based network for neurodegenerative disorders can be decisive for the implementation of precision medicine strategies across different specialised centres. Such networks can contribute to accumulate and share the overall information derived from

the combination of patients' participation (*participative medicine*) and the broad medical expertise provided by physicians. On this subject, an excellent example of a multidisciplinary, web-based platform is the Italian IRCSS Network of Neuroscience and Neurorehabilitation, whose aim includes standardising and optimising patients' clinical care and the therapeutic strategies applied to neurodegenerative disorders [36].

1.2.4 Physiatry

Nowadays, all the population knows the health benefits of a physically active lifestyle, from clinicians to ordinary people [37]. Such a lifestyle goes together with lower risks of morbidity and mortality from a great number of chronic diseases, such as coronary artery disease, diabetes, and colon cancer. In addition, for people with a physical disability, a physically active lifestyle could improve everyday functioning, reduce disability, and decrease the risk of secondary health problems [38–43]. However, people with disabilities are, in general, less physically active than people without inabilities [44].

The World Health Organization (WHO) defines disability as problems that an individual may experience in functioning. The relation between functioning, disability, physical activity, and determinants of physical activity is described in more detail in the "Physical activity for people with a disability", from Van der Ploeg *et al.* [45].

In 1997, only 12 % of US adults with a disability aged more than 17 participated in moderate physical activity for at least 30 minutes per occasion for at least five days per week, compared with the 16 % of people without a disability [44]. For leisure-time physical activity, the difference between people with and without disabilities is even larger: 56 % and 36 % respectively did not engage in leisure-time physical activity [44]. Consequently, persuading people with disabilities to become more physically active is probably even more important than it is for the general population.

For two reasons, rehabilitation provides an excellent opportunity to start promoting a physically active lifestyle. Firstly, for many people, rehabilitation is the start of learning to live with a disability. It may thus be an effective strategy to integrate physical activity into the new everyday routine after rehabilitation. Secondly, sports and other physical activities are often a component of rehabilitation programmes. It is probably easier to integrate these activities directly into everyday life than to become physically active when sedentary [45].

Physical rehabilitation of people affected by some diseases should also be personalised and tailored while allowing a patient to obtain objective indices for assessing the progress.

Although it may seem out of context with respect to the canonical meaning of precision medicine, objectifying the measurements and grades of disabilities, and addressing the right rehabilitation treatment according to the patient's physical capabilities, is fully part of the precision medicine scope.

1.3 Objective and Research Question

Based on preceding paragraphs, the adoption of precision medicine in the current clinical practice is far from being practically realised. In fact, despite the considerable effort made by the scientific community, which made the Computer-Aided Diagnosis (CAD) systems an established and rapidly growing field of research, the proposed architectures are not yet used in the current clinical practice, from a precision medicine perspective and a diagnostic one.

In fact, CAD systems have to meet and satisfy several demands to be used widely in everyday activities. According to literature, the four major requirements, especially for the radiologists' practice [46], are:

- CAD system should improve clinicians' performance;
- CAD systems should allow clinicians to save time;
- CAD systems must be seamlessly integrated into the workflow of the diagnostic procedures;
- CAD systems should not impose liability concerns and the incremental costs should be negligible or reimbursed, thus sustainable for the healthcare system.

Unfortunately, most CAD systems today do not meet all these requirements, and this is why most applications described in the rapidly growing body of scientific literature on CAD systems are not widely used in the clinical practice. The requirements mentioned above are only partially satisfied having sufficient performance, no increase in reading time, seamless workflow integration, regulatory approval, and efficiency in terms of costs. Performance is still the major bottleneck for many CAD systems due to unavailable performing processing systems or badly designed algorithms.

Novel ways of using CAD systems, extending the traditional paradigm of displaying markers for a second look, could be the key for using the technology effectively. Also, the most promising strategy to improve CAD development is the creation of publicly available databases for training and validating the algorithms. The solution to these open problems

can identify the most fruitful new research directions, and provide a platform to combine multiple approaches for a single task to create superior algorithms. Finally, the advent of Deep Learning strategies is leading diagnosis support to a profound change with respect to the traditional way of dealing with this task.

Also, considering the precision medicine scenario, accurate and reliable systems for supporting the clinical diagnosis are necessary for the current clinical practice and in all the fields included in the healthcare system. Moreover, Decision Support Systems (DSS) in medicine are also included into the Health Technology Assessment (HTA), which is a multidisciplinary approach for the systematic assessment of medical technologies regarding their effectiveness, appropriateness, efficiency, and social and ethical aspects and implications; thus they actually contribute to the global evaluation of the healthcare system [47]. Improving the performance of such systems is of fundamental importance.

Since CAD systems are shown to be compelling tools for supporting clinicians in performing more accurate diagnosis, the objective of this research work has been to investigate and develop innovative methodologies and approaches for further improving the state-of-the-art performance in such systems, focusing on both classification and segmentation tasks considering different body districts. Also, each methodology has been applied to real-world cases in order to test and validate the proposed algorithms working on biomedical signals.

Furthermore, the treatment of ageing-related disorders and neurodegenerative diseases need innovative solutions for supporting the objective assessing of the clinical course over time, making available systems for de-hospitalising the treatment and the monitoring of the disease.

1.4 Contribution

The main contributions of this thesis regard the analysis, design and development of innovative procedures to support diagnosis in order to allow the quantitative and objective assessment of pathologies, ranging from diagnosing in the radiological area to rehabilitation support using innovative devices.

Specifically, the design of diagnosis support systems based on biomedical signals and images is investigated following both the traditional methodologies and those based on Deep Learning strategies.

Concerning the systems for supporting rehabilitation, instead, Virtual and Augmented Reality, as well as, low-cost and wearable devices are examined, with the aim to develop

smart systems for remotely monitoring and assessing the disease course and rehabilitation efficacy and progresses.

1.5 Part Outline

After providing an introduction reporting the reference context in this Chapter, the following Chapter 2 describes the state-of-the-art of the tools and methods for CAD systems for supporting clinical diagnosis, and methodologies for assessing the clinical evaluation of cognitive status and physical rehabilitation. Chapter 3 reports the original contribution beyond the existing literature for the clinical area of radiology and pathology, focusing on Decision Support Systems based on image analysis. Chapter 4, instead, shows the research works conducted in the neurology, psychology and physiatry areas, focusing on the applications for supporting diagnosis and assessment of different pathologies. Finally, the conclusions of the research work described in this thesis are reported in Chapter 5.

Chapter 2

Decision Support Systems

As evidenced in the previous Chapter, precision medicine necessitates the development of clinical Decision Support Systems (DSS) that assist physicians with the decision-making process at the point of care, independently from the clinical domain or clinical area under investigation.

Computer-Aided Diagnosis (CAD) systems are powerful tools for assisting doctors in the interpretation of biomedical signals, supporting them in making clinical decisions. Biomedical signals, including mono and multi-dimensional signals, yield a great deal of information that the medical professional has to analyse, correlate with data from other domains, and evaluate comprehensively and, possibly, in a short time. CAD systems process digital signals in order to make evident some interesting sections or areas, also suggesting possible diseases or clinical outcomes, in order to offer input to support a decision made by the professional clinician.

Several alternative terms have been used for identifying and describing CAD systems over time, mainly based on the CAD specific purpose; these include expert systems (ES), computer-aided evaluation or diagnosis, computerised sound analysis (e.g., Computerised Lung Sound Analysis (CLSA) and Computerised Heart Sound Analysis (CHSA)), computerised biomedical signal analysis [48–52].

Also, there are many ways in which CADs may work. An automatic system may emulate the decision workflow of a human expert for making a diagnosis by means of diagnostic rules. Advanced CAD systems, based on intelligent algorithms, may have learning capabilities, meaning that these systems can analyse clinical data and infer new knowledge [53]. In turn, this new knowledge can enhance current diagnostic rules and enable such systems to improve their performance over time. However, in order to infer new knowledge correctly, these

systems must implement feedback mechanisms with the aim of allowing the human expert to assess and validate the new knowledge.

The increasing amount and complexity of data produced during the modern clinical practice led to investigate intelligent CAD systems making use of Artificial Intelligence (AI), Data Mining algorithms and Machine Learning (ML) approaches for supporting decision making. In addition, knowledge about human anatomy and physiology increased significantly during the last century, as well as the technological improvements of the human body examination tools, including X-rays devices, ultrasounds, or Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), have improved dramatically.

All these improvements led to developing CAD systems offering great support in making and supporting diagnostics decisions for a considerable number of diseases and clinical conditions [54–56]. This aspect is of fundamental importance, also considering that, in recent years, the study of diseases progressed considerably, making, in some cases, more complex and challenging the medical diagnosis.

2.1 A Brief History of Decision Support Systems in Medicine

In the late 1950s, researchers started using computers to investigate problems in several fields, including medicine and biology. Several works of that period deal with medical diagnosis performed by computerised systems [57–59]. In the beginning, automatic diagnostic systems, known as expert systems in medicine, were based on production rules generating the diagnostic results using a knowledge base containing the associations between patients' symptoms and laboratory test outcomes [60, 61].

However, starting from the 1970s, all the limitations in performing accurate diagnosis by means of such "simple" systems emerged. Although those systems introduced some automatisms in clinical diagnosis, the intelligent module was too simple for dealing with complex and delicate problems as a clinical diagnosis. In fact, simple flow charts, statistical pattern matching or association rules did not satisfy at all the requirements from the medical area [62–66]. These analyses revealed that the proposed systems were too far from being automatic and accurate approaches to introduce them in the daily clinical practice.

Thanks to the limitations of the current algorithmic approaches evidenced by Karp [67], the research work conducted in the field of automatic systems for supporting the clinical diagnosis changed perspective. In fact, some researches started to investigate the possibility to employ artificial intelligence and specialised computer algorithms (i.e., pattern recognition and classification algorithms) to support the diagnosis of diseases based on the data acquired

from patients [68]. Today, a CAD system is considered as an important part of a diagnostic process which also actively involves human experts [69].

Besides the previous developments, during the 1960s several radiologists had also started working on building an early form of CAD systems for automatically detecting abnormalities in medical images [70–72]. Today diagnostic radiology and medical image analysis are some of the most active research and application areas in CAD development [69].

In the meantime, medical malpractice litigation has significantly increased since the mid-1980s [73]. Rising medical malpractice liability insurance costs had a negative effect on the healthcare industry, and thus the cost of healthcare has increased dramatically for ordinary users [74, 75]. One of the motivations to develop CAD included helping physicians to avoid medical malpractice cases and thus to reduce medical healthcare costs.

Today, more and more real-world applications employed in the current clinical practice are CAD systems supporting clinicians in everyday diagnostic procedures, offering a cheap and suitable alternative to double reading as a mean for reducing diagnostic errors [76]. Double reading is a procedure in which two readers read and interpret an imaging examination, that, in general, reduces errors and increases sensitivity [77]. The shortage of clinical staff and overwork may reduce physician's availability to offer a "second opinion" about risky (or doubtful) cases. Having automatic systems doing this may be crucial for improving the overall diagnostic performance for the healthcare system in general.

2.2 Modern Systems for Computer-Aided Diagnosis

Based on the previous premises, CAD systems have become very popular in the academic literature of the last decades. In fact, the most common search engines about academic research show that this topic reached more than one thousand papers indexed by from 2008 to 2019. As an example, Fig. 2.1 shows the number of papers per year dealing with Computer-Assisted Diagnosis systems applied to Medical Imaging only.

As evidenced in the previous section, from the late 1970s, CAD systems have been going on supporting the clinical diagnosis in the different clinical areas. Research activities allowed to develop intelligent systems based on signal processing algorithms, procedures for extracting features from input data, and machine learning algorithms for developing intelligent classifiers for the different clinical objectives, from detection to segmentation. Designing performing CAD systems requires the contribution of several professionals, from domain experts to system developers.

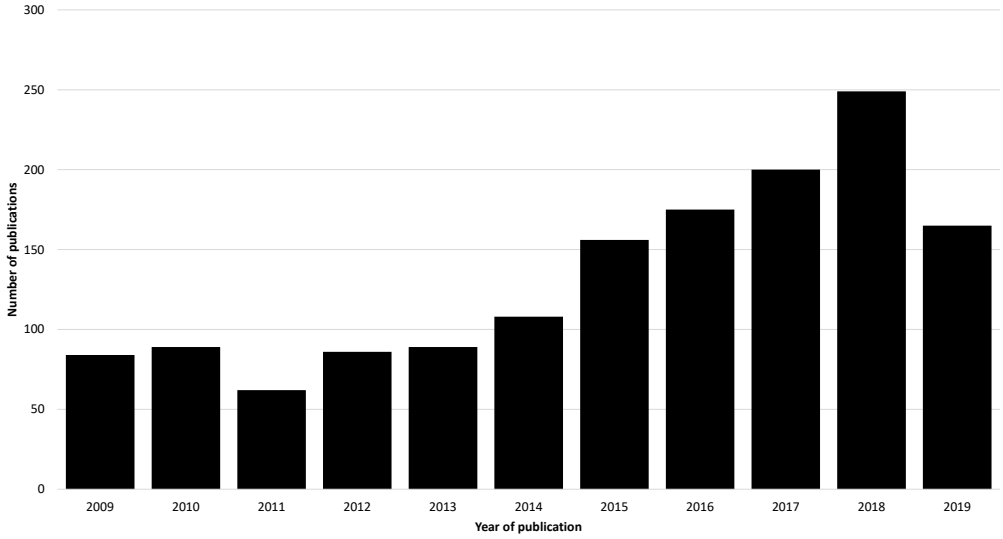


Fig. 2.1 Number of publications per year from 2009 to 2019. Topic: Computer Aided Diagnosis & Medical Imaging. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.

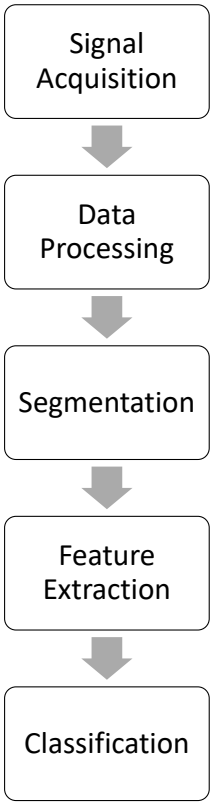


Fig. 2.2 The steps of the pipeline usually implemented by traditional CAD systems.

A traditional CAD system requires designing, developing and integrating at least the modules reported in Fig. 2.2, which include: signal acquisition, data processing, feature extraction and classification. Furthermore, some clinical problems require the CAD system to perform the segmentation of specific areas of interest of the input signals, thus requiring specific modules dedicated to this task. The following section investigates all the features characterising each module of the described pipeline.

From the end-user point of view, CAD systems could be considered as "black-boxes", taking data as input and providing the desired output; however, these are very complex systems composed of sub-modules dedicated to specific tasks. Based on the specific clinical domain interested by the decision support system, each module specialises its specific working methodology and procedures. For example, the signal acquisition module may work on one- or multi-dimensional data or the classifier may be linear or not linear, based on the features pattern complexity and dimensionality. Although all the steps are crucial for accurate detection and classification, or segmentation, of the pathology and areas under investigation, the extraction of sets of significant features, able to describe the input for classifying the

phenomenon under investigation, strongly influences the performances of the CAD system. In fact, the way of extracting features from input data has been changing significantly during the last years, thanks to the widespread of Deep Learning methodologies.

Finally, in order to assess the performance of the decision support systems, the design of CAD systems requires validation procedures aimed at verifying the correct functioning of the automatic system during the normal working conditions [69, 78, 79].

2.3 The Traditional Pipeline of Modern CAD Systems

To better understand the way of working of CAD systems, it is necessary to organise them in order to accent the peculiarities for each specific domain. As already discussed in the previous paragraph, modern decision support systems in medicine are complex architectures based on machine learning approaches for making decisions. From the perspective of the clinical objective, CAD systems may be divided between those aimed at classifying, i.e. detection or diagnosis suggestion, and system for segmentation purposes, i.e. quantification or objectification of measurements.

From the methodological point of view, instead, the training of the intelligent system may be afforded in supervised or unsupervised methods. Before proceeding to detail the characteristics of each category, it is necessary to describe the pipeline for traditional CAD systems aimed at classifying or performing the segmentation fo interesting areas. As already reported in Fig. 2.2, the workflow of traditional CAD systems includes the following steps:

- Data Processing
- Segmentation
- Extraction of Features
- Classification

Data Processing After deciding the clinical domain, developing CAD systems requires a campaign for collecting and organising data, based on which decisions can be taken. Since data may be heterogeneous, e.g. come from different sources, different modalities of acquisition or different clinical facilities with different protocols, data need to be pre-processed in order to have an homogeneous knowledge base in the design phase. Also, processing data is fundamental in order to obtain clean data to be used in the subsequent phases [80].

Automatic systems tend to be misled by differences that humans can readily and easily ignore. If a CAD system is trained with and tested only on data from one kind of source, as is commonly the case for studies reported in the scientific literature, pre-processing may not be necessary, but the results may not be generalised and may not be consistent with different settings.

Based on the defects affecting data, there are a lot of strategies and algorithms to apply in order to clean and uniform them [81, 82]. For example, images may need to be calibrated, data may have to be resampled, or noise removal algorithms, or other filtering procedures, may have to be applied. The goal of applying pre-processing procedures to the input data is to remove differences between data from different sources or obtained with different protocols, in order to allow the training algorithms to work on comparable data for making decisions.

For example, in case of CAD systems working on images, a pre-processing phase is needed after the acquisition of the input data; this step is fundamental for improving the quality of the images and removing possible artefacts [83]. This is a crucial step in order to reach an optimal result in subsequent phases, since the output of this phase affects the performance of the whole workflow. As far as medical imaging is concerned, in literature there is a huge number of useful algorithms for pre-processing images [84–88].

Segmentation In the case of images, a further step allowing to segment the input data may be required. In fact, the Regions of Interest (ROI) to be described by means of features could include only a limited section of the image; thus, the ROIs need to be segmented.

Segmentation is the most important step in processing medical imaging, aiming at separating images into regions that are meaningful for a specific task, such as the detection of organs or the computation of some metrics. Further details will be given in Section 2.4.3.

The detected ROIs could be then processed in the subsequent step. More in detail, a further step for the extraction of features could be necessary for implementing intelligent systems supporting the clinical decision.

Feature Extraction After processing the input data, some candidate regions of the input signals are selected for being subsequently classified. In order to use automatic systems, the selected regions need to be described by extracting some features. Almost all systems use the vector space paradigm for describing Regions of Interest (ROI). This means that for each candidate region, a fixed number of characteristics are computed. Traditionally, the set of features generated for each ROI depended on hand-crafted procedures for evaluating some characteristics of the input signal; in very recent years, this approach has been changing since

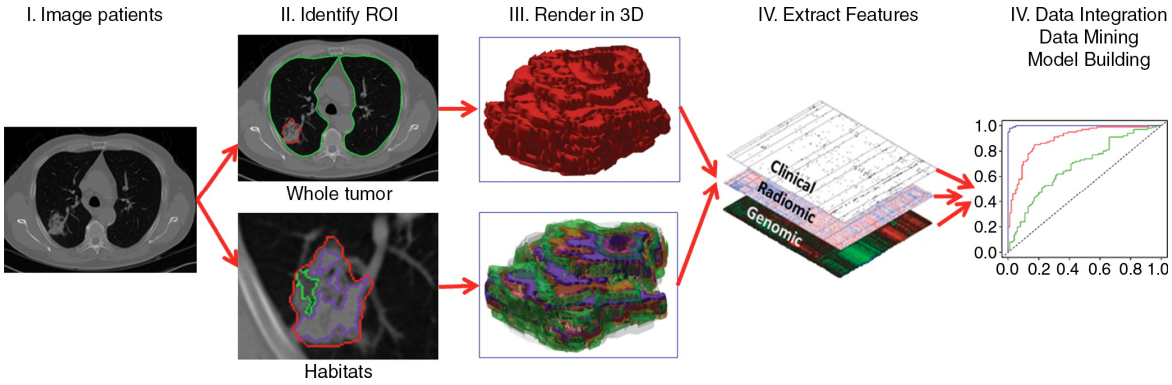


Fig. 2.3 Flowchart shows the process of radiomics and the use of radiomics in decision support. Image from [96]

the set of features is computed automatically by Deep Learning (DL) architectures [89, 90], as further described in Section 2.5. Traditional features may include be the mean value over the region, the standard deviation of the values, the border gradient (in case of images), or other more complex mathematical descriptors of the candidate region and its surroundings.

According to literature, there are several sets of features that could be used to characterize regions of interest starting from input images. From a general point of view, the features may be distinguished between *global* and *local*, based on the localisation of the information used to compute them: global features are computed from information extracted from the whole image, whereas local features are a function of a local area of the image. Among global features, the most used characteristics are those from Haralick [91] (reported in Table 2.1), Local Binary Patterns (LBP) [92] and Threshold Adjacency Statistics (TAS) [93]. Regarding local features, the most famous set of local features are the Speeded-Up Robust Features (SURF) [94].

Describing neoplasias or lesions based on shape descriptors, computed from Regions of Interest, revealed to be fundamental in the Precision Medicine context [95, 96]. Extracting discriminative features, in conjunction with the information from clinical domains other than radiology, can be correlated with clinical outcomes data and used for evidence-based clinical decision support (Fig. 2.3). Fig. 2.4 shows some examples of tumour classification between benign and malignant, based on the shape of the lesion. Border sharpness description and quantification is fundamental for making such a classification.

Regardless of the nature of the features extracted, the number of such descriptors may be too high, or generally may exceed the number effectively required for the subsequent step of decision making. This high dimensionality of the created dataset could lead to relevant problems in the subsequent classification phase, seriously affecting the performance both

Table 2.1 The 14 statistics that can be calculated from the co-occurrence matrix with the intent of describing the texture of the image.

Name	*Description*
Angular Second Moment	$\sum_i \sum_j p(i,j)^2$
Contrast	$\sum_{n=0}^{N_g-1} n^2 \{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)\}, \lvert i-j\rvert = n$
Correlation	$\frac{\sum_i \sum_j (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$ where μ_x, μ_y, σ_x and σ_y are the means and std. deviations of p_x and p_y the partial probability density functions
Sum of Squares: Variance	$\sum_i \sum_j (i-\mu)^2 p(i,j)$
Inverse Difference Moment	$\sum_i \sum_j \frac{1}{1+(i-j)p(i,j)}$
Sum Average	$\sum_{i=2}^{2N_g} i p_{x+y}(i)$ where x and y are the coordinates (row and column) of an entry in the co-occurence matrix, and $p_{x+y}(i)$ is the probability of co-occurence matrix coordinates summing to $x+y$
Sum Variance	$\sum_{i=2}^{2N_g} (i-f_8)^2 p_{x+y}(i)$
Sum Entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i) log\{p_{x+y}(i)\} = f_8$
Entropy	$-\sum_i \sum_j p(i,j) log(p(i,j))$
Difference Variance	$\sum_{i=0}^{N_g-1} i^2 p_{x-y}(i)$
Difference Entropy	$-\sum_{i=0}^{N_g-1} p_{x-y}(i) log\{p_{x-y}(i)\}$
Info. Measure of Correlation 1	$\frac{HXY-HXY1}{max\{HX,HY\}}$
Info. Measure of Correlation 2	$(1-\exp -2(HXY2-HXY))^{\frac{1}{2}}$ where $HXY = -\sum_i \sum_j p(i,j) log(p(i,j))$, HX, HY are the entropies of p_x and p_y, $HXY1 = -\sum_i \sum_j p(i,j) log\{p_x(i)p_y(j)\}$, $HXY2 = \sum_i \sum_j p_x(i)p_y(j) log\{p_x(i)p_y(j)\}$
Max. Correlation Coeff.	Square root of the second largest eigenvalue of Q where $Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$

Fig. 2.4 Tumour diagnosis from contours of breast masses: (b) benign masses, (m) malignant tumours.

during the design phase (for both time and computational resources) and in the classification phase. For these reasons, before proceeding to the design phase of the most appropriate classifier for a given classification task, the obtained dataset is further processed to identify a new feature space. Following this approach, the new dataset is reduced, or generally transformed, with respect to the previous one, while maintaining the same informative content of the extracted data. In this regard, several techniques allow to properly process a dataset in order to reduce its dimensionality, leading to a general increase in the performance of the subsequent classifier; these include feature selection methods, e.g. Information Gain [97], or feature reduction methods, e.g. Principal Component Analysis or Independent Component Analysis [98–106].

Designing good sets of features is a crucial step in developing CAD systems; the design of the classification phase and, thus, the performance reachable by the classification methods, depends on the information described and extracted by the designed set of features.

Classification After describing the areas of interest by means of features, the aim of the classification step is making a decision based on the obtained specific values, e.g. discriminate between benign and malignant tumour or determine the grade of a tumour. In pattern recognition, the problem of identifying regions in a feature space where different kinds of candidates are to be discriminated is solved by training a classifier. Generally speaking, training a classifier for making a decision requires a training set, a validation set and a test set. The first one, used for training the classifier, consists of correctly classified candidates, the so-called Ground Truth. The validation set, instead, is used for validating the classifier during the training phase. A test set is composed of samples used to evaluate the generalisation performance of the classifier, typically constituted by samples not used during the training. The training set is usually created by asking a human expert to provide a reference standard with correct classifications, for example by indicating lesions on chest radiographs with prior knowledge of their location on computed tomography (CT) scans. More details about classifiers supporting decision used in CAD systems are reported in the following Section.

2.4 Machine Learning for CAD Systems

Based on the workflow reported in the previous section (Fig. 2.2), the classification is the last step of the operating workflow for CAD systems. After taking the input data, CAD systems make decisions about a pathology, compute a score for the quantitative classification of tumours, or provide other forms for staging the degree of progression of a disease.

Based on the model designed for making the decision, different kinds of CAD systems can be differentiated based on the training techniques employed. In particular, there are CAD systems based on supervised learning, and other based on unsupervised learning. In the following, some details about the two families of decision support systems will be given.

2.4.1 CAD Systems based on Supervised Learning

Decision Support Systems based on supervised learning are automatic architectures producing a decision by performing classification or regression tasks starting from a knowledge base of annotated samples, meaning that it is known the label (or class) of each sample. Specifically,

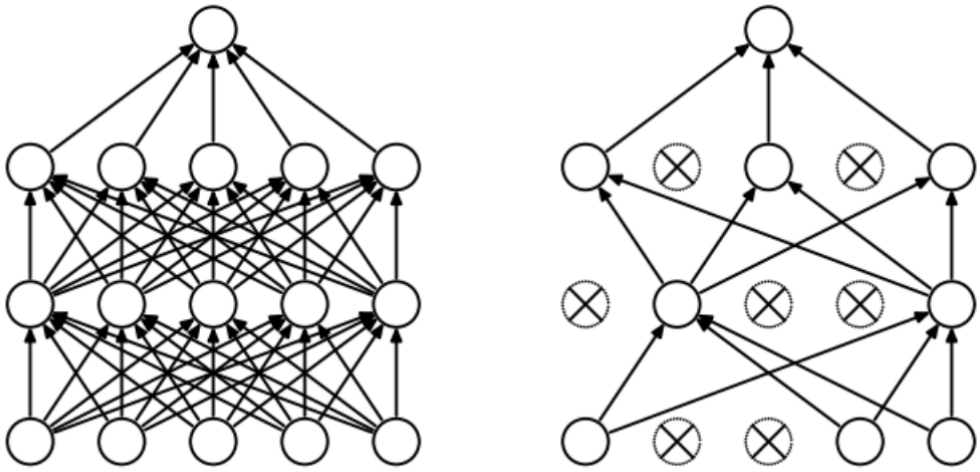
- a classification problem implies the output variable is discrete, or categorical, such as "presence" and "absence" of disease, or "benign" and "malignant" tumour;
- a regression problem implies the output variable is a real value, such as "amount of dollars" or "people weight".

In the last years, a relevant number of studies in the clinical area have been proposed, based on supervised learning approaches, performing classification tasks for supporting the decision making process. In most cases, the architecture designed for performing classification employed Artificial Neural Networks (ANN) or Support Vector Machines (SVM), as well as Swarm Intelligence or Linear Discriminant Analysis (LDA) classifiers built on the radiologists' gold standard labelling [107–123].

Since Artificial Neural Networks have been used in several works described in the following Chapter, detailing the research contribution of this thesis, we focus on this robust machine learning algorithm.

In literature, there are several ways for classifying ANNs, i.e. based on:

- the function that the ANN is designed to serve, e.g. pattern association or clustering (in fact, some implementations are employed also for unsupervised learning approaches; more details in Section 2.4.2);
- the degree of connectivity of the neurons in the network architecture; it may be fully-connected, when all the neurons of a layer are connected with all the neurons of the subsequent layer; partially-connected, where a dropout procedure, for preventing over-fitting during the training, delete some connections between consecutive layers [124] (Fig. 2.5);



(a) Standard Neural Net (b) After applying dropout.

Fig. 2.5 ANN models; Left: A traditional neural network with 2 hidden layers. Right: the same neural network after dropout, where some connections are missing.

- the direction of flow of information within the network, identifying recurrent and feed-forward neural networks [125];
- the type of learning algorithm, which represents a set of systematic equations that use the outputs obtained from the network, along with an arbitrary performance measure, to update the internal structure of the ANN, in terms of connection weights;
- the learning rule, also known as the driving engine of the learning algorithm;
- the degree of learning supervision needed for ANN training.

In general, supervised learning involves the training of an ANN with the correct answer (i.e, target outputs) being given for every sample, and using the deviation error of the ANN solution from the corresponding target values to determine the required amount by which each weight should be adjusted.

The development of an ANN requires the partitioning of the available data (dataset) into:

- **training set**: it should include all the available data belonging to the problem domain and is used in the training phase to update the weights of the network;
- **validation set**: used after selecting the best network to further examine the network or confirm its accuracy before the network is implemented in the neural system and/or delivered to the end user;

- **test set**: is used at the end of the learning process to check the degree of generalisation of the network on data that have not been used during the training phase.

The data used in each set (training, validation and test) should be different. There are no mathematical rules for the determination of the required sizes of the training, validation and test sets. However, there are some rules of thumb derived from the designer experience and the analogy between ANNs and statistical regression [126].

For correctly training an Artificial Neural Network, Cross-Validation (CV) is a popular strategy for algorithm selection. The main idea is to resample the input dataset, once or several times, thus training the neural architecture considering each subset of training and validation sets. The popularity of CV mostly comes from the "universality" of the data splitting heuristics. Nevertheless, some CV procedures have been proved to fail for some model selection problems, depending on the goal of model selection, estimation or identification. Furthermore, many theoretical questions about CV remain widely open, as reported in [127–133].

Concerning with the topology of Artificial Neural Networks, in literature, two different classes of ANNs could be identified on the basis of the number of hidden layers: Shallow and Deep Neural Networks. In details, ANNs with few hidden layers are named Shallow Neural Networks, whereas a Deep architecture has grater number of hidden layers, typically three or more (Fig. 2.6).

Recent studies have evaluated the performance differences between shallow and deeper neural networks by highlighting the strengths and weaknesses of both architectures [134, 135]. According to the literature, there is not an real motivation to prefer Deep or Shallow Neural Networks. In fact, both the architectures could approximate any (reasonable) function, where the quality of the final generalisation properties strictly depends on the significance and classes-balance of the available training data. However, Shallow Neural Networks could reach an extremely high number of neurons in their hidden layers, leading to very wide architectures, thus making the number of parameters to be tuned during the training phase considerably high, with the risk of data overfitting [136]. On the other side, the introduction of multiple layers makes ANNs able to learn features at different levels of abstraction, leading to stronger capabilities of generalisation if compared to a shallow architecture with the same computing power, in terms of the number of neurons and connections. But, also deeper architectures requires the introduction of different strategies for preventing overfitting, including dropout or other regularisation methodologies [137].

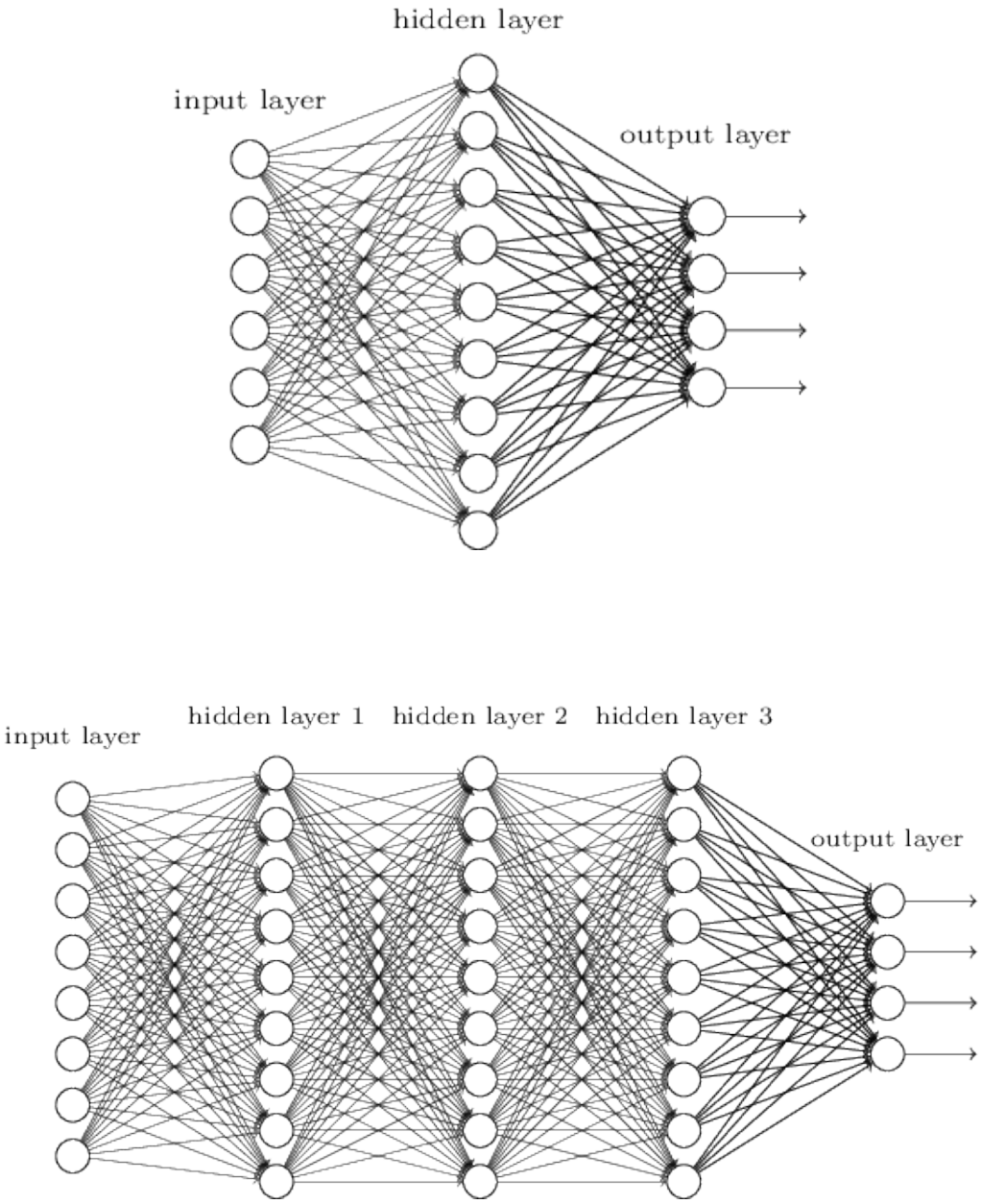


Fig. 2.6 Architectural differences between (a) shallow and (b) deep neural networks.

.

2.4.2 CAD Systems based on Unsupervised Learning

Differently from supervised learning, the unsupervised learning does not require the labelling of the training samples, i.e. does not require a knowledge base with annotated data. Instead, the network arranges examples into clusters based on their similarity or dissimilarity [126], by exploring the structure underlying the data and the correlation between the examples [138–140].

Typically, medical applications performing visual recognition and classification tasks use labelled data, as required by all algorithms belonging to the approaches based on supervised learning. However, for tasks where manually generating labels corresponding to large datasets is laborious and expensive, the use of unsupervised learning methods can be of significant value.

The use of unsupervised learning approaches requires the knowledge of some complementary information provided with the data to improve learning, which, differently from other domains, may not be available for several classification tasks in medical imaging [141–143].

Although the approach based on supervised learning seems to be the ideal one in problems of medical interest, several works based on unsupervised learning deal with medical detection and diagnosis cases, even though these are less frequently used for CAD systems than supervised algorithms [144].

Some of the most common algorithms used in unsupervised techniques include Clustering, e.g. k-means or or hierarchical clustering, Artificial Neural Networks, e.g. autoencoders, Generative Adversarial Networks (GAN) or Self-Organizing Maps (SOM), and methods for learning latent variable model, e.g. Principal Component Analysis (PCA) or Independent Component Analysis (ICA).

The widespread of Deep Learning methodologies in the last years, as described in Section 2.5, allowed to return to make use of unsupervised methodologies to solve medical problems, especially autoencoders architectures and GANs. In fact, autoencoders, by design, reduce data dimensions by learning how to ignore the noise in the data. This make autoencoders particularly suitable to solve tasks linked to medical imaging. The typical architecture of an autoencoder is represented in Figure 2.7.

Several works dealing with unsupervised architecture have been proposed in the literature so far [145]. The approach by Shin *et al.* used stacked autoencoders for multiple organ detection in MRI scans [146]. Vaidhya *et al.* presented a brain tumor segmentation method with a stacked denoising autoencoder evaluated on multi-sequence MRI images [147]. In a work by Sivakumar *et al.*, the segmentation of lung nodules is performed with unsupervised clustering methods [148]. In another study, Kumar *et al.* used features from an autoencoder

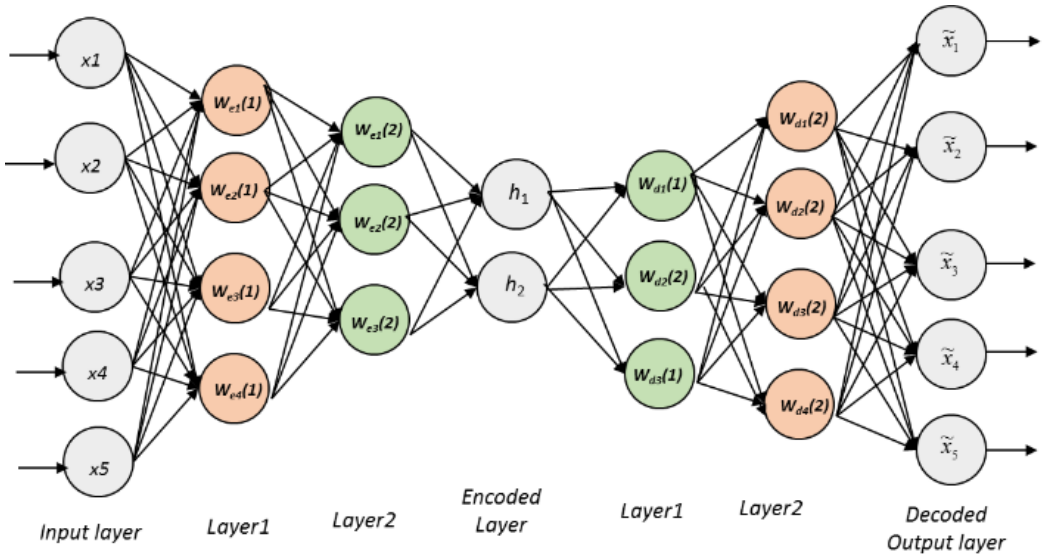


Fig. 2.7 Representation of an autoencoder.

for lung nodule classification [149]. These auto-encoder approaches, however, did not yield satisfactory classification results. Other than these, unsupervised deep learning has also been explored for mammographic risk prediction and breast density segmentation [150]. Unsupervised feature learning remains an active research area for the medical imaging community, more recently tackled with Generative Adversarial Networks (GAN) [151].

2.4.3 CAD Systems for the Segmentation of Images

Alongside classification purposes, segmentation is one of the main objectives for which CAD systems processing biomedical images can be employed in the daily clinical practice. Although segmentation and classification systems may be considered different, in terms of algorithms or purposes, they are strictly linked together. In fact, the segmentation process always includes a classification procedure, meaning that the segmentation pipeline, in the end, allows assigning a label to each pixel of the input image.

As well as the classification tasks may be performed following different approaches (e.g., supervised or unsupervised), there are plenty of algorithms allowing to perform the segmentation of images, from the simplest ones to more complex pipelines. There are different categories of segmentation approaches; this classification is mainly based on the features involved in discriminating the different parts of the images, in terms of type and processing methods. It has to be noticed that features of interest may include pixel intensities, gradient magnitudes, or measures of texture. Focusing on the segmentation techniques

dealing with these features, they can be broadly classified into three categories: region-based, edge-based, or classification-based [87]. Specifically, region-based and edge-based segmentation techniques explore intra-region similarities and inter-region differences among features in order to assign a label to each pixel. Classification techniques, instead, label individual pixels or voxels based on the values of the features, considering 2D or 3D space, respectively.

In some cases, depending on the domain of the problem under investigation, grey-level thresholding could be a simple but effective segmentation method [152]. Concerning thresholding, it may be performed at a global or local level, i.e., thresholds can be selected equal to a constant value throughout the image, or can vary spatially by computing different values for each subsection of the image. Thresholding methods can also be categorized as point-based or region-based techniques. Region-based methods compute the value of a proper threshold based on the grey-level of an individual pixel and the properties of its neighbourhood. Whether local or global, point-based or region-based, thresholds are typically estimated from the intensities histogram using different approaches.

In other cases, a priori knowledge could be necessary to perform a good segmentation. Noise, artefacts or other kinds of interferences could make the segmentation a tricky task, not simply achievable using only information coming from the grey levels of the input image. To overcome such kinds of problems, deformable and active models or atlas-based methods have been employed over time, depending on the task to address [153–156].

Finally, more complex approaches, including Region Growing, Bayesian approaches, Clustering or Deformable models are segmentation techniques used for performing the segmentation of medical images [157–159].

There are segmentation approaches based on classification algorithms, too. The classification approach for segmentation, based on supervised learning, requires training data from users (i.e., the labelled ground truth) enabling classifiers to learn how to label each pixel of the input images. On the other side, unsupervised classifiers, starting from unlabelled data, perform cluster analysis to discriminate natural structures in the input images. In recent years, however, segmentation methods based on Deep Learning strategies have been introduced [157, 160–162]. Section 2.5 deal with a detailed discussion about Deep Learning architectures used also for segmentation purposes.

Regardless of the segmentation method, the output from the segmentation task is, in the most cases, a binary (or multi-labelled) mask allowing to filter out all the undesired regions of the input image [163], in order to perform the subsequent actions, such as 3D reconstruction of areas of interest or organs. An example of a CAD system developed for the segmentation

Table 2.2 Confusion Matrix

		True Condition	
		Positive	***Negative***
Predicted Condition	***Positive***	*TP*	*FP*
	Negative	*FN*	*TN*

of MR images from patients affected by Autosomal Dominant Polycystic Kidney Disease (ADPKD) [164] is reported in Fig. 2.8. Specifically, starting from input MR slices, kidneys are automatically segmented. The implemented CAD system also allows superimposing the ground truth segmentation to those obtained with the automatic approach in order to compare and evaluate the segmentation error.

2.4.4 Evaluating the Performance of a CAD System

In order to evaluate the performance of CAD systems, several metrics are considered in the literature, which depend on the aim of each specific CAD. In the following sections, the metrics for evaluating both the classification and the segmentation approaches will be investigated.

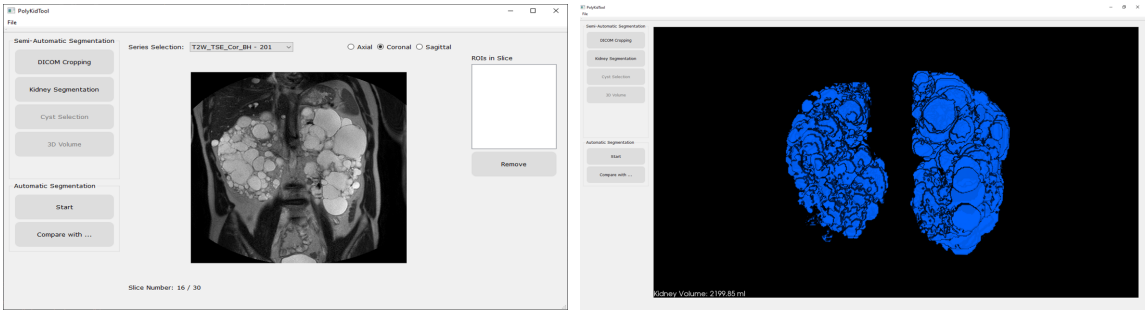
2.4.4.1 Classification Performance

For the sake of the definiteness, let us consider a binary classification task, with Positive (P) and Negative (N) classes. The results obtained from a classification system may be organised in the so-called *Confusion Matrix*, as the one reported in Table 2.2.

Specifically, *True Positive (TP)* indicates the number of instances labelled as Positive, and correctly classified as Positive; *True Negative (TN)* refers to the number of instances labelled as Negative and correctly classified as Negative; *False Positive (FP)* refers the number of instances labelled as N but misclassified as Positive; *False Negative (FN)* indicates the number of instances labelled as P but classified as Negative.

Starting from the confusion matrix, several metrics could be computed in order to evaluate the performance of a CAD system. Equations 2.1, 2.2 and 2.3 report how to compute Accuracy, Specificity and Sensitivity, which are the metrics most used in the subsequent chapters to compare the results obtained with the state of the art.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2.1}$$



(a) (b)

(c)

Fig. 2.8 The representation of a CAD system for the segmentation and 3D reconstruction of Polycystic Kidney Disease from MR Images, (a) input MR images; (b) the output obtained from a fully-automated segmentation procedure; (c) the superimposition of the ground truth to the obtained 3D volume.

$$Specificity = \frac{TN}{N} = \frac{TN}{FP+TN} \tag{2.2}$$

$$Sensitivity\ (or\ Recall) = \frac{TP}{P} = \frac{TP}{TP+FN} \tag{2.3}$$

$$Precision = \frac{TP}{TP+FP} \tag{2.4}$$

Although a correctly designed classifier should maintain TPs and TNs higher than FPs and FNs, real applications may produce prediction errors. A false positive error is a result that indicates a given condition exists, when it does not; a false positive error is a type I error where the test is checking a single condition, and wrongly gives an affirmative (positive) decision. A false negative error, on the contrary, is a test result that indicates that a condition does not hold, while in fact it does. In other words, erroneously, no effect has been inferred; a false negative error is a type II error occurring in a test where a single condition is checked for and the result of the test indicates erroneously that the condition is absent [165].

Based on the final aim of the decision support system, type I or type II errors should be avoided respectively. In fact, test for screening the population, aimed at identify pathologies in people without any symptoms, should avoid type II errors, increasing, as much as possible, the Sensitivity of the system. As a result many of the positive results are false positives. This is acceptable, particularly if the screening test is not harmful or expensive. On the contrary, diagnostic tests, aimed at providing a definitive diagnosis, should avoid both type I and II errors, increasing the specificity of the test, with more weight given to diagnostic precision and accuracy than to the acceptability of the test to patients [166].

The advent of new competitive imaging modalities for the same diagnostic problem has led to performing many studies involving comparisons among the information obtained from the different imaging techniques, especially in radiology [167]. Several of these comparisons have used Receiver Operating Characteristic (ROC) curves [168–171]. The main goal of these studies is to judge the discrimination ability of various statistical methods that combine different clues and test results for predictive purposes. A ROC curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is created by plotting the True Positive Rate (TPR), or Sensitivity, against the False Positive Rate (FPR), obtained as in Equation 2.5, at various threshold settings. The Area Under the ROC Curve (AUC), ranging from 0 to 1, allows at

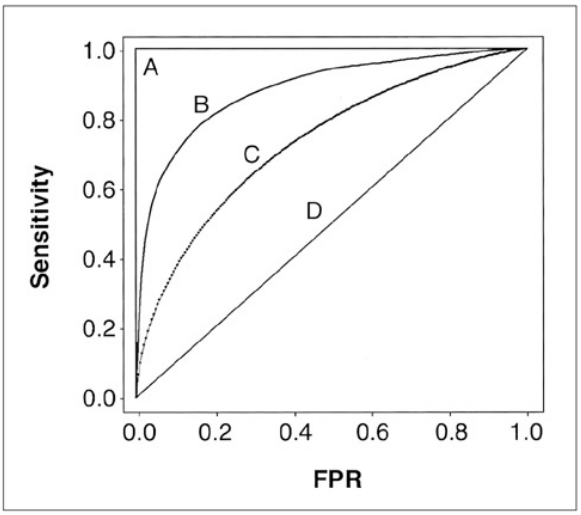


Fig. 2.9 Four ROC curves with different values of the area under the ROC curve. A perfect test (A) has an area under the ROC curve of 1. The chance diagonal (D, the line segment from 0, 0 to 1, 1) has an area under the ROC curve of 0.5. ROC curves of tests with some ability to distinguish between those subjects with and those without a disease (B, C) lie between these two extremes. Test B with the higher area under the ROC curve has a better overall diagnostic performance than test C. Image from [167]

discriminating the test performing better. A perfect test has AUC equal to 1, while the chance has an AUC equal to 0.5. Some examples are reported in Fig. 2.9.

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{N} = \frac{FP}{FP+TN} = 1 - Specificity \tag{2.5}$$

2.4.4.2 Segmentation Performance

Regarding the segmentation tasks, the Confusion Matrix should also be considered, but the metrics reported in Equations 2.6, 2.7 and 2.8 have to be used for evaluating the segmentation performance. In particular, the *Boundary F1 Score (BF)* measures how closely the predicted boundary of an object matches the corresponding ground truth; it is defined as the harmonic mean of the Precision (Eq. 2.4) and Recall (Eq. 2.3) values. The resulting score spreads in the range [0, 1], from a bad to a good match. The *Jaccard Similarity Coefficient*, instead, is the ratio between the number of pixels belonging to the Positive class classified correctly (TP) and the sum of the number of pixels belonging to the Positive class ($P = TP + FN$) and the Negative pixels wrongly predicted as Positive (FP). Since, in some research work, the

object detection problem has also been investigated, the *Average Precision* (Eq. 2.4) and the logAverageMissRate were evaluated, considering the *Miss Rate* (MR) according to Eq. 2.8.

$$Boundary\ F1\ Score\ =\ \frac{2*Precision*Recall}{Precision+Recall} \tag{2.6}$$

$$Jaccard\ Similarity\ Coefficient\ =\ \frac{TP}{TP+FP+FN} \tag{2.7}$$

$$Miss\ Rate\ =\ \frac{FN}{FN+TP} \tag{2.8}$$

2.5 Deep Learning: A New Perspective for CAD Systems

The development of hardware technology, such as general-purpose computing GPUs, enabled performing complex operations in shorter computation time. For example, training a Deep Neural Network (DNN), or more complex architectures, for classification or segmentation purposes, became a more manageable task than in the past. Deep learning models now can generate more meaningful and powerful features after analysing a large amount of uncategorised data and training the model for accurate prediction by using these features. This process, implemented by some deep architectures, is surprisingly similar to the self-organization of humans for obtaining or discovering knowledge. These breakthroughs have led to relevant improvements in performances in various research fields, such as speech recognition, image classification, and face recognition, also thanks to the scientific community which made available several open-source deep learning libraries such as Caffe, Microsoft Cognitive Toolkit (CNTK), Tensorflow, Theano, and Torch [172–176].

Different learning strategies and architectures have been introduced in the literature so far. Nowadays, these automatic learning systems have tremendous success, mainly thanks to the use of (deep) Convolutional Neural Networks (CNNs) and Deep Learning algorithms in the field of image processing, classification and also segmentation. In fact, these kinds of architectures are able to take a decision (i.e. classify, detecting objects or segment) working directly on the raw images given as input to the network [177–179].

Unlike traditional neural architectures, a Convolutional Neural Network may have different kinds of layers, which are generally combined in several ways depending on the specific implementation and the task for which have been designed [172]. Specifically, the three main types of layers to build these architectures are Convolutional Layer, Pooling Layer, and Fully-Connected Layer. Fig. 2.10 reports an example of the implementation of a CNN

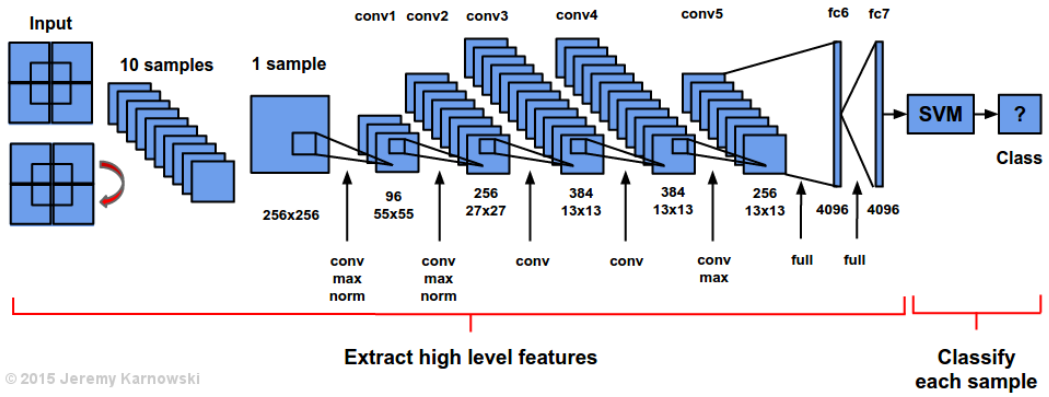


Fig. 2.10 The implementation of Alexnet CNN combined with SVM classifier. This architecture is composed of 5 convolutional layer, some of them coupled with max-pooling and normalization layers, and 2 fully-connected layers. Finally, the SVM classifiers is used for class discrimination.

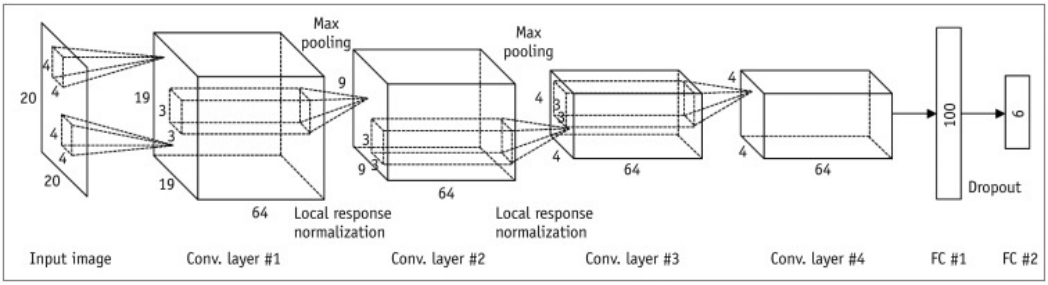


Fig. 2.11 Architecture of convolutional neural networks, including input, Convolutional Layers, and Fully Connected layers.

.

architecture, showing the well-known AlexNet CNN architecture combined with a final SVM classifier [180] as classification layer. The representation of another CNN architecture is reported in Fig. 2.11, which clearly shows the composition of the network about convolutional, pooling layers and fully connected layers. These kinds of DL architectures are specifically designed for image classification tasks.

More in details, the primary purpose of a convolutional layer is to detect distinctive local motif-like edges, lines, and other visual elements. The parameters of specialized filter operators, termed as convolutions, are learned. This mathematical operation describes the multiplication of local neighbours of a given pixel by a small array of learned parameters called a kernel (Fig. 2.12A). By learning meaningful kernels, this operation mimics the extraction of visual features, such as edges and colours, similar to those noted for the visual cortex.

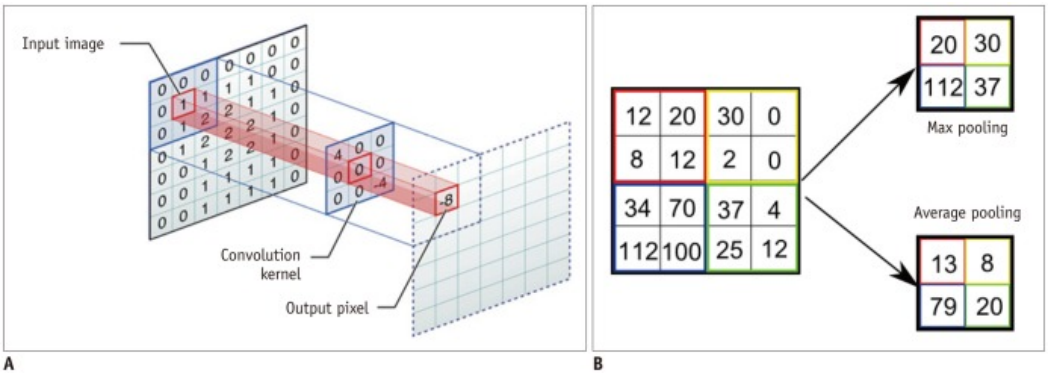


Fig. 2.12 Illustration of convolution and pooling methods.

The network performs this process by using filter banks. Each filter is a square-shaped object that travels over the given image. The image values on this moving grid are summed using the weights of the filter. The convolutional layer applies multiple filters and generates multiple feature maps. Convolutions are crucial components of CNNs and are vital for succeeding in image processing tasks such as segmentation and classification. In order to capture an increasingly large field of view, feature maps are progressively and spatially reduced by pooling the pixels together (Fig. 2.12B). By propagating only the maximum or average activation through a layer of max or average pooling, convolutional layers subsequently become less sensitive to small shifts or distortions of the target object in the extracted feature maps. The pooling layer is used to effectively reduce the dimensions of feature maps, and is robust to shape and position changes of the detected semantic features within the image. Choosing max-pooling, average-polling, or other forms of downsampling the feature map depends on the task performed by the CNN, e.g. natural language processing or computer vision tasks. As reported in Fig. 2.10 and Fig. 2.11, the global architecture of the classifier shows convolutional and pooling layers repeated several times before the fully connected layers. This architecture allows computing feature spaces at different levels, from lower ones (e.g. edges and gradients) to higher abstraction levels. The Fully Connected (FC) layers are incorporated to integrate all the feature responses from the entire input image and provide the final results [178].

A CNN, thus, has the capability to automatically extract some descriptors, or features, from an image, thanks to the so-called "feature learning capability", eliminating the need to design and develop image processing algorithms aiming to the extraction of *hand-crafted* features, necessary to a traditional feature-based classifier, such as an ANN or SVM, to perform the classification task [172, 177, 178].

By using deep CNN architectures to mimic the natural neuromorphic multi-layer network, deep learning can automatically and adaptively learn a hierarchical representation of patterns, from low- to high-level features, and subsequently identify the most significant features for a given task (Fig. 2.11) [178].

Because deep CNN architecture generally involves many layers in the neural network, there may be millions of weight parameters to estimate, thus requiring many data samples for model training and parameter tuning. In general, the minimum requirement of data size depends on the specific application. For example, more than 1000 cases per class could be needed to train deep learning architecture from scratch in classification tasks. However, there are alternative methods to get around the data size criteria. One is data augmentation, and the other is the reuse of a pre-trained network. By using these methods, a smaller number of cases per class could provide a reasonable outcome.

According to literature, Convolutional Neural Networks are powerful architectures that may be used in three different ways [172, 178, 181]:

- **Training from scratch**: as for ANNs, Convolutional Neural Networks may be created from scratch, designing the overall architecture and providing enough samples as input for training. Generally, this process takes a lot of time and computational resources using large datasets with several classes.

- **Transfer Learning or Fine-Tuning**: this approach consists in using an available pre-trained model for classification purposes different from the original target classes. In details, it is possible to fine-tune the classification layer of a CNN to predict new classes given as input. Thanks to the power and versatility of these architectures, it is possible to fine-tune all the layers of the network, or just some of them, maintaining the weights of the other layers. In particular, some authors suggest fine-tuning the higher levels of networks, namely the layers closer to the output, due to the higher generality of the features respect to those computed in the lower levels [172].

- **Features Extractors**: in addition to the previous ways for implementing a CNN, it is possible to get the output at a specific level of a CNN and use it as automatically-generated descriptors of the input data. Since this process is iterative, it is possible to intercept the output at the desired level, based on the desired level of abstraction of the features.

The choice of how to use the CNN depends on the problem under consideration, but is also strictly related to the available amount of data for training the architecture and the computational resources available.

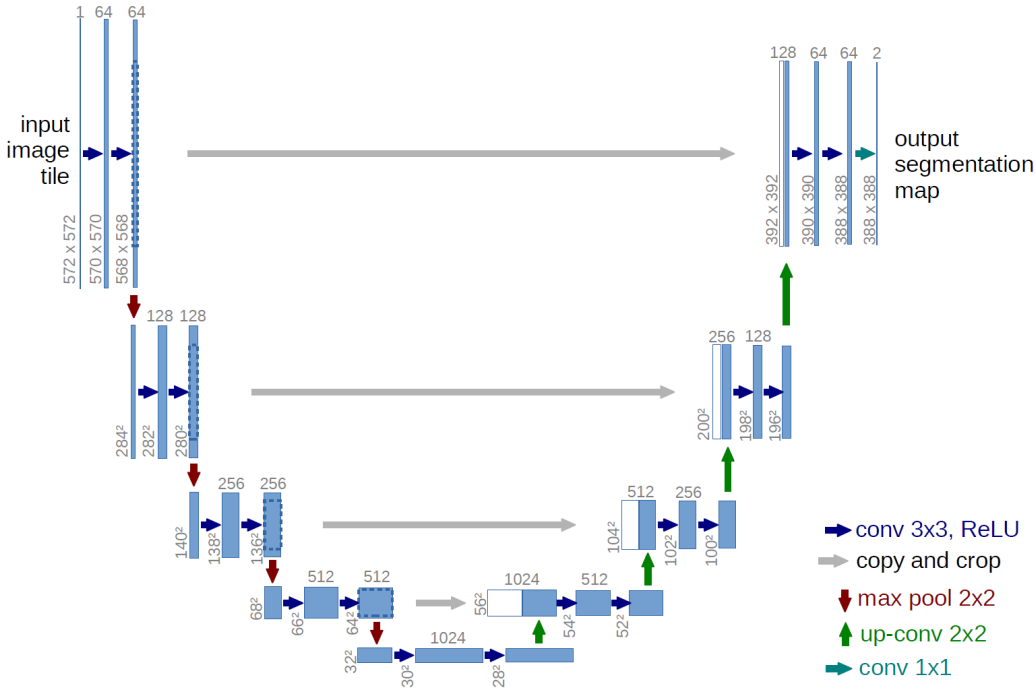


Fig. 2.13 U-net: a CNN architecture performing semantic segmentation.

Convolutional Neural Networks are also employed as intelligent systems for the automatic segmentation of images. Semantic Segmentation could be intended as image classification at a pixel level. A general semantic segmentation architecture can be broadly thought of as an encoder network followed by a decoder network, where the encoder could also be a pre-trained classification network, like VGG/ResNet, followed by a decoder network (Fig. 2.13). The task of the decoder is to semantically project the discriminative features (lower resolution) learnt by the encoder onto the pixel space (higher resolution) to get a dense classification.

Unlike classification, where the end result of the very deep network is the only important thing, semantic segmentation not only requires discrimination at pixel level but also a mechanism to project the discriminative features learnt at different stages of the encoder onto the pixel space. Different approaches employ different mechanisms as a part of the decoding mechanism.

Semantic segmentation demonstrated to be a consistent approach for medical images segmentation, thanks to very powerful architectures performing segmentation also considering three-dimensional data inputs [182–184].

In the following Chapter 3, all the considered methodologies have been employed for different case studies in order to support clinical decisions in several units of the healthcare system.

2.6 Virtual Reality Systems for Clinical Assessment

Virtual Reality (VR) is an artificial, computer-generated simulation or recreation of a real-life environment or situation. It is capable of generating an immersive experience for the users by making them feel like they were experiencing the simulated reality firsthand, primarily by stimulating their vision and hearing. Alongside with Augmented Reality (AR), which includes technologies for adding computer-generated content to the real-world environment, VR constitutes the group of the so-called Mixed Reality (MR) technologies, more general techniques allowing to merge both real and virtual worlds to produce new environments and visualizations.

A Virtual Reality platform is a combination of hardware and software aimed at building VR applications for researchers and developers. In such a system, the hardware components accept input from user-controlled devices and relay multi-sensory feedback to create a virtual world's illusion. Generally speaking, in VR applications, the virtual world may be either a model of a real-world entity, such as a house, or an abstract world that does not exist in a real sense, but is understood by humans, such as a chemical molecule or a data set representation, or it may be part of a completely imaginary reality. Typically, a VR system comprises [185]:

- the graphic rendering system that generates and render the virtual environment at 20–30 frames per second, for avoiding the so-called "motion sickness";
- the database construction and virtual object modelling software for building and maintaining detailed and realistic models of the virtual world. In particular, the system should be able to handle the properties of the models, including geometry, texture, and physical characteristics. Also, intelligent behaviours of the objects should be modelled and included in the virtual world;
- the input tools (trackers, gloves or mice) that continually report the position and movements of the users;
- the output tools (visual, aural and haptic) that immerse the user into the virtual environment.

The correct combination and integration of all the components constituting the generation and presentation systems will allow obtaining more performing systems, suitable for being also used in more serious contexts, other than gaming or ludic applications.

Based on the hardware and software included in a VR system, it is possible to distinguish among:

- **Fully Immersive VR**: with this type of solution the user appears to be fully immersed in the computer-generated environment. This illusion is produced by providing immersive output devices (head-mounted display, force feedback robotic arms, etc.) and a system of head and/or body tracking to guarantee the exact correspondence and coordination of users' movements with the feedback of the environment.

- **CAVE**: this is a room where a computer-generated world is projected on the front and side walls. This solution is particularly suitable for collective VR experiences because it allows different people to share the same experience at the same time.

- **Telepresence**: users can influence and operate in a world that is real but in a different location. The users can observe the current situation with remote cameras and achieve actions via robotic and electronic arms.

- **Augmented**: the user's view of the world is supplemented with virtual objects, usually to provide information about the real environment. For instance, in military applications, vision performance is enhanced by pictograms that anticipate the presence of other entities that are currently out of sight.

- **Desktop VR**: uses subjective immersion on a standard PC screen. The feeling of immersion can be improved through stereoscopic vision. Interaction with the virtual world can be made via mouse, joystick or typical VR peripherals such as a data glove.

In the last years, VR systems were shown to be transversal tools in the clinical field, allowing researchers and clinicians to perform different tasks, from cognitive and physiological assessment to the implementation of protocols allowing the de-hospitalization of rehabilitation treatments [186–188]. Several VR applications for the understanding, assessment and treatment of mental health problems have also been developed in the last years [189–191]. Also, in VR, patients learn to manipulate problematic situations related to the problem afflicting them. For this reason, the most common application of VR in this area is the treatment of anxiety disorders [192].

In this context, Virtual Reality can be described as an advanced form of a human – computer interface allowing the user to interact with and become immersed in a computer - generated environment in a naturalistic fashion. In addition, VR can also be considered as an advanced imaginal system, a medium that is as effective as reality in inducing emotional and cognitive responses. This is achieved through its ability to induce a feeling of "presence" in the computer - generated world experienced by the user [193, 194].

With regard to the psychological aspect of using VR technologies to perform such tasks, it has now been established that VR technology can offer enough capabilities to influence the shape of psychological therapy profoundly. For most clinical applications the key characteristics of virtual environments are the high level of control of the interaction with the device and the enhanced patient experience [195]. Emerging applications of VR in psychotherapy include post-traumatic stress disorder, sexual disorders, pain management, stress management and eating disorders and obesity [196–201].

In fact, immersive VR can be considered an embodied technology for its effects on body perceptions [202]. VR users become aware of their bodies during virtual navigation; for example, their head movements alter what they see, or they could perceive the reproduction of the movement of their limbs. The sensorimotor coordination of the moving head with visual displays produces a much higher level of sensorimotor feedback and first-person perspective (egocentric reference frame). For example, through the use of immersive VR, it is possible to induce a controlled sensory rearrangement that facilitates the update of the biased body perceptions. This allows the differentiation and integration of new information, leading to a new sense of cohesiveness and consistency in how the self represents the body. The results of this approach are very promising; in fact, different experimental researches showed that VR is effective in producing fast changes in body experience [203] and in body dissatisfaction [204].

Another medical field in which VR has been successfully applied is neuropsychological testing and rehabilitation. Here, the advantage of VR over traditional assessment and intervention is provided by three key features: the capacity to deliver interactive 3D stimuli within an immersive environment in a variety of forms and sensory modalities; the possibility of designing safe testing and training environments; and the provision of cueing stimuli or visualisation strategies designed to help guide successful performance to support an error-free learning approach [186, 187, 205, 206].

Future VR clinical applications will also include online virtual worlds (such as Second Life, There or Active Worlds): computer-based simulated environments characterised by the simultaneous presence of multiple users within the same simulated space, who inhabit and

interact via avatars [207]. Online virtual worlds can be considered as 3D social networks, where people can collaboratively create and edit objects, besides meeting each other and interacting with existing objects [208]. Over the last few years, the number of virtual worlds' users has dramatically increased and today Second Life, the largest 3D online digital world, counts about 12 million subscribers.

Beyond clinical applications, VR has been revealed to be a powerful tool for behavioural neuroscience research. Using VR, researchers can carry out experiments in an ecologically valid situation, while still maintaining control over all potential intervening variables. Moreover, VR allows us to measure and monitor a wide variety of responses made by subject [209].

Chapter 3

Case Studies: Computer Aided Diagnosis Systems based on Medical Imaging

After the introduction about the clinical context highlighting the importance of having accurate diagnosis systems in the precision medicine era, and the description of both methodologies and technologies for achieving the set goals, this Chapter reports the research works conducted in the field of interest of this thesis during the Ph.D. activities, focusing on studies concerning the medical imaging.

In details, as well as the importance of a quantitative and objective assessment of different diseases has been investigated for several clinical fields in Chapter 1, the designed and implemented decision support systems and the technological frameworks for objectively assessing different pathological conditions will be analysed in the clinical areas of radiology and pathology. In all cases, we will point out the importance of quantitatively and objectively measuring pathological indexes and how it may be fundamental for the new era of precision medicine.

3.1 Motivations

In the last decades, the amount of deaths due to cancer has significantly increased, overcoming the number of deaths caused by heart attacks and stroke, as emphasized in the reports of the World Health Organization (WHO) [210, 211].

More in detail, in industrialised countries, several kinds of neoplasias exist with a high incidence. An early, and possibly non-invasive, diagnosis and staging could luckily prevent unfavourable prognosis, reducing the risk of clinical complications in some cases, also improving the life expectancy of people affected by such diseases.

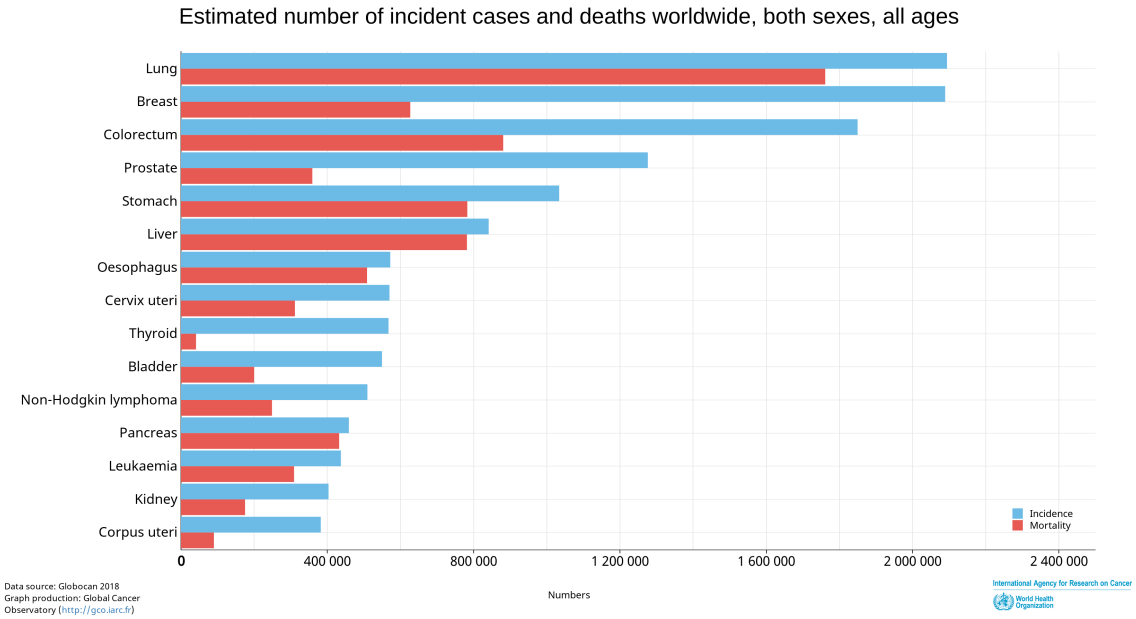


Fig. 3.1 Top fifteen tumours for incidence cases and deaths worldwide in 2020.

In this thesis, four cancer sites have been considered and evaluated among the top fifteen in the world for estimated number of incident cases and deaths [211], namely the breast, liver, blood neoplasias and kidneys. This can be well noticed in Fig. 3.1, where the number of people affected by forms of the liver tumour was estimated at the sixth position all over the world in 2018; the incidence of breast cancer is at the first position all over the world. In addition, leukaemia and kidney tumours show lower incidence with respect to the previous ones, but they still figure among the most diffuse tumoural forms. It can be noticed that the same malignant forms are among the top fifteen in Europe too, making breast cancer the most diffused tumour form, as reported in Fig. 3.2.

Medical imaging is a fundamental methodology for representing the internal organs of the human body; it permits the non-invasive and accurate diagnosis of several diseases, including also neoplasias [212], and to assess the course of the disease over time.

It should be pointed out that there are different imaging techniques able to highlight the characteristics of the human body, on the basis of the sensors used to acquire information and produce the representation of each internal organ or body structure [213]. Moreover, besides the diagnostic capabilities, medical imaging is also crucial for staging and monitoring the clinical course of each disease under investigation [214–216].

The rest of this Chapter is organised as follows: Section 3.2 shows the contribution to the state of the art in the radiological field reporting the designed and implemented innovative

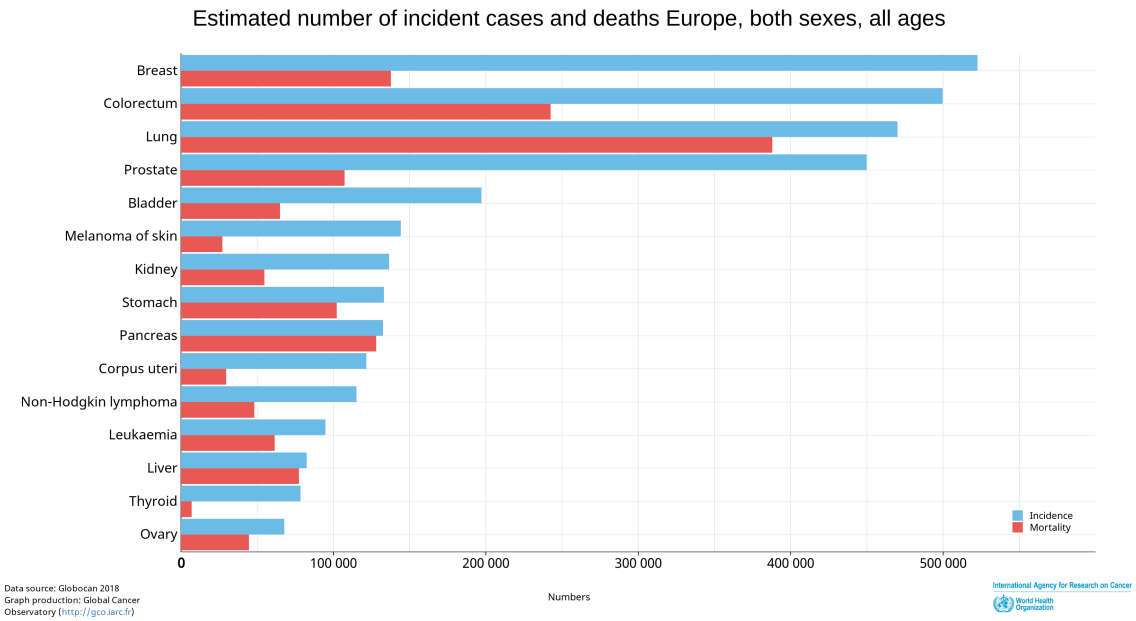


Fig. 3.2 Top fifteen tumours for incidence cases and deaths in Europe in 2020.

CAD systems, whereas the researches in the field of pathology, focusing on nephrology and haematology, are reported in Section 3.3.

3.2 Radiology

Concerning the radiological area, several works have been conducted in order to support the clinical diagnosis of different pathologies. In this section, we will investigate three different sites of the body, namely breast, liver and kidneys. Breast cancer and liver carcinoma will be described and analysed from the classification point of view, in order to detect and stage the tumours. Kidneys, instead, will be investigated considering a genetic disorder, namely the Autosomal Dominant Polycystic Kidney Disease (ADPKD), using diagnostic imaging techniques, thus intelligent algorithms for image processing and segmentation, in order to assess and estimate the volume occupied by the diseased organs.

Specifically, breast acquisitions included images from Magnetic Resonance (MR) and Digital Breast Tomosynthesis (DBT), whereas the liver acquisitions and analysis included images from Computed Tomography (CT) using Contrast Medium for implementing a triphasic protocol. Kidney images, instead, were obtained from MR imaging protocols, without any contrast medium for image enhancing.

3.2.1 Breast Cancer

The high incidence of breast cancer in women (more than 25 % of cancers affecting women is breast cancer [211]) and the ever-increasing life expectancy of the population require an accurate assessment of the breast glands with imaging techniques. In this field, mammography still represents the gold standard imaging tool [217]; in fact, mammographic examinations are currently employed in several screening programs all over the world, thanks to their capability to perform a very early detection; Magnetic Resonance (MR), Computer Tomography (CT) or Digital Breast Tomosynthesis (DBT) techniques, instead, become necessary to perform a more in-depth analysis of risky subjects in diagnostic tests, or for the follow-up of treated patients [218].

Although over the years the detection and classification of breast lesions have been accomplished with traditional image processing techniques and Machine Learning algorithms, several works have been presented dealing with breast lesions characterisation considering innovative Deep Learning (DL) strategies, especially using Convolutional neural Networks (CNN) for classify images or Regions of Interest (ROI).

In [219] Samala *et al.* designed a DL-CNN architecture for breast micro-calcification classification. The authors compared a DL-CNN architecture, whose optimal structure was obtained by varying among 216 combinations of parameters in the network (e.g., the number of filters and the filter kernels) and analysing the effects of their variation in the parameter space, and a previously designed Artificial Neural Network (ANN) performing convolutions on the input images. The results reported in the work using the approach based on Deep Learning showed a statistically significant improvement since the Area Under the Curve (AUC) improved from 0.89 for the ANN approach to 0.93 obtained with the DL-CNN.

Kallenberg *et al.* presented a method capable of learning from features, at multiple scales hierarchy, using unlabelled data, addressing two different tasks: (i) breast density segmentation and (ii) scoring of mammographic texture [150]. The authors reported that the scores obtained by performing the proposed approach, based on automatic learning, have a high correlation with those obtained with the investigated manual approach making use of cumulative thresholds set by expert users. Furthermore, the features scores learned by the automatic systems texture revealed to be also predictive of breast cancer.

3.2.1.1 Research Contribution

In this section, two different strategies designed and implemented for detecting and classifying breast lesions will be analysed and discussed. In particular, a supervised approach

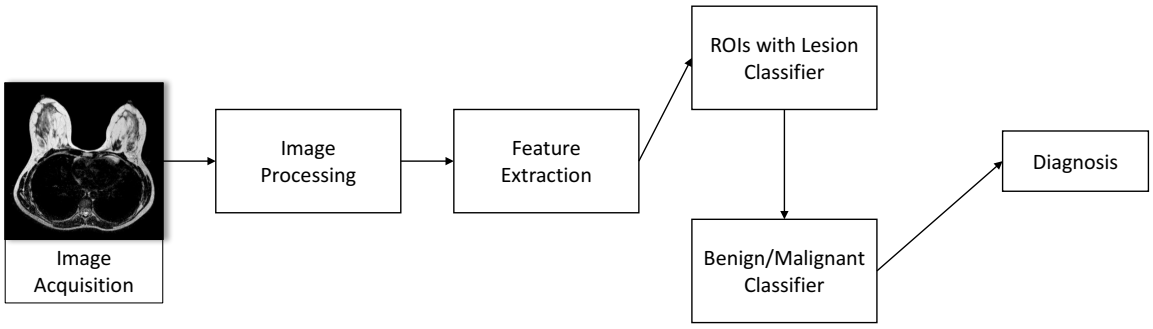


Fig. 3.3 Workflow for breast lesion classification.

for the detection and classification of breast lesions from MR images will be introduced in the first paragraph, whereas a supervised approach based on Deep Learning architectures for the extraction of automatically-computed features which allow one to classify using simpler non-neural strategies will be discussed in the subsequent one. In this latter case, some comparisons with neural approaches and classifiers working on hand-crafted features will also be reported.

Classification based on Magnetic Resonance Images

This section describes a Decision Support System for detecting and classifying breast lesions in images acquired via Magnetic Resonance. This work has been published in [107]. The workflow implemented in this work is reported in Fig. 3.3. After the acquisition of images, there are four different steps in order to allow discriminating between benign and malignant breast lesions, namely preprocess of the input images, extraction of features describing some regions of interest, detection of lesions among candidate regions and classification of the lesions.

The acquisition phase was conducted following the diagnostic protocol for breast cancer diagnosis, which consists in the following procedure:

- Transverse short T1 inversion recovery (STIR) turbospin-echo (TSE) sequence (TR/TE/TI = 3.800/60/165 ms, field of view (FOV) = 250x450 mm (APxRL), matrix 168x300, 50 slices with 3-mm slice thickness and without gaps, 3 averages, turbo factor 23, resulting in a voxel size of 1.5 x 1.5 x 3.0 mm^3; acquisition time: 4 minutes);
- Transverse T2-weighted TSE (TR/TE = 6.300/130 ms, FOV = 250x450 mm (APxRL), matrix 336x600, 50 slices with 3-mm slice thickness and without gaps, 3 averages, turbo factor 59, SENSE factor 1.7, resulting in a voxel size of 0.75x0.75x3.0 mm^3; acquisition time: 3 minutes);

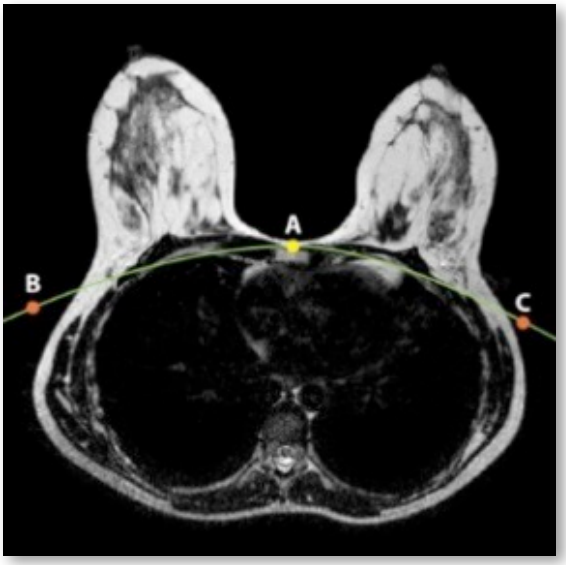


Fig. 3.4 Algorithm output for thorax masking. The reference points for parabola generation are A, B and C.

- Three-dimensional dynamic, contrast-enhanced (CE) T1-weighted high resolution isotropic volume (THRIVE) sequences (TR/TE= 4.4/2.0 ms, FOV = 250x450x150 mm (APxRLxFH), matrix 168x300, 100 slices with 4-mm slice thickness, spacing between slices: 2 mm; turbo factor 50, SENSE factor 1.6, 6 dynamic acquisitions, resulting in 1.5 mm^3 isotropic voxels, a dynamic data acquisition time of 1 min 30 s, and a total sequence duration of 9 min).

In details, the first step of the research work consisted in preprocessing the acquired images; after rescaling the grey levels of the pixels, an algorithm for masking the thorax in all the slices was performed, thus filtering out all the structures external to the breasts. As reported in Fig. 3.4, three points were computed to generate a parabola with the vertex on the sternum (point A) and passing through the side edges of the chest (points B and C).

A segmentation phase, necessary for the removal of all the uninteresting parts of the image, was subsequently performed. In particular, a thresholding operation was executed, considering the 95th percentile of the grey levels histogram of the images acquired without contrast medium (CM), considering the Transverse T2-weighted TSE sequence. Then, areas with diameter below 5 mm were removed and the obtained mask was applied to the starting image in order to extract the regions of interest. The remaining areas, candidate regions to be lesions, were characterised considering 10 features, which were:

- **F1**: size in mm^2 of the suspicious lesion;

- **F2, F3**: average value of the grey levels of images with and without CM in the ROIs, to determine areas with the grey intensities different from the standard ones;
- **F4, F5**: standard deviation value of the grey levels of images with and without CM in the ROIs, respectively;
- **F6**: circularity expressed as $4\pi(\frac{A}{p^2})$ where $A = ROI_{area}$ and $p = ROI_{perimeter}$;
- **F7**: aspect ratio, intended as $\frac{majorAxis}{minorAxis}$;
- **F8**: eccentricity of the ellipse whose second order moments coincide with those of each ROI;
- **F9**: solidity, defined as ratio of the area of the ROI and the area of convex hull;
- **F10**: convexity (or edge roughness), given by the ratio between the perimeter of the convex hull and the one of the ROI.

After the extraction of the features, two different Artificial Neural Networks (ANN) were designed using the evolutionary strategy reported in a previous work [108]. The first ANN was designed to discriminate among lesions and other structures (e.g., vessels) for detecting lesions among all the candidate regions identified after the image processing steps. The second classifier, instead, was used to discriminate benign and malignant lesions, taking as input the same set of features of the regions classified as lesions in the previous step.

Regarding the first classifier, the image processing and segmentation pipeline revealed more candidate regions (i.e., False Positives) respect to the real lesions, generating an unbalanced dataset for adequately training a neural classifier. Using the Synthetic Minority Over-sampling Technique (SMOTE), input dataset was balanced by increasing the number of patterns characterising lesions up to the same amount of negative cases, in order to achieve better classifier performance. [220]. For both the implemented classifiers, performances were measured in terms of Accuracy, Sensitivity and Specificity, according to the equations reported in Section 2.4.4.

The results obtained in the two classifications steps of the pipeline are reported in Table 3.1, and 3.2, respectively. In details, the reported tables show Accuracy, Sensitivity and Specificity as mean values obtained performing 100 repetitions of the training, validation and test of the ANN, considering a different random permutation of the input dataset at each iteration.

The reported results show that a supervised machine learning approach for the detection of breast lesions from MR images and the subsequent classification between benign and

Table 3.1 Results obtained in for the discrimination between ROIs with and without lesions.

	Min	**Max**	**Mean**
Accuracy	*0.9624*	*0.9849*	0.9736 ± 0.0044
Sensitivity	*0.9592*	*0.9958*	0.9791 ± 0.0075
Specificity	*0.9459*	*0.9892*	0.9684 ± 0.0075

Table 3.2 Results obtained for the discrimination between benign and malignant lesions.

	Min	**Max**	**Mean**
Accuracy	*0.7308*	*1*	0.8977 ± 0.0584
Sensitivity	*0.6923*	*1*	0.8908 ± 0.1021
Specificity	*0.7692*	*1*	0.9046 ± 0.0875

malignant is consistent, and shows good performance, especially from the False Negative reduction perspective.

Classification based on Digital Tomosynthesis Images

Digital Breast Tomosynthesis (DBT) has been recently introduced for breast cancer screening and detection, and is a promising innovative radiological technique for early diagnosis and staging [221]. DBT produces a limited angle cone-beam tomosynthesis of the breast glands and has demonstrated to have a higher accuracy if compared to the most commonly used bi-dimensional imaging techniques, such as mammography, CT or MR [217, 222–224]. After the acquisition of multiple thin and high-resolution images, a 3D model of the breast is created, also reducing the effect of tissue superimposition with respect to the other imaging techniques [225]. DBT also improves the visualisation of masses and architectural distortions. In particular, the edges of the breast lesions are better defined, thus leading to an improvement of the final diagnosis performance [226].

Starting from DBT images, a CAD system to support the detection and classification of three different kinds of lesions was designed and developed. The workflow implemented, allowing to discriminate among different kinds of breast lesions, is represented in Figure 3.5. Specifically, the workflow includes a pre-processing images phase, fundamental for extracting regions of interest, a step for extracting features and, lastly, the classifier design for discriminate the different lesions. This research work has been published in [227].

As for the approach based on MR images reported in the previous paragraph, a preliminary processing step performing the segmentation of images was needed to extract candidate

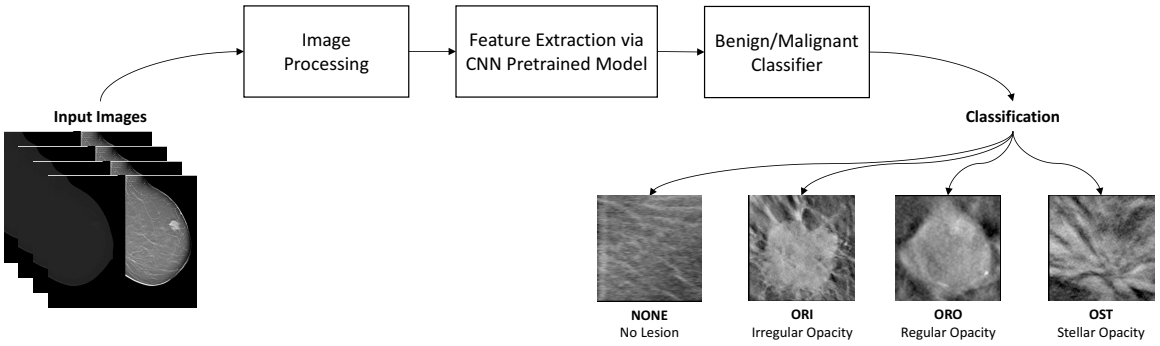


Fig. 3.5 Workflow for breast lesion classification.

regions of interest containing suspicious areas. In details, the extracted ROIs were labelled according to the classification by radiologists in four classes:

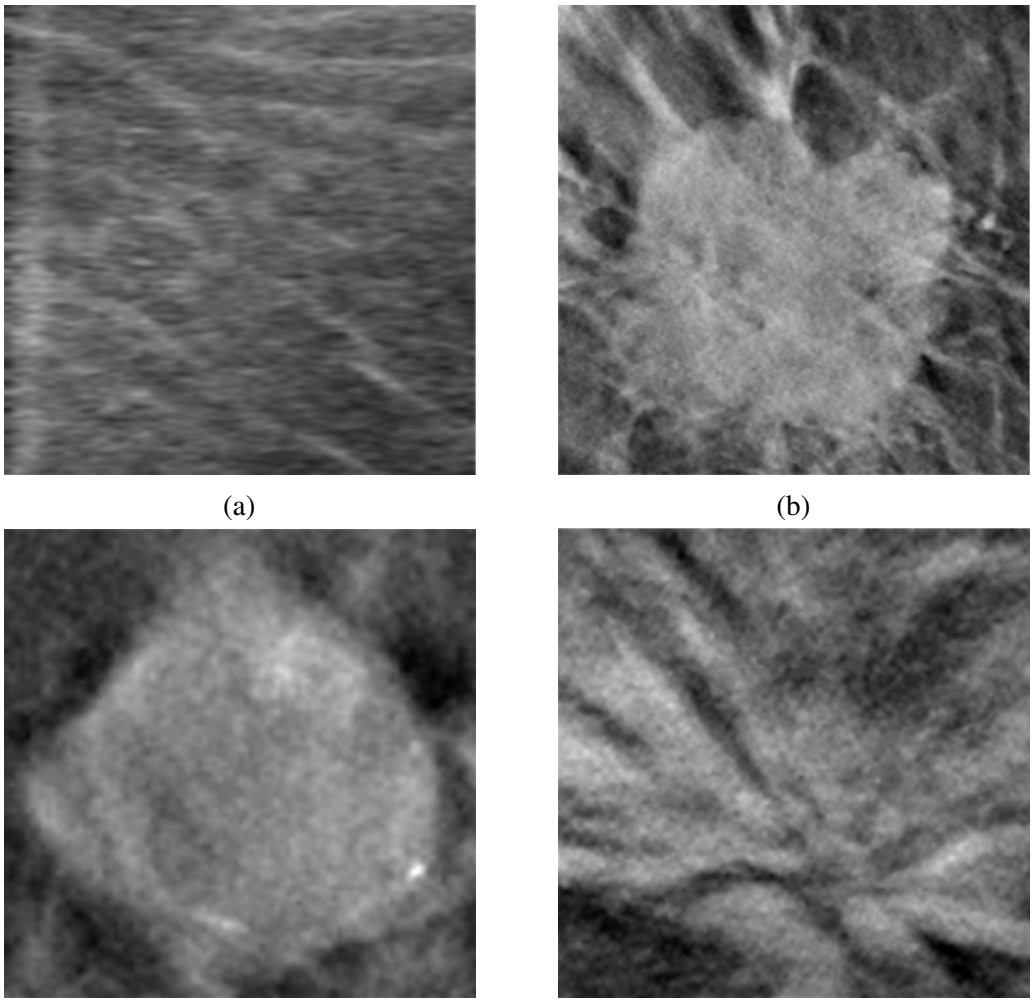
1. **None**: segmented ROI not containing any kind of lesion (Fig. 3.6 (a));
2. **Ori**: segmented ROI containing an irregular opacity (Fig. 3.6 (b));
3. **Oro**: segmented ROI containing a regular opacity (Fig. 3.6 (c));
4. **Ost**: segmented ROI containing a stellar opacity (Fig. 3.6 (d)).

Differently from the work described in the previous paragraph, the classification of lesions was investigated following two different approaches:

- Using several pre-trained models of Convolutional Neural Networks as automatic generator of features. The extracted pattern of features was then classified by means of non-neural algorithms.
- Morphological and textural features were computed, by analysing the Grey Level Co-occurrence Matrix (GLCM) [91, 228, 229]. In this case, the features extracted were classified by Artificial Neural Networks.

Both the investigated approaches have been implemented and compared.

Hand-Crafted Features and ANN Classification Concerning with the approaches using Artificial Neural Networks as classifiers, the ROIs extracted were described using both morphological and textural features computed by means of the Grey Level Co-occurrence Matrix (GLCM). In fact, these features are very frequently used in similar research



(a) (b)

(c) (d)

Fig. 3.6 Images extracted after the segmentation phase: (a) ROI with no lesions; (b) ROI with irregular opacity; (c) ROI with regular opacity; (d) ROI with stellar opacity.

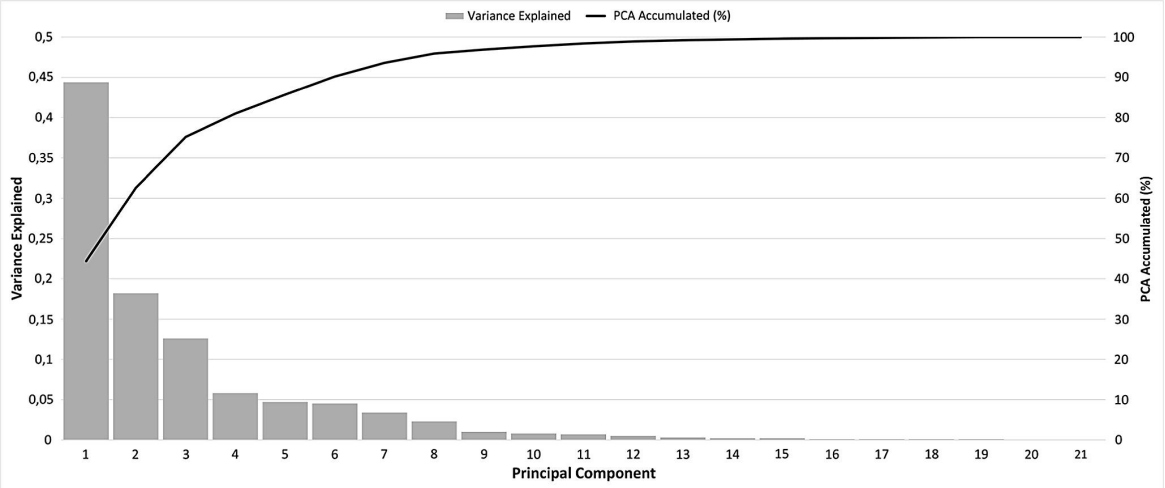


Fig. 3.7 PCA explained variance of the principal components computed from the initial dataset. Bars show the variance for each component, whereas the black line indicate the cumulative variance for each principal component.

works [163, 230]. In particular, 26 different features were computed and the arithmetic mean of pixels' grey levels in the region was also considered.

In order to reduce the number of features to be considered for the classification, dimensionality reduction was performed using the eigen-decomposition Principal Component Analysis (PCA) [231, 232]; finally, a set of 13 features in the new subspace was selected to cover an overall variance of 99.3 %, as could be seen in Fig. 3.7.

Two different Artificial Neural Networks were designed, by adopting an evolutionary approach, to discriminate among the four classes. In a first step, Ori, Oro and Ost classes were grouped as a single class (labelled as Positives - P), while None was the second class (labelled as Negative - N). In this case, a binary classifier was used obtaining the performance reported in Table 3.3.

Subsequently, a second ANN was employed to discriminate among the four classes, but the obtained Accuracy was lower than those obtained with the binary classifier, reaching 74.84 % $\pm$ 4.89.

Analysing the obtained results, this behaviour was predictable and reasonable, considering the way in which ANNs were designed and optimised: the extraction process of the most discriminant features heavily influenced the overall capabilities of the classifier. In this case, the features used as inputs achieved good results in discriminating samples between two classes, while they did not provide enough information to correctly describe, thus discriminate, all the different kinds of lesions in the multi-class approach.

Table 3.3 Results obtained for binary classification.

	Accuracy	**Sensitivity**	**Specificity**
Mean	*84.19 %*	*85.90 %*	*82.82 %*
Standard Deviation	*3.06*	*5.33*	*5.17*

CNN-based Features and Non-Neural Classification As already reported in Section 2.5, Deep Learning strategies, specifically Convolutional Neural Networks, are often employed for automatically generating features from input images [233]. Regarding CNNs as feature extractors, in this work several pre-trained models were considered, namely GoogLeNet, ResNet, AlexNet, VGG-verydeep, VGG-F, VGG-M and VGG-S [178, 234–240].

In this approach, the dataset for classifying lesions was constituted by the set of features extracted using the CNN models. The output of the CNN models (in other words, their final activations) was used to train several non-neural learners, which were: Linear Support Vector Machine (Linear SVM), K-Nearest Neighbor (KNN), Naïve Bayes, Decision Tree and Linear Discriminant Analysis (LDA) [241–246].

Final tests were performed considering all the investigated CNN architectures. In order to improve the overall classification performance, several tests for processing the dataset before training were performed. Specifically, Activation Normalization and Dataset Augmentation were both explored leading to a slight improvement in all the architectures employing both the strategies. In fact, normalization allows each layer of a network to learn by itself a little bit more independently of other layers, whereas the dataset augmentation allows enriching the dataset with generated samples, increasing the generalisation capabilities of the architecture [247, 248]. Finally, since VGG-F, VGG-M and VGG-S achieved the higher mean accuracies and the lowest processing time, they were considered for the final evaluation.

The results are represented in Table 3.4 and show that the Naïve Bayes classifier is not recommended in this classification task; Decision Trees allowed to improve the mean accuracy in comparison to Naïve Bayes, but it was still far to be considered as a reliable classifier for this problem. The linear classifiers (SVM and LDA) further increased the performance, but better results were obtained from the KNN classifier. In this case, the mean performance reached very good levels of accuracy, specificity and sensitivity, also with low variability among the training and test repetitions. Furthermore, to substantiate the high level of performance in terms of accuracy, it is worth to mention that sensitivity and specificity

Table 3.4 Results of the selected pre-trained CNNs used as features extractor, training several learners with normalization and augmented images.

	KNN	***LDA***	***LINEAR SVM***	***NAÏVE BAYES***	***DECISION TREES***
VGG-F	91.63 ± 0.41	64.57 ± 0.66	67.29 ± 2.02	43.82 ± 0.59	59.68 ± 1.07
VGG-M	90.74 ± 0.48	66.25 ± 0.60	69.50 ± 2.16	42.85 ± 0.57	57.03 ± 0.98
VGG-S	92.02 ± 0.48	65.24 ± 0.80	68.84 ± 1.89	44.89 ± 0.60	56.16 ± 0.93

Table 3.5 Sensitivity and Specificity for the lesions evaluated through 1-vs-all approach.

	Ori vs all		**Oro vs all**		**Ost vs all**	
	Sensitivity	***Specificity***	***Sensitivity***	***Specificity***	***Sensitivity***	***Specificity***
VGG-F	*98.67 %*	*97.07 %*	*96.01 %*	*95.98 %*	*97.24 %*	*96.93 %*
VGG-M	*98.14 %*	*96.64 %*	*95.00 %*	*95.76 %*	*97.18 %*	*96.62 %*
VGG-S	*98.36 %*	*97.13 %*	*95.61 %*	*96.33 %*	*96.67 %*	*97.25 %*

for positive samples were higher than 95 %, as reported in Table 3.5, where the results were calculated using a 1-vs-all approach.

3.2.2 Liver Carcinoma

As for breast cancer, liver carcinoma shows an extremely high mortality worldwide [211]. In recent years, several works have been presented dealing with detection and classification of hepatic tumours considering different strategies for their discrimination. In fact, segmentation of liver is crucial for several clinical procedures, including radiotherapy, volume measurement and computer - assisted surgery.

The design and implementation of a CAD system consisting of both liver and tumour segmentation, feature extraction and classification modules was presented in [249], characterising the liver tumour from CT images between haemangioma and hepatoma. The experimental results showed that the classification accuracy of a Fast Discrete Curvelet Transform (FDCT)-based feature extraction and classification approach was higher than a previously investigated wavelet based method. Performance measurements could also be improved by increasing the number of considered samples. In the considered work, a pattern recognition network was used, a feed-forward network with tan-sigmoid transfer functions in both the hidden layer and the output one. The performance of the classifier was evaluated by

calculating accuracy (93.3 %), sensitivity (90 %) and specificity (96 %) from the obtained confusion matrix.

A Computer Aided Diagnosis system based on texture features and a multiple classification scheme for the characterisation of four types of hepatic tissue from Computed Tomography images has been presented in [250]. The proposed system achieved a total classification performance of about 97 %. Regions of Interest corresponding to normal liver, cyst, haemangioma, and hepatocellular carcinoma were drawn by an experienced radiologist on abdominal non-enhanced CT images, for labelling the ground truth. For each ROI, five distinct sets of texture features were extracted using the following methods: first order statistics, spatial grey level dependence matrix, grey level difference method, Laws texture energy measures, and fractal dimension measurements. Since the dimensionality of some feature sets was high, a feature selection strategy, based on a Genetic Algorithm (GA), was applied to extract the optima subset of features. Classification of the ROI was carried out by a system of five neural networks, each using as input one of the above feature sets. The members of the ANN system (primary classifiers) are 4-class ANNs trained by the back-propagation algorithm with adaptive learning rate and momentum. The final decision of the CAD system was based on the application of a voting scheme across the outputs of the individual ANNs. The multiple classification scheme using the five sets of texture features resulted in significantly enhanced performance.

A pilot study for hepatocellular carcinoma grading and the evaluation of microscopic vascular invasion was proposed in [251], where a shallow artificial neural network was compared to linear models. The results obtained from the ANN, in terms of AUC and Accuracy, were higher than the ones obtained from the linear logistic model in both Hepatocellular Carninoma grading and Microvascular Invasion presence evaluation.

In [252], biorthogonal wavelet based texture features were extracted and used to train the Probabilistic Neural Network (PNN) to classify the liver tumour, cholangio-carcinoma, hepatocellular adenoma and haemangioma with better performance to help radiologists and medical specialists during their medical decision process.

In [253], 164 liver lesions (80 malignant tumours and 84 haemangiomas) were evaluated. The suspicious tumour region in the digitised CT image was manually selected and extracted as a circular sub image. The proposed pre-processing adjustments for sub-images were used to equalize the information needed for a differential diagnosis. The auto-covariance texture features of sub-images were extracted, and a support vector machine classifier identified the tumour as benign or malignant. The Accuracy of the proposed diagnosis system for classifying malignancies was 81.7 %, the Sensitivity was 75 %, the Specificity was 88.1 %,

the Positive Predictive Value (PPV) was 85.7 % and the Negative Predictive Value (NPV) was 78.7 %, where PPV and NPV were evaluated according to Equations 3.1 and 3.2 respectively, considering the Confusion Matrix reported in Table 2.2.

$$Positive\ Predictive\ Value\ (PPV) = \frac{TP}{TP+FP} \tag{3.1}$$

$$Negative\ Predictive\ Value\ (NPV) = \frac{TN}{TN+FN} \tag{3.2}$$

Also for liver classification and segmentation, several Deep Learning approaches have been investigated in the last years, after the problem have been investigated following the traditional methodologies of image processing and Machine Learning [254–259]. For instance, a recent survey by Kumar *et al.* [260] on the topic of CAD systems for diagnosing livers affected by some pathologies does not consider the modern approaches based on deep learning classification techniques, but limit its analysis to traditional methodologies [261–271]

A method to segment liver and lesions automatically in CT and MR abdomen images was proposed in [272] by using Cascaded Fully Convolutional Neural Networks (CFCNNs), which enabled the segmentation of large-scale medical trials and quantitative image analyses. In particular, the authors focused on CT images and applied a CFCNN on CT slices for segmenting liver and lesions, leading to a significantly higher segmentation quality and showing interesting performance on a public challenge dataset.

Moreover, a Deep Convolutional Neural Network (DCNN) was developed to segment the liver in CT slices via an automatic procedure also in [273]. The same model was subsequently employed for the classification of lesions, by considering images from the previous classification as inputs. The developed models were evaluated using the Liver Tumour Segmentation Challenge dataset (LiTS) for the liver segmentation tasks. From the performance point of view, a DICE coefficient equal to 0.67 was reached, but the lesion classification performance was still low, as reported by the author himself.

A new method for the automatic detection and segmentation of unknown cancers in longitudinal liver CT studies and for a burden quantification of tumours has been presented in [274]. The considered inputs were the baseline/follow-up CT scans, the baseline delineation of tumours, and a tumour appearance prior model. The outputs, instead, were the new segmentations of tumours in the follow-up scan, tumour burden quantification in both scans, and the tumour burden change. The method developed in [274] aimed at finding new neoplasias by integrating information from the scans, the baseline tumours delineation, and

a tumour appearance prior model in the form of a global CNN classifier. Reported results showed that this method was superior to existing stand-alone/follow-up methods, with both high true positive rate and precision of about 86 % on a dataset with 37 longitudinal liver CT studies including the labelling of 246 tumours, among which 97 new ones.

A fully automatic framework for segmentation of the liver along with its tumour on contrast enhanced abdominal CT scans was developed in [275], based on three steps: i) liver localization by a simple CDNN model; ii) liver fine segmentation via a deeper CDNN model with reduced kernel sizes; iii) a tumour segmentation by a CDNN model with enhanced liver region as an additional input feature. The considered CDNN models were fully trained in an end-to-end mode with minimum pre- and post-processing efforts.

Furthermore, two CNNs were used for liver segmentation and metastases detection in CT examinations by Ben-Cohen *et al.* [276]. Authors used CNN architectures based on a VGG 16-layer net as in [237]. In particular, the final classification layer was removed and substituted with a [1×1] convolution, with channel dimension equal to 2, in order to predict scores for lesion or liver at each of the coarse output locations, followed by a deconvolution layer to upsample the coarse outputs to pixel-dense outputs, whereas all the intermediate fully connected layers were substituted with convolutional layers. Starting from the previous architectures, two variants were obtained linking lower layers to the final layer, which were FCN-8s DAG and FCN-4s DAG, respectively; 3D information was also considered by giving adjacent slices as input to the network. Moreover, regarding the segmentation phase, Dice index, Sensitivity and Positive Predictive Values were used to evaluate the segmentation performance. In particular, higher values for all three indexes were reached using the FCN-8s DAG combining information from the 2 adjacent CT slices of the considered one. Regarding the detection of metastases, instead, the True Positive Rate (TPR) and False Positive per Case (FPC) metrics were used for performance evaluation. In this case, 3D information allowed the authors to reach the highest performance with the FCN-4s architecture. Although the results obtained in this work were very promising, the dataset considered was too small for the algorithm to be considered robust and accurate.

A novel application of Convolutional Neural Networks was presented in [277] to segment liver tumours. In the considered work, CNNs architectures were tested on 30 CT images using leave-one-out cross validation. The reported experiments showed that the CNN model produced an accurate and robust liver tumour segmentation, if compared to traditional machine learning methods, such as AdaBoost and SVM. Limitations of CNNs were still found on segmenting tumours with inhomogeneous density and unclear boundary, especially noticing an under-segmentation in the tumour adjacent to structures with similar densities.

3.2.2.1 Research Contribution

In the following sections, two specific works for detecting and classifying hepatocellular carcinoma (HCC) will be analysed and discussed. In particular, a supervised approach based on the extraction of hand-crafted features for the detection and classification of HCCs from a triphasic CT protocol will be introduced in the first paragraph, whereas a second approach based on a supervised Convolutional Neural Network will be discussed in the subsequent one.

Hepatocellular carcinoma is the most common type of primary liver cancer in adults, and is the most common cause of death in people with cirrhosis [278]. As with any cancer, the treatment and prognosis of HCC vary depending on the specifics of tumour histology, size, how far the cancer has spread, and overall health. Methods of diagnosis in HCC have evolved with the improvement in medical imaging. The evaluation of both asymptomatic patients and those with symptoms of liver disease involves blood testing and imaging evaluation. Although historically a biopsy of the tumour was required to prove the diagnosis, imaging (especially MRI and CT) findings may be conclusive enough to obviate histopathologic confirmation.

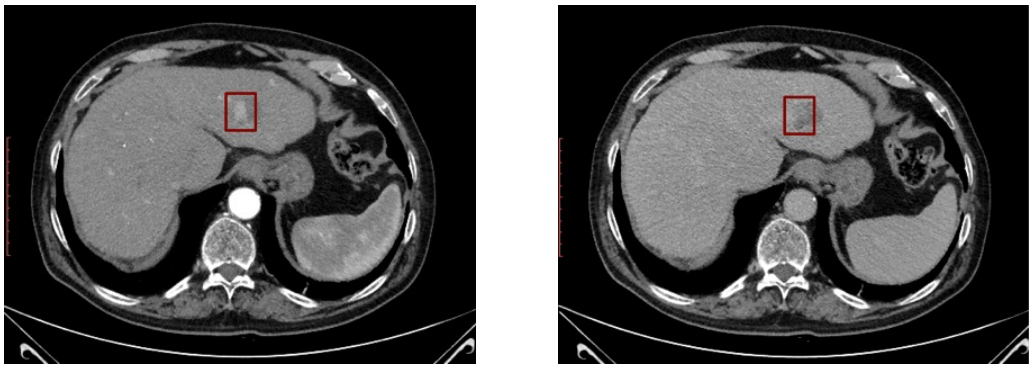
In the considered approaches for HCC staging, the acquisition system was based on Computer Tomography, a medical imaging technique that is widely used for Hepatocellular Carcinoma detection. In fact, CT, as well as MR, are always required to determine the disease extension. Both techniques are considered as the gold standard for non-invasive evaluation of focal and diffuse diseases of the liver and the biliary tract [279, 280].

The CT scans were acquired with a 320 slices Scanner (Toshiba Aquilion One) after an automated injection of 1.5 ml/kg of iodinated contrast medium (Iomeprole 400 mgI/ml) through a 16G Needle in antecubital vein at a flow rate of 4ml/sec with the following protocol:

1. **arterial dominant phase** acquired 20 seconds after the aortic peak calculated by a bolus tracking system with a ROI positioned in the abdominal aorta at a trigger density of 150 Hounsfield Units (HU) (Fig. 3.8 (a));
2. **portal phase** acquired 70 seconds after contrast injection;
3. **equilibrium phase** acquired 180 seconds after contrast injection (Fig. 3.8 (b)).

Detection and Segmentation Considering Hand-Crafted Features

In this work, published in [281], a double-step segmentation was carried out, after an image



(a) (b)

Fig. 3.8 CT images acquired in the different phases: (a) Arterial phase; (b) Equilibrium phase. The square indicates the lesion in both phases.

pre-processing phase performed through contrast enhancement, in order to improve the contrast of the CT images, and a cropping phase to reduce the amount of data to be processed. In particular, liver and HCC segmentations were performed.

Concerning liver segmentation, the histogram of the slice with the largest connected portion of the liver was analysed to identify the typical liver grey intensity. By means of this evaluation, local thresholding and morphological operations were performed, allowing to remove all the structures external to the liver. The segmentation result of the first CT slice was dilated and used as a binary mask for the previous and the following slices. Those slices were further segmented, and the method was iteratively propagated in both directions until the final slices were reached. Representation of the segmentation steps is reported in Fig. 3.9.

A similar procedure was performed also for the HCC segmentation. Hepatocellular carcinomas were subsequently segmented using an innovative two-dimensional region growing algorithm which took into account images from both arterial and equilibrium phases. The description of the implemented algorithm is reported in Fig. 3.10. Some morphological operation were performed to refine the obtained Regions of Interest. The output of the HCC segmentation procedure is reported in Fig. 3.11

Textural features were then extracted using grey level co-occurrence matrices, as proposed by Haralick *et al.* [91] and derived from his works [228, 229] to describe the ROIs. Each HCC was texturally described considering the information from the two corresponding slices in both the arterial and the delayed phases.

Three different subsets of features were generated from the dataset:

1. 44 features obtained considering ROIs extracted from the two phases (specifically, 22 features extracted from each phase);

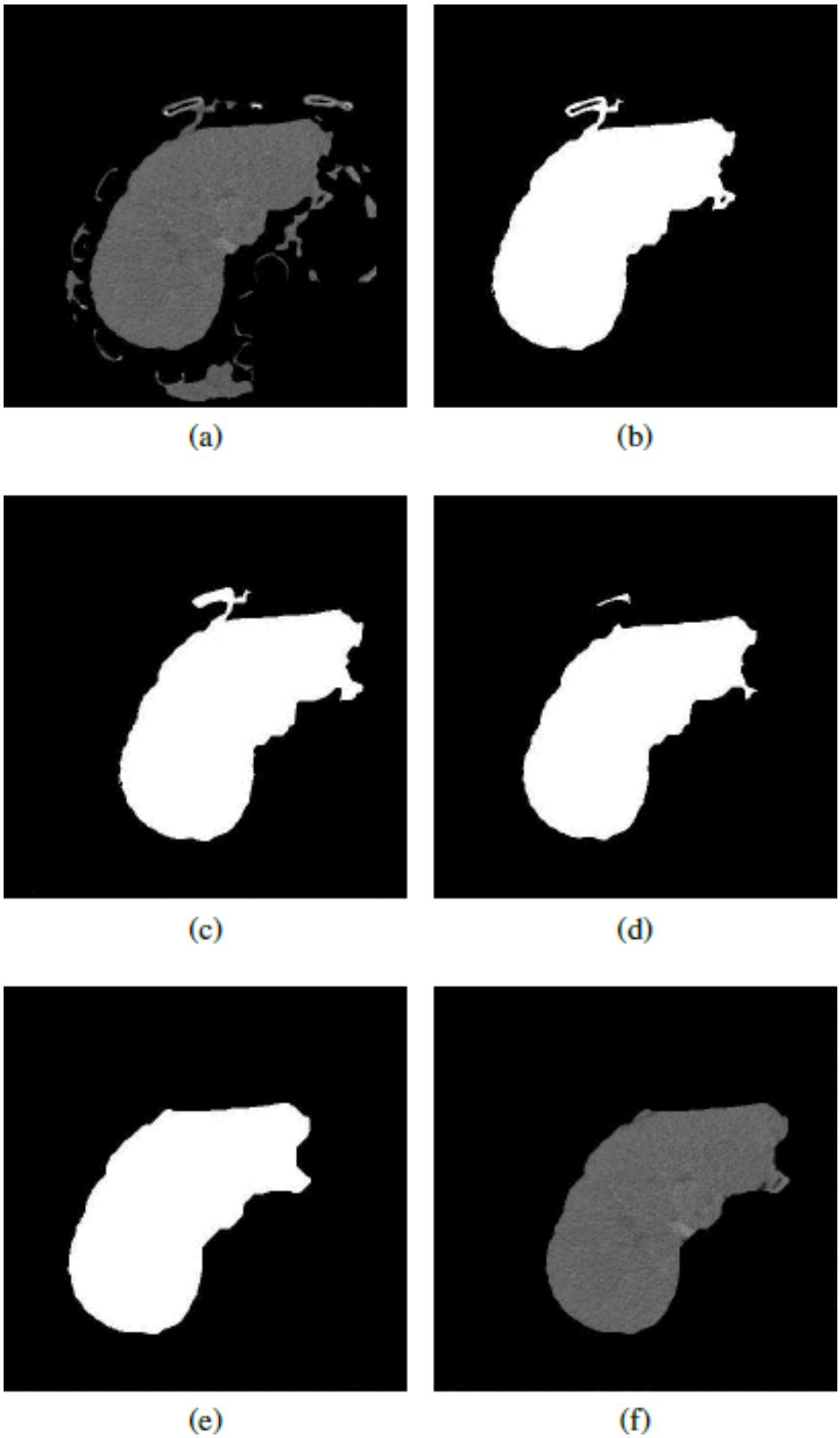


Fig. 3.9 Output of liver segmentation steps: (a) Local thresholding; (b) Largest connected component extraction; (c) Hole filling; (d) Erosion; (e) Largest connected component extraction, dilation and closing; (f) Segmented liver.

Data: P_1, P_2 images of two CT phases; S a set of seed points
Result: R region including the liver
Estimate typical liver intensities $(L1, L2)$ as highest peaks, not including background, of histograms belonging to P_1 andP_2;
foreach *seed point* $s \in S$ **do**
  Add a 3x3 neighborhood of s to the region R;
  Get 9x9 neighborhood of s from P_1 andP_2 and add them to the respective computation sets $(R1, R2)$
end
Add neighbors of R to candidate pixel set (C);
Calculate *mean* (μ_1, μ_2) and *standard deviation values* (σ_1, σ_2) of $R1$ and $R2$;
if $L_1 - \mu_1 > \alpha$ **then**
  characterize the lesion as hyperintense;
else
  characterize the lesion as isointense;
end
while R *continues to grow* **do**
  if *lesion is hyperintense* **then**
    Set $G = \{(x,y)|(x,y) \notin R \wedge (x,y) \in C \wedge ((P_1(x,y) \geq L_1 \wedge P_2(x,y) < \mu_2+\sigma_2) \vee (P_1(x,y) > \mu_1-\sigma_1 \wedge P_2(x,y) \leq L_2))\}$;
  end
  if *lesion is isointense* **then**
    Set $G = \{(x,y)|(x,y) \notin R \wedge (x,y) \in C \wedge P_1(x,y) \geq L_1 - \beta \wedge P_2(x,y) < \mu_2 + \sigma_2\}$;
  end
  if $G \neq \emptyset$ **then**
    Add members of G to R;
    Add neighbors of R to C;
    Set $R1 = \{P_1(x,y)|(x,y) \in R\}$ and $R2 = \{P_2(x,y)|(x,y) \in R\}$;
    Recalculate *mean* and *standard deviation values* of $R1$ and $R2$;
  end
end

Fig. 3.10 The Developed Region Growing Algorithm.

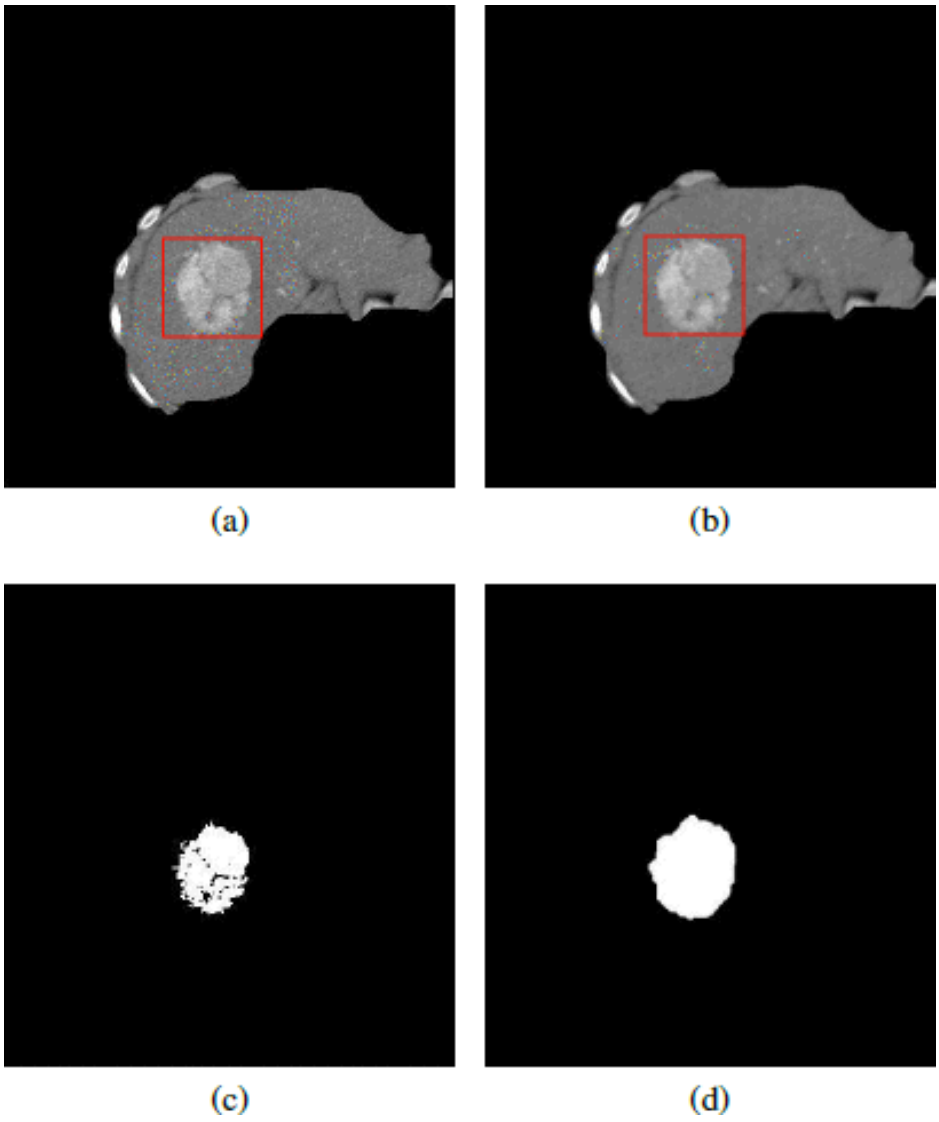


Fig. 3.11 Output of HCC segmentation steps: (a) Segmented liver; (b) Median filtering; (c) Region growing; (d) Closing, dilation and opening.

Table 3.6 Results obtained for HCC classifier

Dataset	Accuracy	Sensitivity	Specificity
44 Features	0.758 ± 0.062	0.755 ± 0.122	0.739 ± 0.141
22 Features	0.763 ± 0.063	0.824 ± 0.089	0.698 ± 0.156
10 Features	0.799 ± 0.073	0.795 ± 0.015	0.804 ± 0.126

2. 22 features coming from a ranking algorithm based on the relative entropy, also known as Kullback-Leibler distance or divergence [97];

3. 5 features in Haralick *et al.* (Contrast, Correlation, Energy, Homogeneity and Entropy) which have been used in similar previous work for blood vessels and tubules classification [163] and have shown good discrimination capabilities.

HCC grades, which could be classified into four classes (from grade 1 to grade 4) [282], have been grouped into two classes: grade 1/2 were addressed as Negative Class, while grade 3/4 were addressed as Positive Class.

Since the optimisation of classifiers, especially Artificial Neural Networks training and topology, was quite common in literature [283–289], in the considered work a binary classifier was designed to discriminate the two groups mentioned above using a single-objective genetic algorithm (GA) [108]. Results, which are expressed in terms of mean values ($\pm$ standard deviation), considering 100 repetitions of training validation and test of the ANN classifier, for accuracy, specificity, and sensitivity, are reported in Table 3.6.

The results obtained in this work show that HCC classes can be discriminated by using the proposed set of extracted features: the HCC wash-in and wash-out dynamic suggests that this type of lesion can be characterized by processing textural differences considering the two fundamental phases in the HCC dynamic.

The Deep Learning Approach

The Deep Learning approach was investigated considering the images acquired with the acquisition protocol as described in paragraph 3.2.2.1.

To better understand the grading process, the following block diagram of the approach, proposed in [290] is reported in Fig. 3.12.

Regarding liver and HCC segmentations, the same procedures described in the previous section were applied, thus obtaining binary masks that were applied to the CT images acquired during the arterial phase. A Convolutional Neural Network, specifically the Google

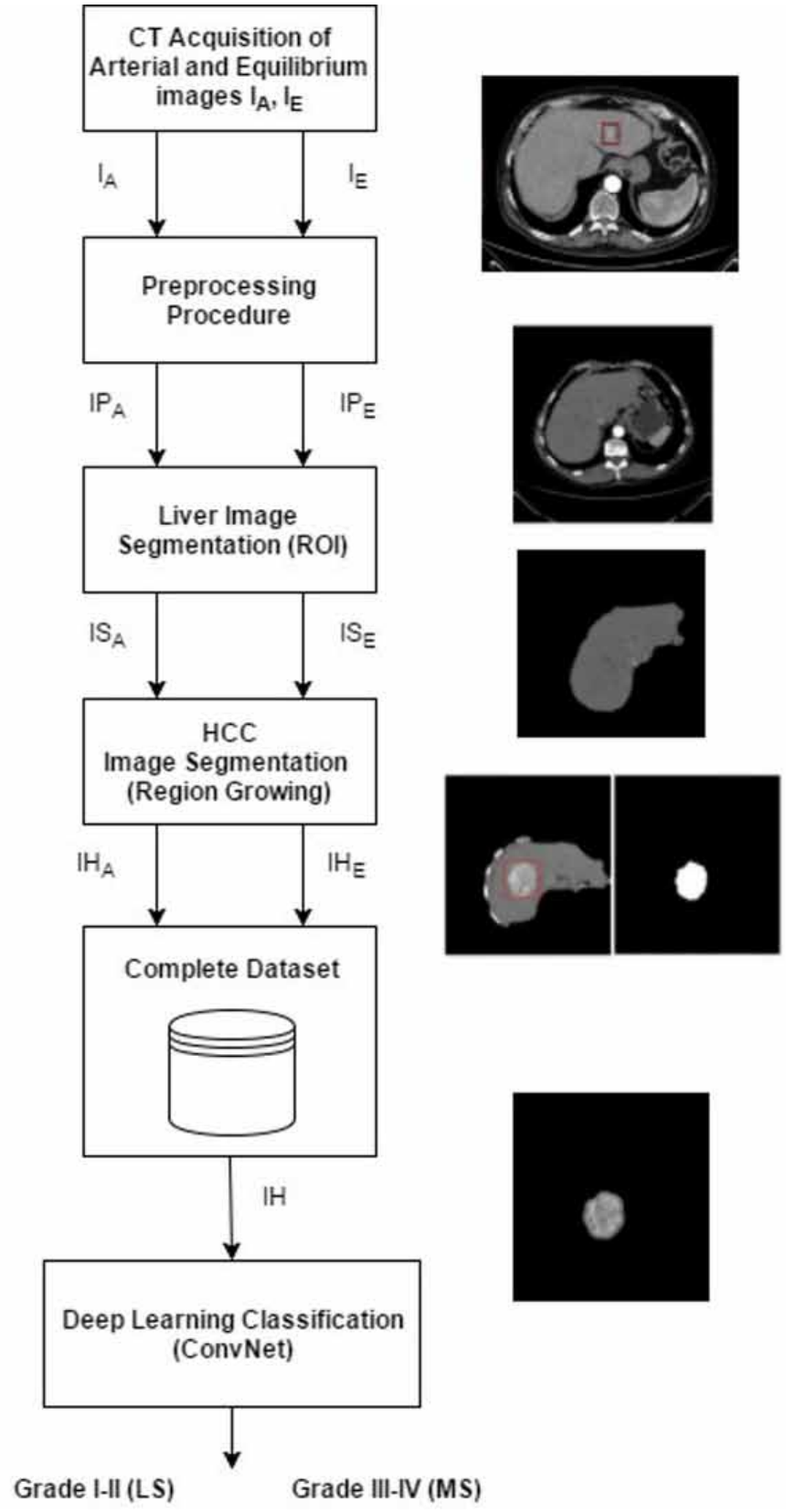


Fig. 3.12 Block diagram of the proposed approach for hepatocellular carcinoma grading with samples of output images at each step.

Table 3.7 Results obtained for CNN classification.

	Accuracy	**Sensitivity**	**Specificity**
Mean	*0.928*	*0.935*	*0.921*
Standard Deviation	*0.055*	*0.075*	*0.089*

Inception v3 implementation reported in [291], was used by re-training the model in order to classify using the created dataset of images as input.

The implemented CNN has been trained, validated and tested 100 times considering different permutations of the input dataset. The results reported in Table 3.7 show that the implemented approach reached high generalization performance regardless of the input permutation; in fact, Accuracy, Sensitivity and Specificity were higher than 90 % with low standard deviation.

3.2.3 Autosomal Dominant Polycystic Kidney Disease

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is a hereditary disease characterised by the onset of renal cysts that lead to a progressive increase of the Total Kidney Volume (TKV) over time. Specifically, ADPKD is a genetic disorder in which the renal tubules become structurally abnormal, resulting in the development and growth of multiple cysts within the kidney parenchyma [292]. The mutation of two different genes characterises the disease. The ADPKD type I, which is caused by the PKD1 gene mutation, involves 85 - 90 % of the cases, usually affecting people older than 30 years. The mutation of the PKD2 gene, instead, leads to ADPKD type II (affecting 10 - 15 % of the cases), which mostly regards children developing cysts already when in the maternal uterus and dying within a year. However, the clinical features of patients with PKD1 or PKD2 mutations are the same, but the latter is associated with a milder clinical phenotype and a later onset of End-Stage Kidney Disease (ESKD). In all cases, the size of cysts is extremely variable, ranging from some millimetres to 4 - 5 centimetres [293]. Currently, there is no specific cure for ADPKD and the TKV estimation over time allows one to monitor the disease progression. Tolvaptan has been reported to slow the rate of cysts enlargement and, consequently, the progressive kidney function decline towards ESKD [294, 295]. However, as all the pharmacological treatments aimed at slowing the growth of the cysts, a non-invasive and accurate assessment of the renal volume is of fundamental importance for the estimation of the ADPKD severity over time.

There are several methods in the literature performing TKV estimation; traditional methodologies, requiring medical image acquisitions, such as Computed Tomography and Magnetic Resonance, include stereology and manual segmentation [296, 297]. Also, several studies tried to correlate this metric with body surface and area measurements in order to have a non-invasive estimation of TKV [298, 299].

Stereology consists in the superimposition of a square grid, with specific cell positions and spacing, on each volume slice; the TKV estimation is obtained merely counting all the cells which contain parts of the kidneys to get the bi-dimensional area on each slice, and interpolating all the slices considering the thickness of acquisition to get the final volume. Manual segmentation, instead, requires manual contouring of the kidney regions contained in every slice; several tools supporting this task have been developed, introducing digital free-hand contouring tools or interactive segmentation systems to assist the clinicians while delineating the region of interest.

Considering both the phenotyping of the disease and the introduced approaches, the segmentation of biomedical images of kidneys is a tricky and troublesome task, strictly dependent on the human operator performing the segmentation, requiring expert training. In fact, co-morbidities and the presence of cysts in neighbouring organs or contact surfaces make challenging achieving an accurate assessment of the TKV. To overcome the limitations of the manual methodologies, several approaches to perform the semi-automatic segmentation of kidneys and TKV computation have been investigated, such as the mid-slice or the ellipsoid methods, allowing to estimate the TKV starting from a reduced number of selected slices [300–302]. Although the reported methodologies are faster and more compliant than the previous ones, these are far from being accurate enough to be used in clinical protocols [303, 304].

In recent years, innovative approaches based on Deep Learning strategies have been introduced for the classification and segmentation of images. In details, deep architectures, such as Deep Neural Networks or Convolutional Neural Networks, allowed image classification tasks, detection of Regions of Interest or semantic segmentation to be performed [56, 305–307], reaching higher performance than traditional approaches [308]. The architecture of DL classifiers avoids the need to design of procedures for extracting hand-crafted features, as the classifier structure itself computes the most characteristic features automatically for each specific dataset. These peculiarities let DL approaches be investigated in different fields, including medical imaging, signal processing or gene expression analysis [309–313].

Recent studies suggested that MR images should be preferred to other imaging techniques for assessing the risk of ADPKD [314]. However, several research works allowed estimating TKV starting from CT images thanks to the higher availability of the acquisition device, the

more accurate and reliable measurement of TKV and cysts volume. On the other side, CT images are always contrast-enhanced using a contrast medium that is harmful for the patient under examination. On these premises, it is necessary to work on the segmentation of images from MR acquisition for improving the TKV estimation capabilities.

Starting from a preliminary work performed on a small set of patients [164], in the following paragraph two different approaches based on DL architectures to perform the automatic segmentation of polycystic kidneys starting from MR acquisitions are presented. Subsequently, a fully-automated pipeline for the detection and segmentation of areas containing kidney in MR images is presented, including a preliminary step for filtering images not containing the kidney.

ADPKD Segmentation: Semantic Segmentation and R-CNN Approaches

In the first section of this paragraph, several Convolutional Neural Networks, designed and evaluated for classifying the images pixelwise are presented according to the workflow proposed in Fig. 3.13. Subsequently, we describe the object detection approach using the Regions with CNN (R-CNN) technique [315] to automatically detect ROIs containing parts of the kidneys, with the aim to subsequently perform a semantic segmentation on the extracted regions. The latter approach is described by the workflow proposed in Fig. 3.14.

Analyses were conducted on MR images acquired from February to July 2017, on 18 patients affected by ADPKD (mean age 31.25 $\pm$ 15.52 years) which underwent Magnetic Resonance examinations for assessing the TKV. The acquisition protocol was carried out by the physicians from the Department of Emergency and Organ Transplantations (DETO) of Bari University Hospital.

MR examinations and images acquisition were performed on a 1.5 TMR device (Achieva, Philips Medical Systems, Best, The Netherlands) by using a four-channel breast coil. The protocol did not use contrast material intravenous injection and consisted of:

- Transverse and coronal short T1 inversion recovery (STIR) turbo-spin-echo (TSE) sequences (TR/TE/TI = 3.800/60/165 *ms*, field of view (FOV) = 250 x 450 *mm* (AP x RL), matrix 168 x 300, 50 slices with 3-*mm* slice thickness and without gaps, 3 averages, turbo factor 23, resulting in a voxel size of 1.5 x 1.5 x 3.0 mm^3; sequence duration of 4.03 minutes);

- Transverse and coronal T2-weighted TSE (TR/TE = 6.300/130 ms, FOV = 250 x 450 *mm* (AP x RL), matrix 336 x 600, 50 slices with 3-*mm* slice thickness and without

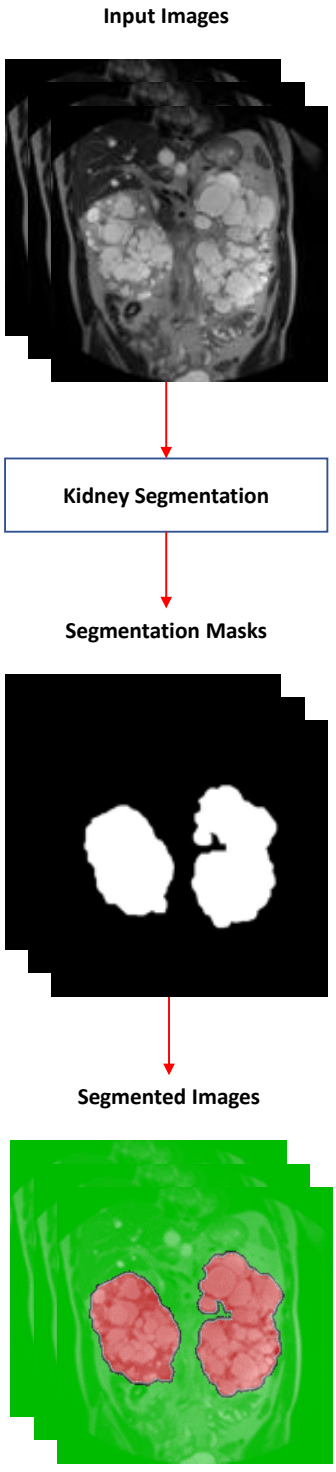


Fig. 3.13 Workflow for the semantic segmentation starting from the full image.

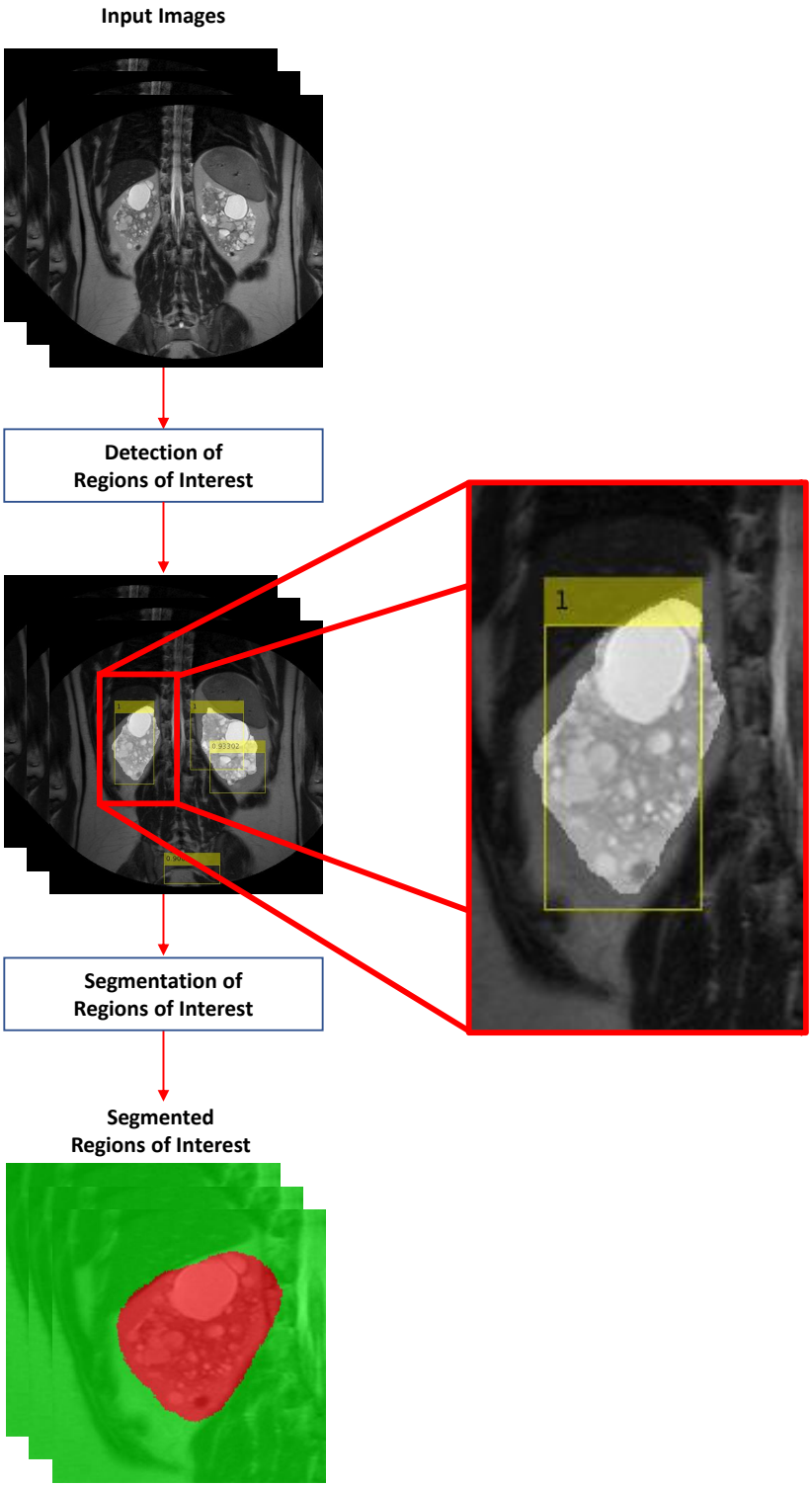


Fig. 3.14 Workflow for the semantic segmentation of ROIs automatically detected with R-CNN.

gaps, 3 averages, turbo factor 59, SENSE factor 1.7, resulting in a voxel size of 0.75 x 0.75 x 3.0 mm^3; sequence duration of 3.09 minutes);

- Three-dimensional (3D) T1-weighted high resolution isotropic volume (THRIVE) sequence (TR/TE = 4.4/2.0 ms, FOV = 250 x 450 x 150 *mm* (AP x RL x FH), matrix 168 x 300, 100 slices with 1.5 *mm* slice thickness, turbo factor 50, SENSE factor 1.6, data acquisition time of 1 min 30 s).

Although patients performed the complete protocol, only the coronal T2-weighted TSE sequence was considered for the processing and classification strategies. In order to have the segmentation ground truth for all the acquired images, a dedicated framework was developed allowing the radiologists to manually draw contours of all the ROIs using a digital tool specifically designed and implemented.

Two different approaches based on DL techniques were investigated to perform a fully-automated segmentation of polycystic kidneys without needing to design any procedure for the extraction of hand-crafted features. In details, the first approach performed the semantic segmentation of the MR images, classifying each pixel as belonging to the kidney or not; the second methodology, instead, performed the detection of reduced areas containing the kidneys before their semantic segmentation.

According to different architectures designed in previous works, such as SegNet [316] and Fully Convolutional Network (FCN) [317], the CNNs performing semantic segmentation tasks featured an encoder-decoder design (as represented in Fig. 3.15). Usually, this kind of architecture includes several encoders interspersed with pooling layers for downsampling the input; each encoder includes sequences of Convolutional layers, Normalisation layers and Linear layers. Based on the encoding part, there are specular decoders with up-sampling layers for reconstructing the input size. Finally, there are fully-connected neural units before the final classification layer able to label each pixel of the input image.

Regarding the first approach, several architectures based on CNNs were designed and tested, varying the number of encoders (and decoders), the number of layers for each encoder, the number of convolutional filters for each layer and the learner used during the training (i.e., SGDM - stochastic gradient descent with momentum, or ADAM [318]). All the investigated architectures included convolutional layers with kernels [3x3], stride [1x1] and padding [1 1 1 1] allowing to keep unchanged the dimensions of the input across each encoder; downsampling (and upsampling) was performed only in the max-pooling layers (upsampling layers for the decoder) having stride [2x2] and dimension [2x2]. To classify each pixel of the

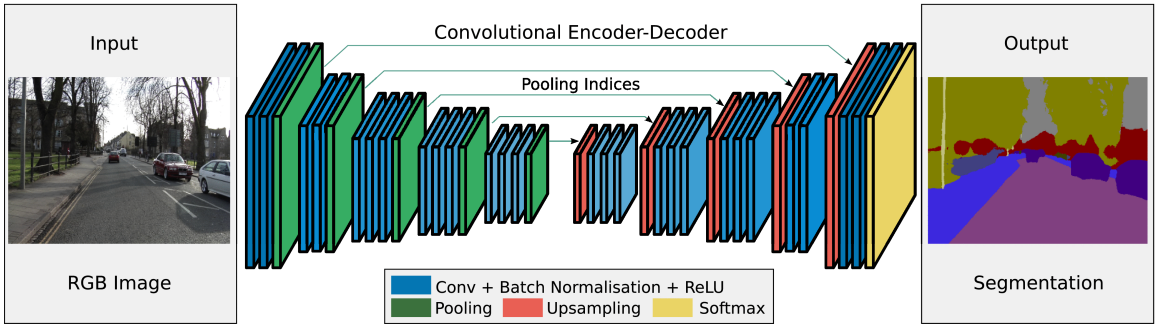


Fig. 3.15 Encoder–Decoder architecture for SegNet.

Table 3.8 Configurations designed and tested for the semantic segmentation of the full image. Each layer is a sequence of a convolutional layer, a batch normalization layer and a ReLu layer.

Network ID	**Number of layers per encoder**	**Number of convolutional filters per layer**	**Learner**
VGG-16	[2 2 3 3 3]	[64 128 256 512 512]	ADAM
S-CNN-1	[3 2 3 3 3]	[64 128 256 512 512]	ADAM
S-CNN-2	[3 2 3 3 3]	[96 128 256 512 512]	ADAM

images, the semantic segmentation was performed considering two classes: "Kidney" and "Background".

For training the classifiers, augmentation of the dataset was also performed according to recent works demonstrating the effectiveness of this procedure for improving the classification performance [319–321]. The following image transformations were considered for data augmentation, randomly generated:

- horizontal shift in the range [-200; 200] pixels;
- horizontal flip;
- scaling with scale factor ranging in [0.5; 4].

Table 3.8 shows the configurations designed and tested for this task, reporting the number of layers per encoder, the number of convolutional filters per layer and the learner employed. The table reports only the three configurations that reached the highest performance among all the investigated architectures.

Due to the presence of cysts in the organs near the kidneys and of very similar structures near the area of interest, which may affect segmentation, a second approach based on the object detection strategy using R-CNN was investigated. In this way, the classifier was able

Table 3.9 Configurations designed and tested for the CNN in the ROI detector. Each layer is a sequence of a convolutional layer, and a ReLu layer.

Network ID	**Number of layers per encoder**	**Number of convolutional filters per layer**	**Learner**
R-CNN-1	[3 3]	[32 32]	SGDM
R-CNN-2	[1 1]	[16 32]	SGDM
R-CNN-3	[3 3]	[64 32]	SGDM

to detect smaller regions inside each image which were subsequently segmented according to the approach described in the previous part of the paragraph.

Object detection is a technique for finding instances of specific classes in images or videos. Like semantic segmentation, object detection is also a well-established process in literature employed in different fields [322, 323]. The CNNs for object detection include a region proposal algorithm, often based on EdgeBoxes or Selective Search [324, 325], as a pre-processing step before running the classification. In literature, traditional R-CNN and Fast R-CNN are the most frequently used techniques [315, 326]. Recently, Faster R-CNN was also introduced, addressing the region proposal mechanism using the CNN itself, thus making the region proposal a part of the CNN training and prediction steps [327].

For tuning the optimal CNN topology also for this task, several networks for detecting areas containing the kidney were investigated, considering the Fast R-CNN approach. The manual contour of each kidney was enclosed in a rectangular bounding box and used for training the network. In this case, the considered CNNs had only the encoding part, where each encoder included the concatenation of Convolutional layers and ReLu layers. Each encoder ended with a max-pooling layer to perform image sub-sampling (size [3x3] and stride [2x2]). At the end of the encoding part, two fully-connected layers were inserted before the final classification layer. Table 3.9 reports the configurations designed and tested for detection (in this case, too, the table reports only the three configurations showing the highest performance).

After the automatic detection of ROIs, the same architectures designed for the segmentation of the whole images, reported in Table 3.8, were considered for performing the semantic segmentation of the ROIs. Since the detected ROIs could have different sizes, a rescaling procedure was performed to adapt all the ROIs to the size required by the CNN for the segmentation task. Image augmentation was performed, as well, considering the image transformations as follows:

- horizontal shift in the range [-25; 25] pixels;

- vertical shift in the range [-25; 25] pixels;
- horizontal flip;
- scaling with scale factor ranging in [0.5; 1.1].

In the end, a fully-automated pipeline for the detection and segmentation of areas containing kidney in MR images from subjects affected by ADPKD has been designed and developed, including a former classification step for the detection of images containing the kidney and a subsequent segmentation phase identifying the areas containing the pixels belonging to the kidney (Fig. 3.16). Specifically, since the detection of slices containing kidneys were shown to be crucial for improving the segmentation performance in the subsequent phase, the design of the CNN architecture to perform the classification task has been optimised through a mono-objective Genetic Algorithm finding the optimal subset of parameters for defining the different layers of the CNN classifier. Based on the architecture obtained for the classification task, a CNN for the segmentation step has been designed by properly creating the respective decoding part of the network for reconstructing the input image.

A two-steps classification strategy made it possible to obtain the final segmentation of images representing kidneys affected by ADPKD. Both steps relied on deep Convolutional Neural Networks. In details, the first classification step was concerned with the detection of the MR slices containing parts of the kidney; the second step, instead, dealt with the segmentation of the images showing the kidney. The following paragraphs report a detailed description of the designed classification approaches.

The first step of the automated pipeline included a CNN for detecting the MR slices representing the kidney. Since the design of a classifier architecture may affect its performance, there are several strategies based on optimisation algorithms allowing one to numerically search for the optimal set of parameters for its design, configuration and tuning [108, 119]. A single-objective genetic algorithm (GA) allowed us to design an optimised topology of the CNN for the classification task [56]. Each genotype described convolutional architectures composed of at least one encoder, up to three; each encoder included up to three groups of the following operators: Bi-dimensional Convolution Layer (conv2d layer), Batch Normalisation Layer, Rectified Linear Unit (ReLU) layer. The chromosome also modelled the number of kernels and the filter size for each convolutional layer. Finally, up to two fully connected layers (with a maximum of 512 neurons per layer) were allowed to precede the softmax layer for the final classification between the two classes. The training algorithm for all the CNNs was ADAM [318], showing more limited memory usage with respect to other

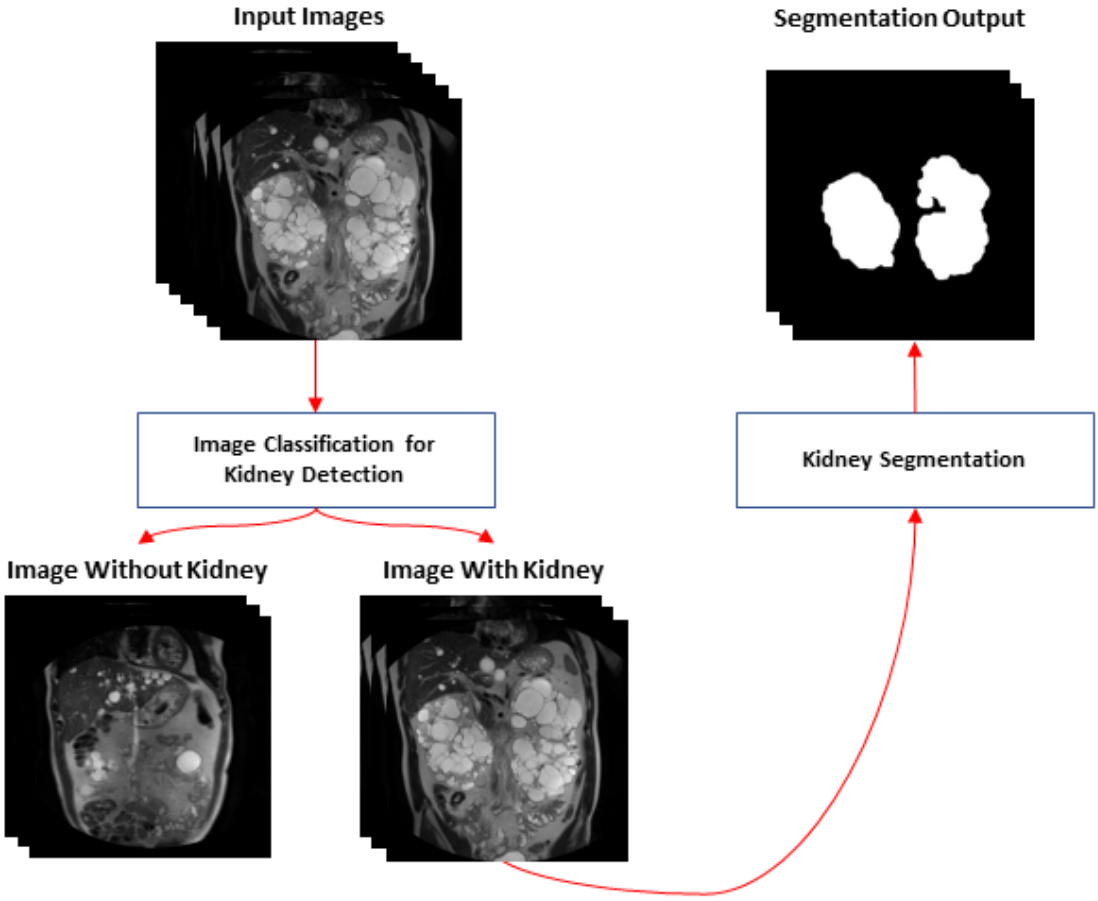


Fig. 3.16 Representation of the implemented automatic pipeline for the segmentation of images containing the kidney

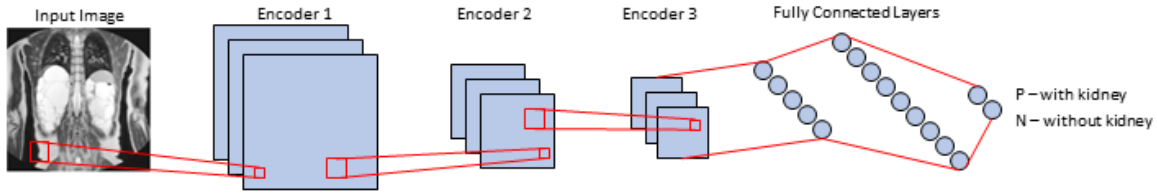


Fig. 3.17 The representation of a Convolutional Neural Network candidate solution in the Genetic Algorithm. Each encoder can have up to 3 sequences of bi-dimensional convolutional layers, batch normalisation layers and Rectified Linear Unit (ReLU) layers. Between two consecutive encoders, there is a max pooling layer with kernel size [4 4].

algorithms. In addition, a max pooling layer (filter size [4x4], stride [4x4]) was included to perform subsampling from one encoder to the subsequent. The training procedure also included dataset augmentation for improving the overall classification performance and the generalisation capabilities of the classifier [308]. The evolutionary optimisation started from an initial random population of 100 individuals. The crossover probability was 0.8 whereas the mutation operator was an adaptive feasible mutation function randomly generating directions adaptive with respect to the last successful or unsuccessful generation (the mutation chooses a direction and step length that satisfies bounds and linear constraints). Fig. 3.17 shows a complete representation of a candidate solution of the GA; Table 3.10 reports the representation of the CNN parameters included in the GA chromosome.

Each individual within the search space was trained, validated and tested on a random permutation of the dataset dividing it into a training set, a validation set and a test set with percentages of 60-20-20, respectively. Each input image was then classified and labelled as Positive (P) if it contained at least one pixel belonging to the kidney, Negative (N) otherwise. The fitness function for evaluating the genotypes was the classification Accuracy (Eq. 2.1) computed considering the performance on the test set. For designing the CNN for segmentation purposes, the optimised CNN topology obtained in the previous step was replicated in the encoding section of the semantic classifier. The decoding part was also generated from it. All images containing at least one pixel of the kidney class constituted the dataset used for the semantic segmentation. Specifically, each image was split into a left and right part, increasing both the number and variety of the sample size. As for image classification, the input dataset was randomly divided into a training set, a validation set and a test set with percentages of 60-20-20, respectively.

Concerning the results obtained in the investigated approaches, several metrics were considered for evaluating the classifiers, including those reported in Section 2.4.4.

Table 3.10 Representation of the parameters optimised by the GA for designing the CNN performing image classification.

Network Configuration			
Layers	***GA Parameters***		
	Number of Layers	*Number of Filters*	*Filter Dimension*
Image Input	-	-	-
Encoder 1	[1 – 3]	[8 – 256]	[1 – 7]
max-pooling	-	-	-
Encoder 2	[0 – 3]	[8 – 256]	[1 – 7]
max-pooling	-	-	-
Encoder 3	[0 – 3]	[8 – 256]	[1 – 7]
max-pooling	-	-	-
	Number of Neurons		
Fully-Connected 1	[0 – 512]		
Fully-Connected 2	[0 – 512]		
Softmax	2		

For each R-CNN architecture reported in Table 3.9, we show the Precision-Recall plot, showing the Precision obtained at different Recall values, and the logAverageMissRate plot, reporting how the miss rate varies at different levels of *FP* per image. Specifically, Fig. 3.19, Fig. 3.20 and Fig. 3.21 show the plots for R-CNN-1, R-CNN-2 and R-CNN-3 respectively. Fig. 3.18 shows the result obtained considering a sample image.

As represented in the plots, R-CNN-1 and R-CNN-3 showed an average precision higher than 0.75, maintaining low the log-average miss rate. Since the aim of detecting ROIs was the identification of regions for the semantic segmentation step, R-CNN-1 was the best candidate among the analysed architectures. In fact, it reached a Recall value of about 0.8 with Precision higher than 0.65, meaning that the classifier was able to detect 80 % of the Positive ROIs, but with a high number of False Positives. However, this problem was overcome in the subsequent step by the semantic segmentation of the detected ROIs.

Concerning the semantic segmentation, Table 3.11 shows the results for each CNN architecture performing semantic segmentation of the MR image. As reported in the table, the architecture performing better for the semantic segmentation of the full image is S-CNN-1, reaching an accuracy above 88 %, higher than the other investigated architectures.

The introduction of an additional layer into the first encoder of VGG-16 allowed the net to create a more significant set of features with respect to a simpler architecture, thus leading to a more accurate classification of the pixels. Conversely, increasing the number

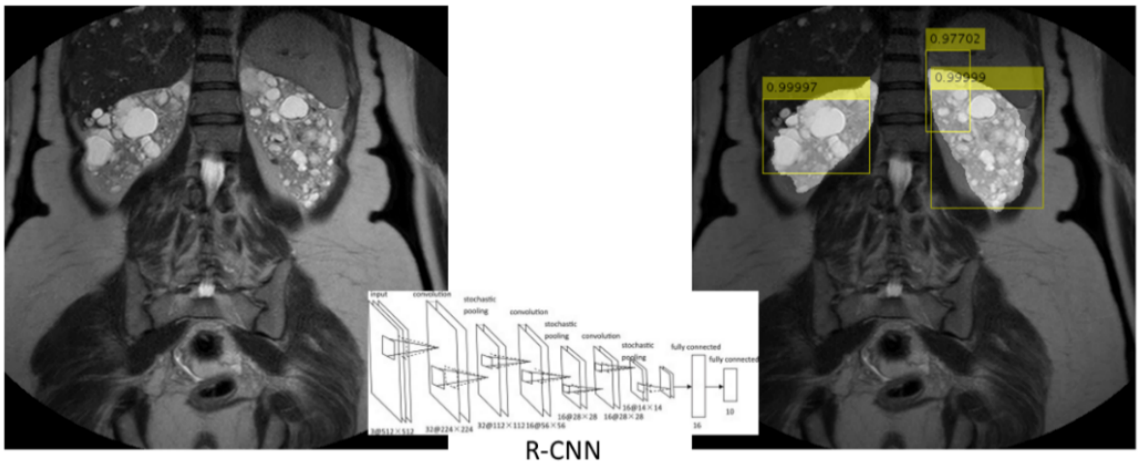


Fig. 3.18 Results from R-CNN classifier. Input image is on the left; the image on the right contains squares on the detected ROIs, each one is associated with a score.

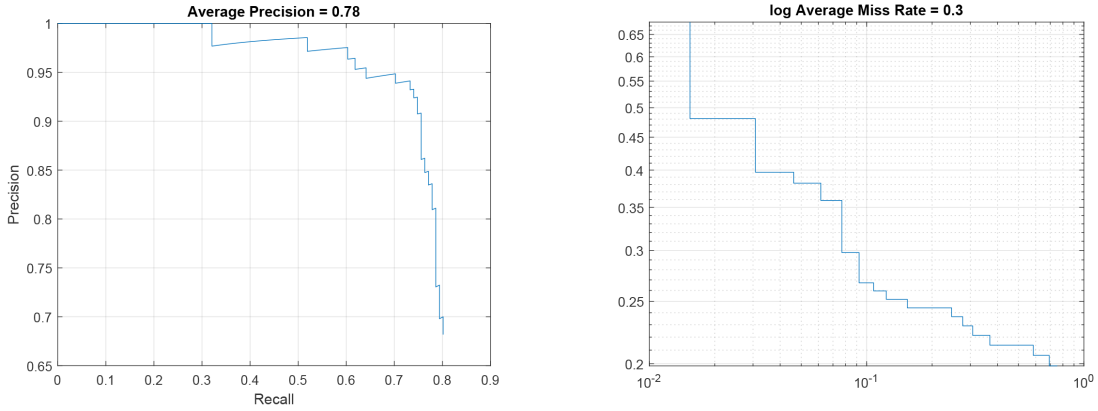


Fig. 3.19 Precision – Recall plot and log Average Miss rate for R-CNN-1.

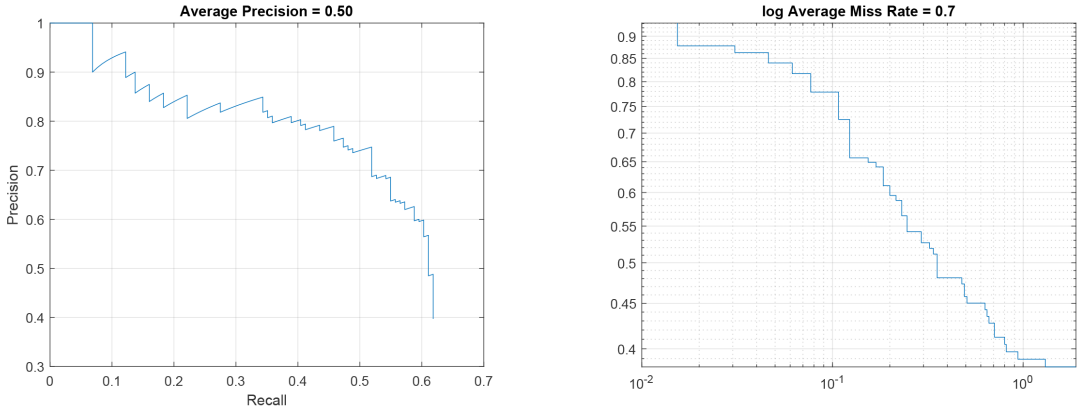


Fig. 3.20 Precision – Recall plot and log Average Miss rate for R-CNN-2.

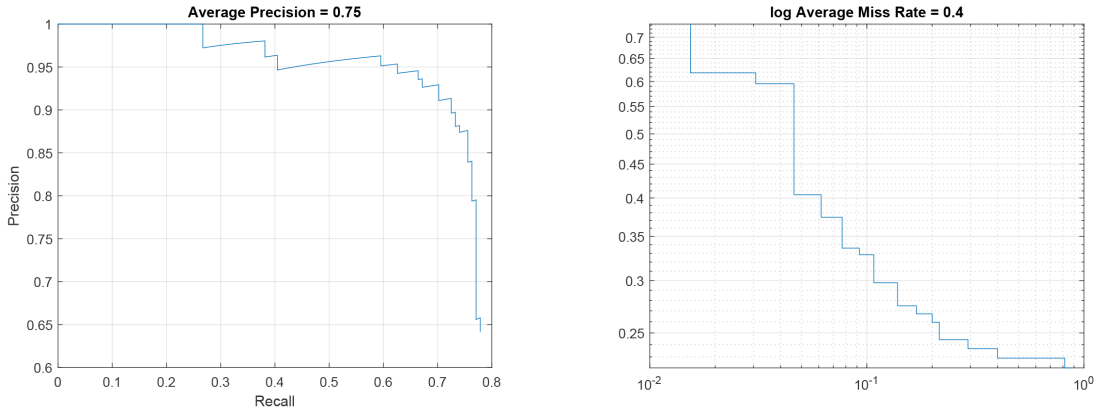


Fig. 3.21 Precision – Recall plot and log Average Miss rate for R-CNN-3.

Table 3.11 Performance indices for the classifiers working on MR images.

Network ID	**Mean Accuracy**	**Weighted IoU**	**Mean BF Score**
VGG-16	0.88076	0.75288	0.41117
S-CNN-1	0.88359	0.76294	0.38205
S-CNN-2	0.79824	0.52781	0.38643

of convolutional filters in the first layer of the first encoder did not improve the overall performance. Table 3.12 reports the normalised confusion matrices obtained for the three cases, whereas Fig. 3.22 shows an example of the semantic segmentation output.

As for the segmentation of the MR images, Table 3.13 reports the performance indices for the semantic segmentation of the ROIs automatically detected by R-CNN-1, which showed an optimal trade-off in detecting ROIs. As in the previous case, the best performance on the semantic segmentation of the ROIs was obtained by S-CNN-1. The normalised confusion matrices for all classifiers are reported in Table 3.14. Fig. 3.23 shows the result obtained for the semantic segmentation of ROI extracted from an image sample.

Finally, concerning the automatic pipeline, the optimised individual found by the GA showed a topology made up of three encoders: the first one included two convolutional

Table 3.12 Normalized Confusion Matrix for VGG-16, S-CNN-1 and S-CNN-2 segmenting the MR images for the test set.

		VGG-16		**S-CNN-1**		**S-CNN-2**	
		True Condition		**True Condition**		**True Condition**	
		Positive	***Negative***	***Positive***	***Negative***	***Positive***	***Negative***
Predicted Condition	***Positive***	*0.96629*	*0.20477*	*0.96146*	*0.19428*	*0.96595*	*0.21611*
	Negative	*0.03371*	*0.79523*	*0.03854*	*0.80572*	*0.03405*	*0.78389*

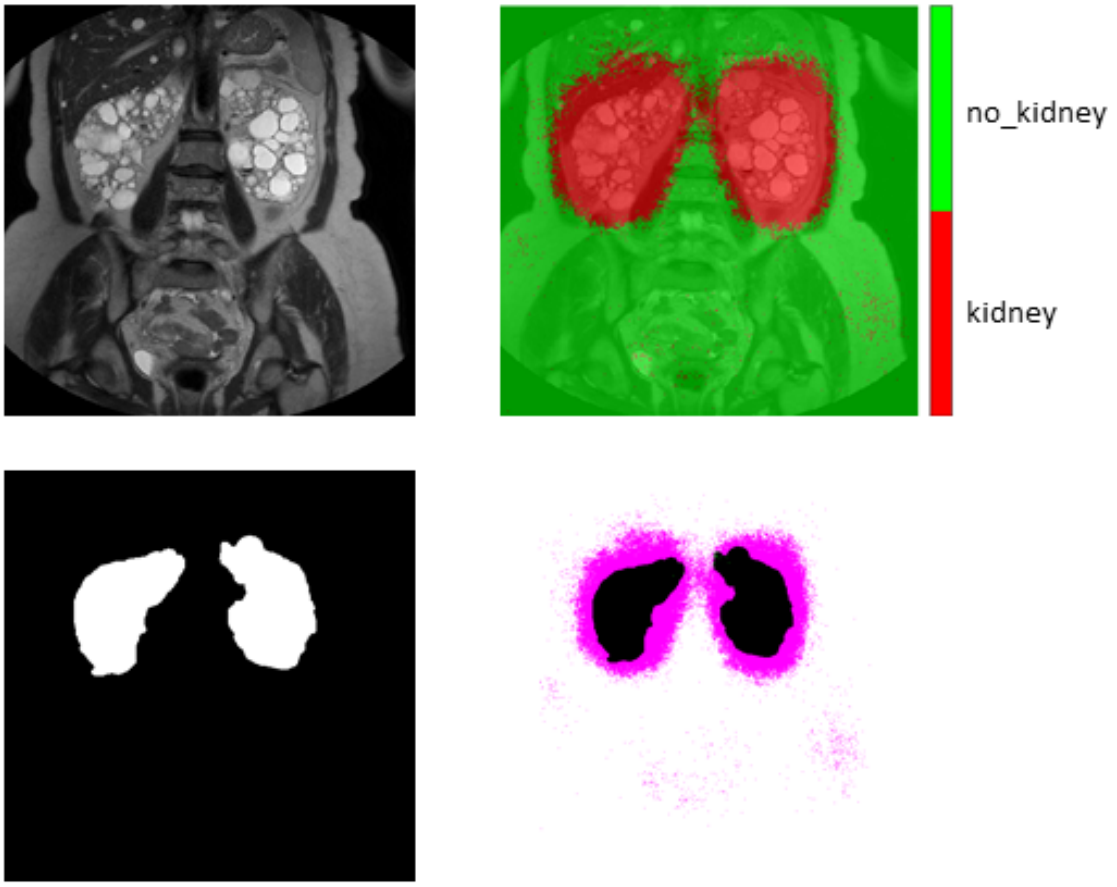


Fig. 3.22 Result of the semantic segmentation considering a sample image. Top left: the MR slice represented in greyscale; top right: the segmentation result; bottom left: the ground-truth mask; bottom right: superimposition of the segmentation result onto the ground-truth mask.

Table 3.13 Performance indices of the classifiers working on the ROIs.

Network ID	**Mean Accuracy**	**Weighted IoU**	**Mean BF Score**
VGG-16	0.86016	0.75426	0.34828
S-CNN-1	0.8726	0.8540	0.4332
S-CNN-2	0.8550	0.82931	0.41515

Table 3.14 Normalised Confusion Matrix for VGG-16, S-CNN-1 and S-CNN-2 segmenting the ROIs detected by R-CNN-1 from the MR images of the test set.

		VGG-16		**S-CNN-1**		**S-CNN-2**	
		True Condition		**True Condition**		**True Condition**	
		Positive	***Negative***	***Positive***	***Negative***	***Positive***	***Negative***
Predicted Condition	***Positive***	*0.88781*	*0.16749*	*0.79742*	*0.05213*	*0.77762*	*0.06762*
	Negative	*0.11219*	*0.83251*	*0.20258*	*0.94787*	*0.22238*	*0.93238*

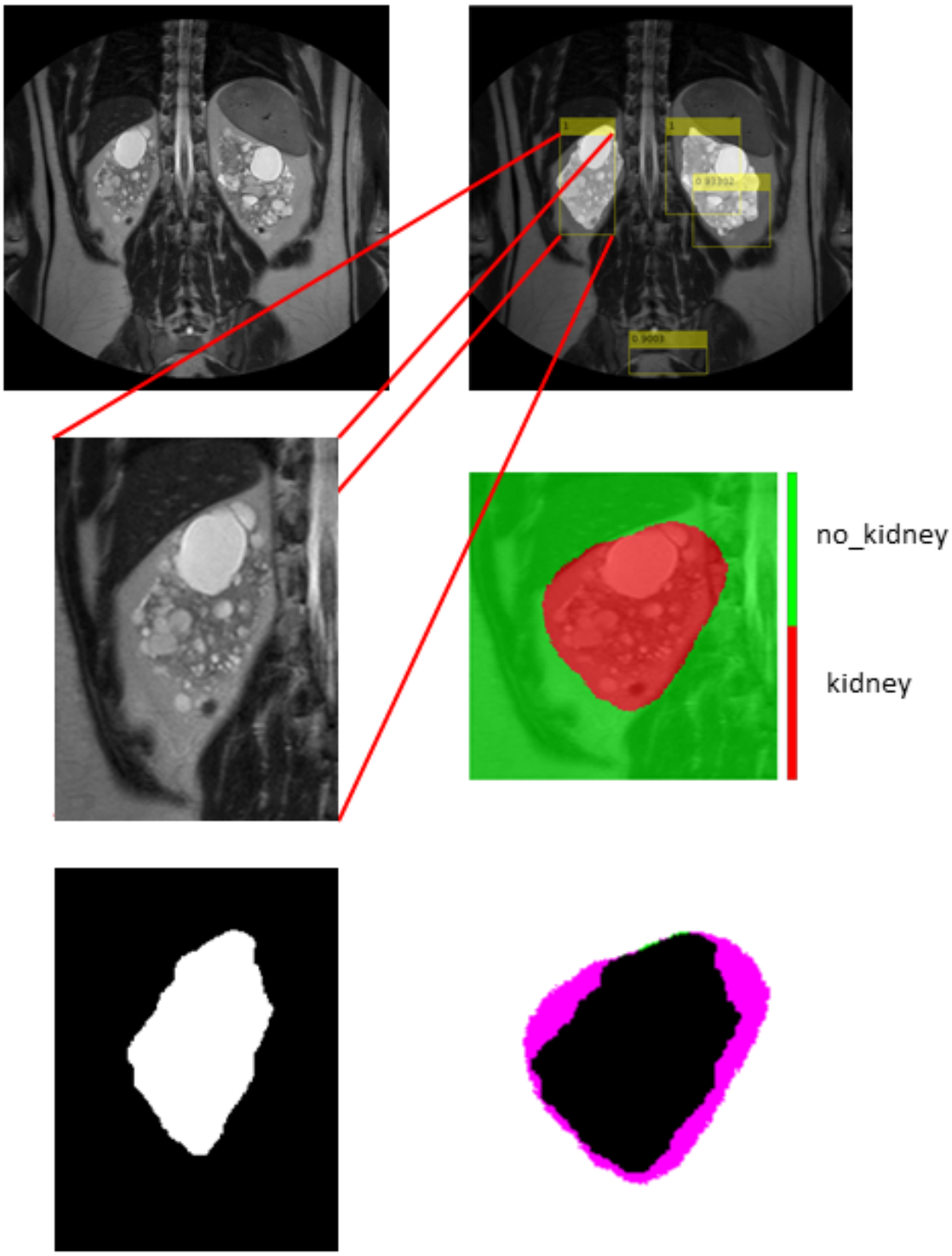


Fig. 3.23 Example result for ROI detection and semantic segmentation. Top left: the MR slice represented in greyscale; top right: the R-CNN detection result; middle left: one of the detected ROIs; middle right the segmentation result; bottom left: the ground-truth mask for the considered ROI; bottom right: superimposition of the classification result onto the ground-truth mask.

Table 3.15 Confusion Matrix computed on the test set for the best individual found by the genetic algorithm.

		True Condition	
		Positive	***Negative***
Predicted Condition	***Positive***	*69*	*2*
	Negative	*3*	*29*

Table 3.16 Confusion Matrix computed on the test set for the best individual found by the genetic algorithm.

		True Condition	
		Positive	***Negative***
Predicted Condition	***Positive***	*0.9441*	*0.1229*
	Negative	*0.0559*	*0.8771*

layers (11 kernels with size [2x2]); the second encoder included two convolutional layers (182 kernels with size [5x5]); the last encoder included one convolutional layer (32 kernels with size [5x5]). The two fully-connected layers included 56 and 232 neurons, respectively. Table 3.15 reports the confusion matrix of the classification results on the test set for the best genotype. As shown, 95.15 % Accuracy is obtained, whereas Sensitivity and Specificity reach 95.83 % and 93.55 %, respectively. Figure 3.24 shows the results obtained for some image samples.

The segmentation task of the pipeline was evaluated considering different metrics. Table 3.16 reports the normalised confusion matrix of the classification performance for the segmentation task. Moreover, the overall performance shows Accuracy reaching 91.06 %, whereas IoU and BF Score reach 0.8296 and 0.5234, respectively.

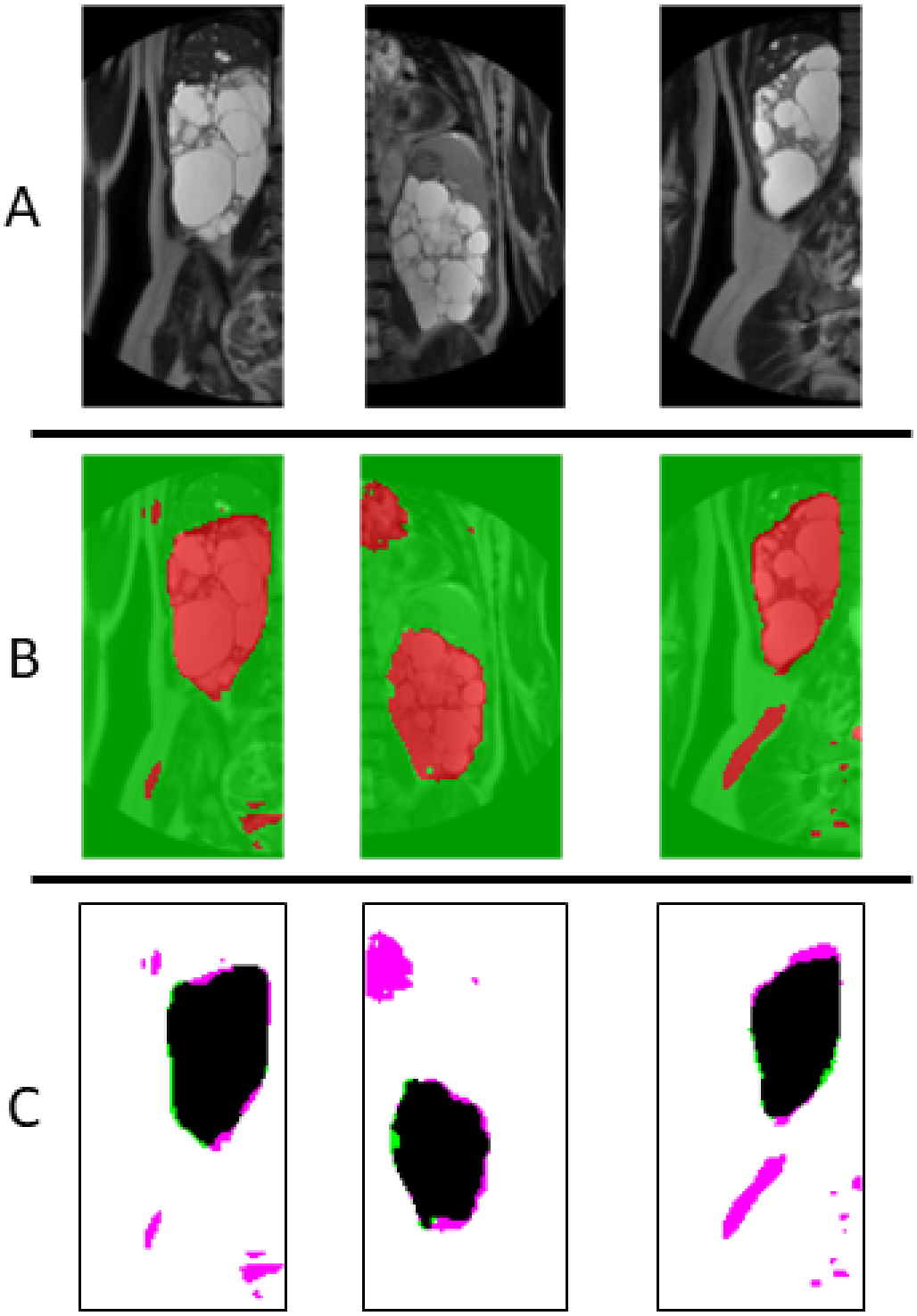


Fig. 3.24 The output of the kidney segmentation by a CNN. (A) Some input images. (B) Superimposition of the output of the segmentation onto the input image: the red pixels are "kidney", whereas the others belong to the background. (C) Superimposition of the segmentation output (purple) onto the ground truth obtained by manual segmentation of the kidney (green); the black pixels are the True Positives.

3.3 Pathology

Concerning the area of pathology, several works have been conducted in order to support the clinical diagnosis. In this section, the analysis of histological images for assessing the suitability for transplant of a kidney, and segmentation and classification of haematological images for supporting white cell counting are described.

3.3.1 Histological Images Analysis

Histological evaluation is a clinical practice of many organ transplant centres when donor age is higher than 60 years or when there are comorbidities affecting the patient, such as diabetes or hypertension, that could have compromised the renal functionality. In these cases, histological analysis of renal biopic tissue is useful to assess whether kidneys are suitable for single or double transplant, or rather have to be discarded. The Karpinski score is an histological evaluation that assesses the degree of chronic lesions in each compartment of the renal parenchyma, by analysing glomeruli, tubules, interstitium and vessels [328]. A semi-quantitative score, from 0 to 3, is assigned to the degree of glomerulosclerosis, tubular atrophy, interstitial fibrosis and vascular sclerosis with narrowing of vascular lumina. High values of these parameters are associated with poor, early and late graft outcomes [329].

3.3.1.1 Research Contribution

Starting from the work in [230], where vessels were segmented by using a procedure for detecting the lumen position and a metric to remove False Positives (FPs), the research work here reported improves on the proposed approach proposing a new procedure to evaluate the relevance of the presence of nuclei in the membrane, making it possible to further reduce the number of FPs. The procedure for classifying blood vessels versus tubules follows the pipeline reported in Fig. 3.25.

Kidney Biopsy Slide (KBS) preparation and digital acquisition have been conducted at the Department of Emergency and Organ Transplantation (DETO), University of Bari Aldo Moro (Bari, Italy). All KBSs have been prepared by expert lab technicians of DETO according to the Renal Biopsy Guidelines of the Ad Hoc Committee appointed by the Renal Pathology Society [330]. Following these guidelines, once it is ready, the KBS can be analysed at the microscope by a nephropathologist and digitally acquired. In details, each KBS was digitally scanned using the Aperio ScanScope CS microscope (http://www.leicabiosystems.com/

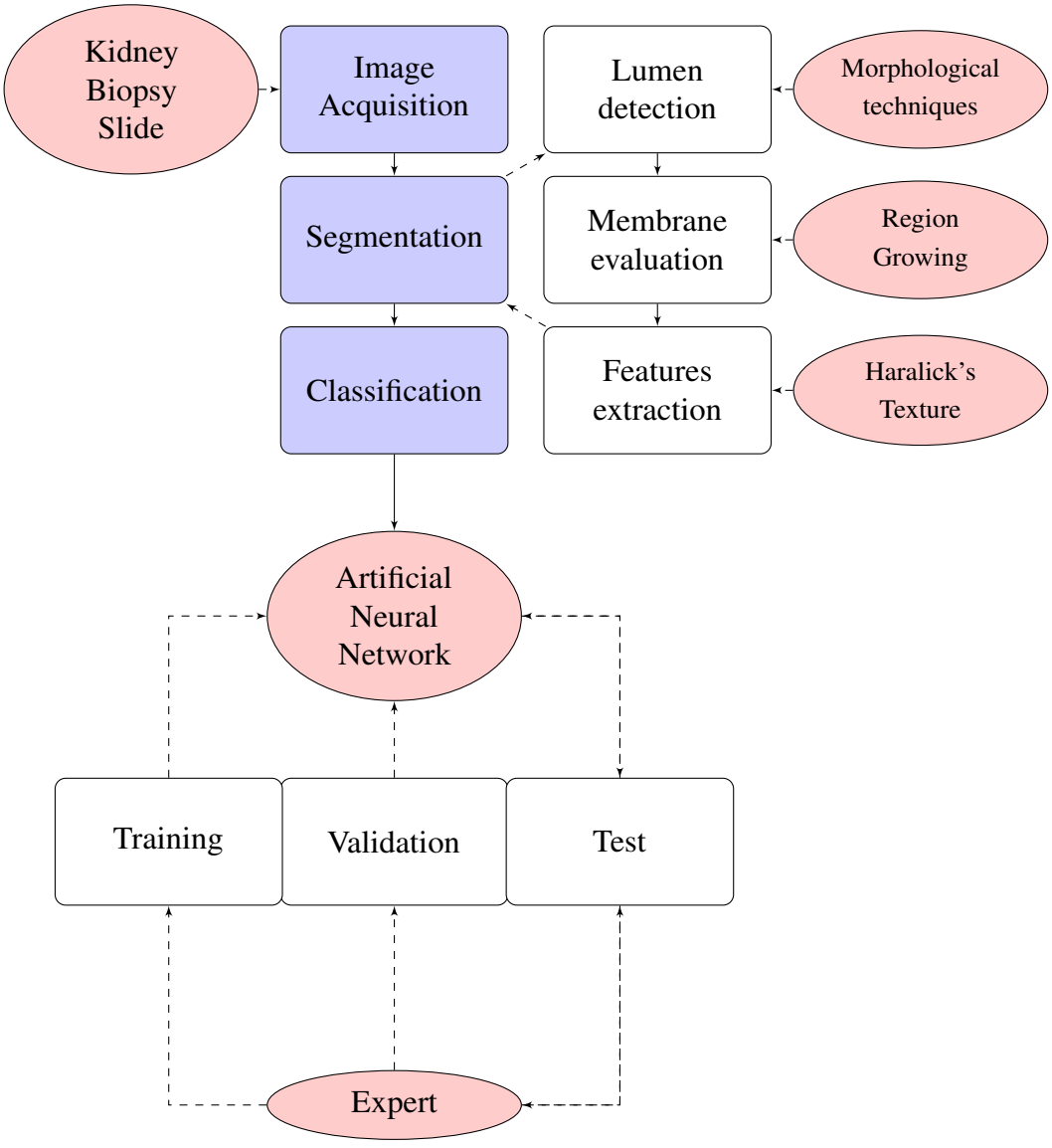


Fig. 3.25 Pipeline architecture.

digital-pathology/aperio-digital-pathology-slide-scanners/details/product/aperio-cs2/) featuring a 20x optical zoom. The acquired RGB images have the following characteristics:

- Resolution: 0.50 μm/pixel;
- Compression Standard: JPEG.

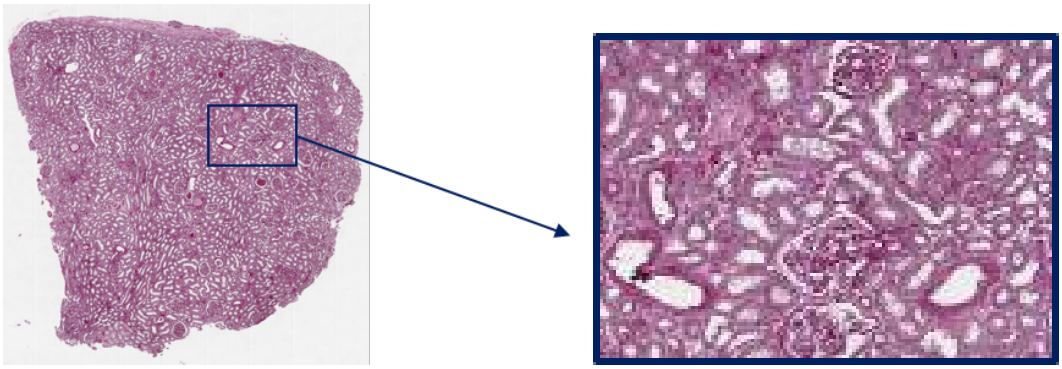
The image file size and the acquisition time depend on the amount of area selected with the ScanScope acquisition interface. All the acquisitions were performed using the automatic focus in order to keep the acquisition durations low [331]. But, at the same time, due to the automatic focus, some areas resulted out of focus. For the evaluation of vessel, the protocol included the Periodic acid–Schiff (PAS) staining slide, which is a staining method used to detect polysaccharides such as glycogen, and mucosubstances such as glycoproteins, glycolipids and mucins in tissues. The reaction of periodic acid oxidises the vicinal diols in these sugars, usually breaking up the bond between two adjacent carbons not involved in the glycosidic linkage or ring closure in the ring of the monosaccharide units that are parts of the long polysaccharides, and creating a pair of aldehydes at the two free tips of each broken monosaccharide ring. The oxidation condition has to be sufficiently regulated so as to not oxidize the aldehydes further. These aldehydes then react with the Schiff reagent to give a purple-magenta color. PAS staining of kidney tissues typically appear as in Fig. 3.26a.

The first step in the segmentation process was to detect the lumen position in the acquired image. In order to identify the position of each lumen, the RGB image was converted into a grey-scale image. Lumens were detected using a combination of a thresholding process and a morphological evaluation [332]. In particular, all the pixels with a value greater than or equal to a threshold were labelled as lumen components. A new black and white mask was created, with white pixels corresponding to lumen pixels, black otherwise. In order to obtain a final lumen mask with a number of connected components equal to the number of lumens, some morphological operations were executed. The result of these steps is shown in Fig. 3.26.

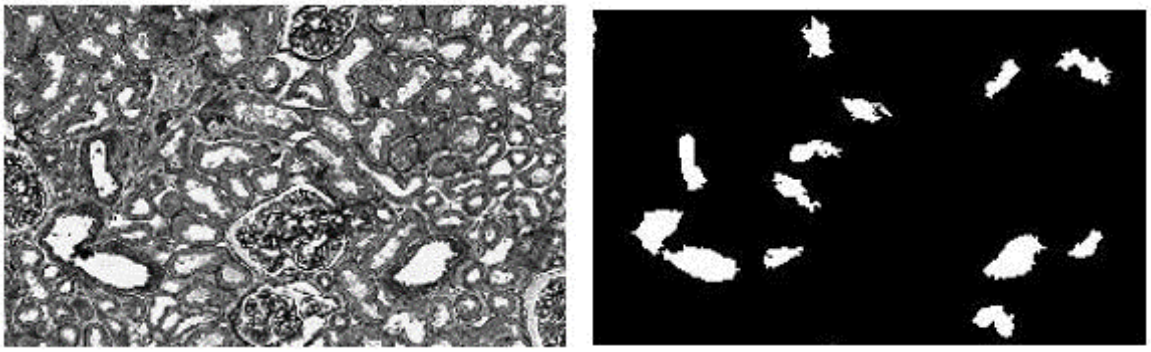
After detecting lumen positions, the mask was processed to remove FPs lumen. In general, a FPs consisted in lumen of tubules.

A metric M has been developed considering 3 factors :

1. Number of the closest connected components found by the previous procedure;
2. Distance among the closest connected component, where the distance between different connected components is the distance between their centroid;
3. Pixel Area ratios with the closest connected components.



(a) Original RGB Image. (b) Sub-ROI of RGB Image (a).



(c) Gray scale image. (d) Lumen mask obtained after the step 3.

Fig. 3.26 Steps of the procedure for the labelling of connected components.

The designed metric M was based on a specific characteristic of the kidney tissue: tubules are very close to each other while the vessels are reciprocally distant. Thus, this metric allowed to find a feature to remove tubules from the mask while preserving the near vessels defined as follows:

$$M = \frac{N_{DC} + N_{CCSA} - N_{VCCSA}}{N_{CN}},\ M \in [0,1]. \tag{3.3}$$

where:

- N_{DC}: *Distant Components Number*, is the number of components that have a distance equal to 65 pixels. The bigger the number of distant components, the more likely the membrane is thicker.
- N_{CCSA}: *Number of Close to each other Components with a Small Area*, is the number of components that have a distance in the range $[30,65[$ pixels and that have a small area compared to one of the component under evaluation (i.e., ≤ 20 %). It is most likely that tubules with small lumen are near the vessels.
- N_{VCCSA}: *Number of Very Close to each other Components with a Similar Area*, is the number of components that have a distance < 30 pixels and which have an area comparable to the component under evaluation (i.e.,≥ 60 %). This parameter is used to compensate the rare situations in which two vessels are close and have a lumen with similar area.
- N_{CN}: *Number of Components in the Neighborhood*, is the number of components in the neighborhood of the component under evaluation.

The result of the application of this metric is shown in Fig. 3.27.

Both tubules and vessels show dark nuclei on their membranes. However, the nuclei of the tubules, unlike those of the vessels, are generally placed uniformly along the centreline, between external and internal edges. For this reason, it was possible to evaluate this feature. Fig. 3.28 and 3.29 clarify the disposition of nuclei and the centreline position.

A Seeded Region Growing (SRG) procedure was developed for extracting the membrane and, for each image, a set of 20 features was computed. Since each histological image was constituted by 5 images, a 101 features pattern was used, including an index of saturation of the membrane itself.

The Vessel vs Tubule classification step was performed by means of a Back Propagation Neural Network (BPNN) [333]. The classifier was trained using all the 101 features

(a) (b)

Fig. 3.27 (a) Mask of Fig. 3.26. (b) Lumen Mask obtained after processing images according to the defined metric.

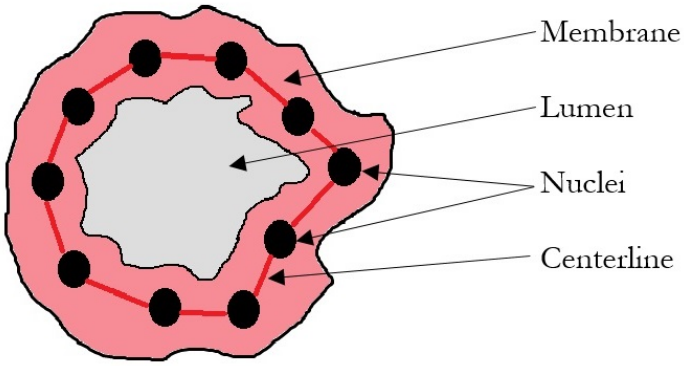


Fig. 3.28 Tubule representation. Membrane, nuclei and centerline.

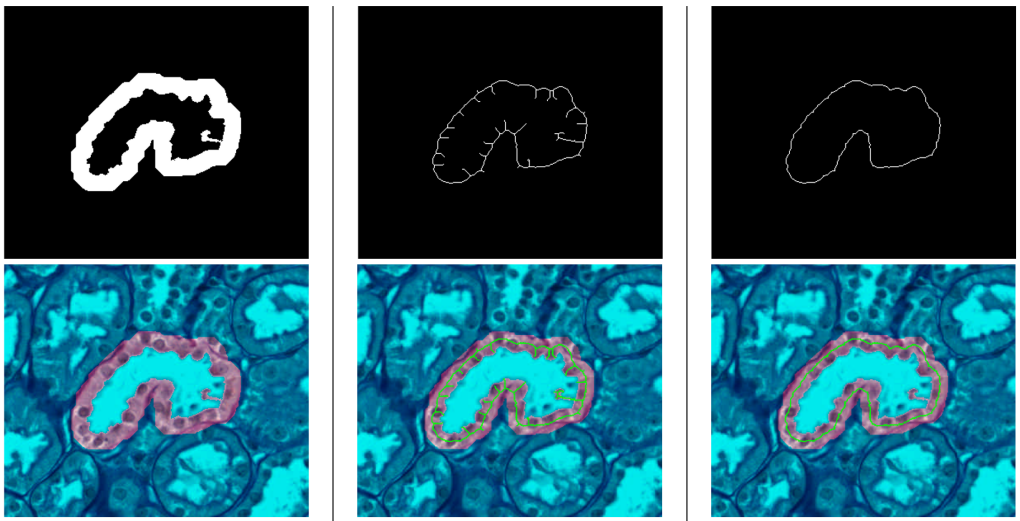


Fig. 3.29 Centerline detection: Membrane, Skeleton and Centerline.

previously extracted from a data set randomly divided in three parts: Training Set (50 %), Validation Set (30 %) and Test Set (20 %).

The classification performance of the validation set showed that the ANN was able to classify Vessels and Tubules with a Precision greater than 88 %, Accuracy greater than 93 % and Recall greater than 95 %. Analysing the performance also on the Test Set, the Neural Network assured also good performance in terms of Precision (91 %) and Recall (91 %), the classifier allowed to achieve good results. It is worth noting that the precision and recall levels have been reached having a small number of examples available for training and validation.

3.3.2 Haematological Images Analysis

The design of CAD systems for supporting clinical diagnosis of blood neoplasias is of fundamental importance since, in the most of cases, the invasiveness of the procedures for their diagnosis can be considerably reduced [334–340].

In the haematological field, the analysis of peripheral blood images is fundamental and should be performed even in case of low resolution images. In fact, thanks to deep learning approaches, several works dealing with the reconstruction of images from low-resolution acquisitions have been introduced. For example, Dong *et al.* propose a deep learning method for single image super-resolution (SR), learning an end-to-end mapping between the low/high-resolution images, represented as a deep convolutional neural network taking the low-resolution image as the input and producing the high-resolution one as output [341]. Then, in a subsequent work [342], different network structures are also investigated, together with parameter settings, to achieve an optimal trade-off between performance and speed.

An SVM classifier performing White Blood Cell (WBC) classification has been presented by Habibzadeh *et al.* [343]. In their work, the authors were particularly interested in classifying and counting five main types of white blood cells (as those in https://github.com/zxaoyou/segmentation_WBC [344]) in a clinical setting, where the quality of microscopic imagery may be poor. The proposed approach was mainly composed of three steps:

1. Image acquisition and discrimination of WBCs from RBCs (Red Blood Cells);
2. Feature extraction by Dual-Tree Complex Wavelet Transform (DT-CWT);
3. Classification by means of a Support Vector Machine (SVM) of the five WBC types.

The reported experimental results indicated that the developed analysis presents remarkable recognition accuracy even in the presence of poor quality samples and multiple classes.

3.3.2.1 Research Contribution

In the following paragraphs, a comparison between two approaches for detecting and classifying white blood cells (leukocytes) from Peripheral Blood Smears (PBS) will be analysed and discussed. In particular, the approach reported in the following paragraph is based on the extraction of hand-crafted features for a subsequent neural classifier making decisions, whereas the second approach, described in the subsequent paragraph, considered CNNs as feature extractors for subsequent classifiers. Also, a comparison between SVM classification and CNN classification considering images as inputs is shown.

Leukocytes Classification: Hand-crafted Features vs Deep Learning Approaches

Observation of PBS under the microscope is fundamental in haematology, as an analysis both of leukocyte formula and of morphological characteristics of blood cells (red blood cells, white blood cells, and platelets) provides useful information from a clinical point of view. The morphological evaluation of the WBC can help specialists diagnose haematological pathologies such as leukaemia and non-haematological ones such as infectious mononucleosis.

In [345], Bevilacqua *et al.* investigated and designed image processing techniques with low computational requirements, together with a CAD system able to recognize all five types of leukocytes (Fig. 3.30), namely Neutrophils, Lymphocytes, Monocytes, Eosinophils and Basophils, improving the work by Hiremath *et al.* [346], who considered only a subset of cell types. The implemented workflow is represented in Fig. 3.31.

The leukocytes segmentation consisted in three main steps able to detect their position, the plasma and the leukocyte edge, respectively. In details, a colour space conversion into a Hue Saturation Value (HSV) space, a thresholding operation, and a morphological dilatation allowed us to generate a mask for the detection of nuclei positions in the images under consideration (Figures3.32 and 3.33).

Then, a thresholding operation considering the grey-scale histogram of each window containing a leukocyte could separate plasma (background) from the cell; finally, leukocytes edges were detected performing morphological operations on a mask obtained after a thresholding operation considering the blue channel of the RGB window. According to the literature [346–349], geometric, chromatic and texture-based features for blood cells

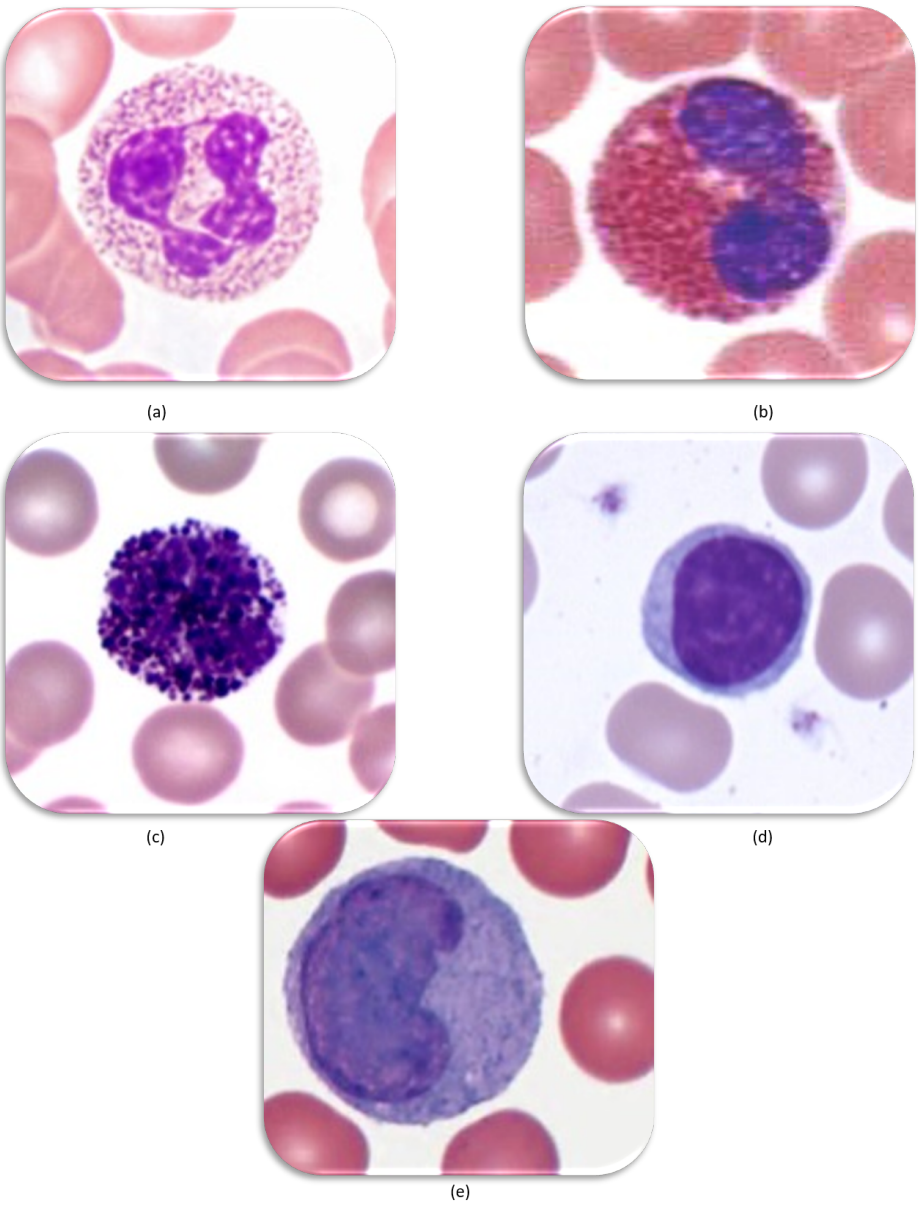


Fig. 3.30 The five types of leukocytes to be classified: (a) Neutrophils, (b) Eosinophils, (c) Basophils, (d) Lymphocytes, (e) Monocytes

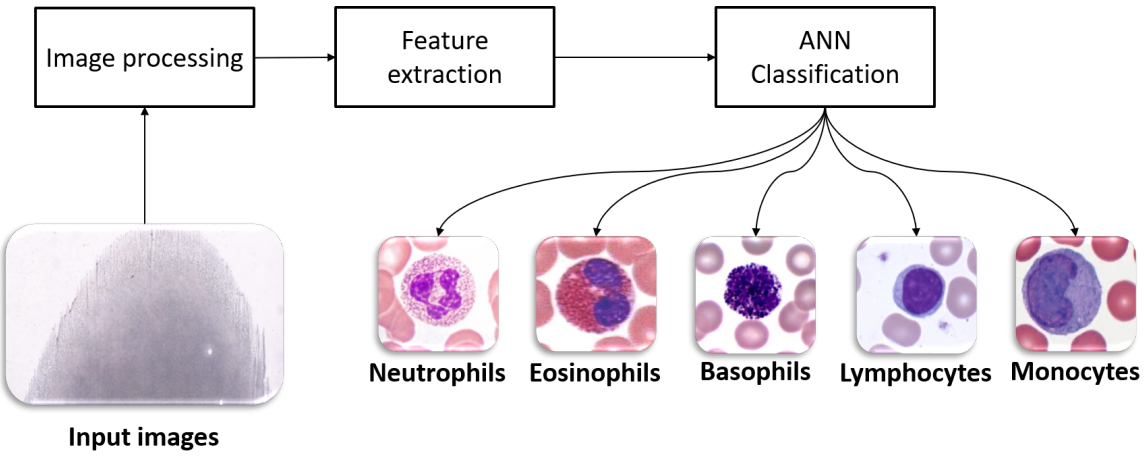


Fig. 3.31 Work flow for leukocytes classification considering hand-crafted features

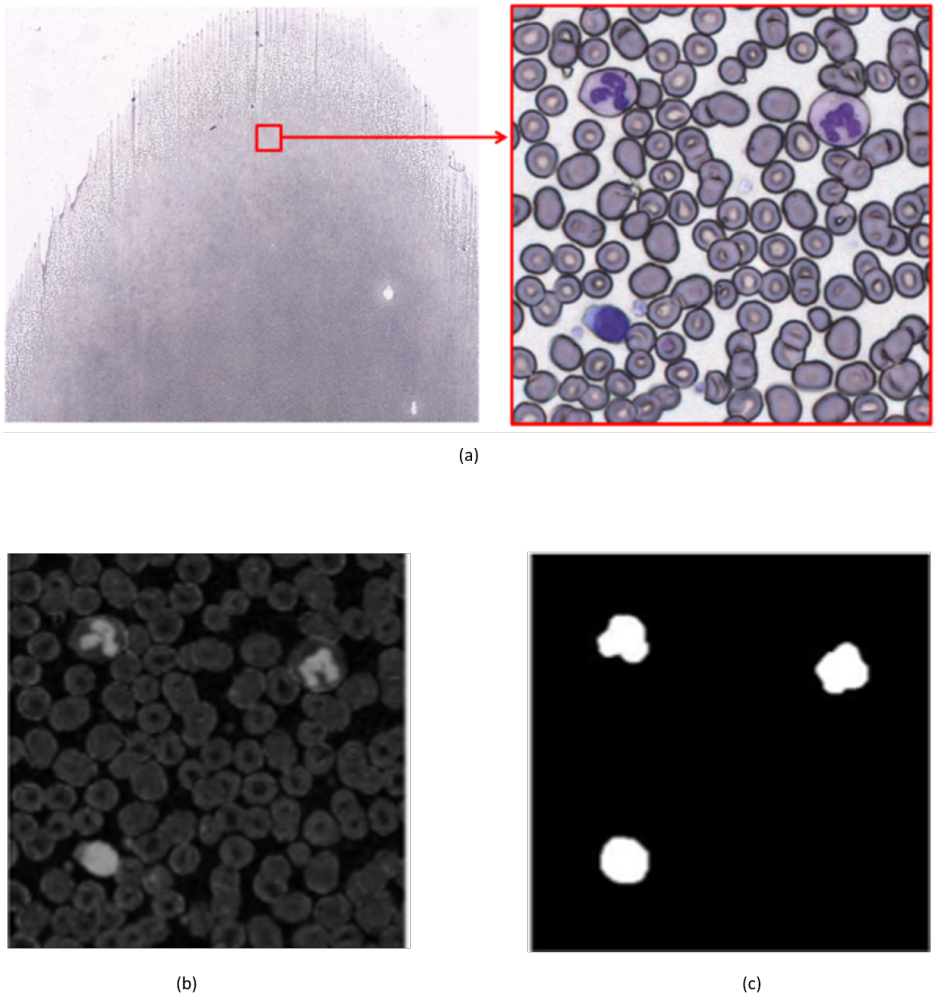


Fig. 3.32 A representation of the steps of the method used to obtain the nuclei mask; (a) Sub-image extraction. (b) S channel of HSV sub-image. (c) Obtained Nuclei mask.

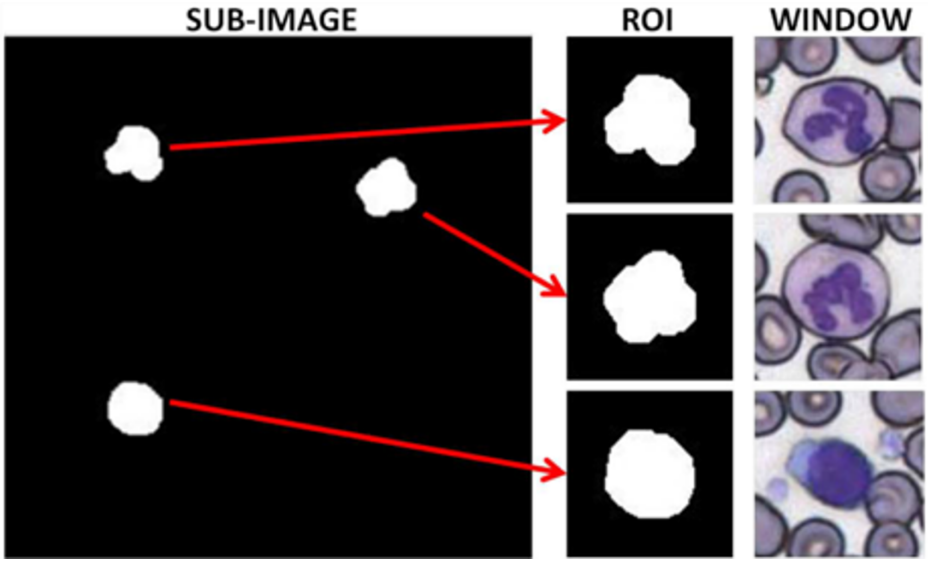


Fig. 3.33 Leukocyte's ROI and window extraction in one sub-image

Table 3.17 Accuracy, Sensitivity and Specificity of leukocyte classification using a 1-vs-all approach.

	Accuracy	**Sensitivity**	**Specificity**
Neutrophils vs all	*96.78 %*	*95.55 %*	*99.73 %*
Lymphocytes vs all	*96.78 %*	*98.81 %*	*96.06 %*
Monocytes vs all	*99.61 %*	*92.59 %*	*99.76 %*
Eosinophils vs all	*99.45 %*	*90 %*	*99.53 %*
Basophils vs all	*100 %*	*100 %*	*100 %*

representation have been firstly extracted; an Artificial Neural Network and a Decision Tree have been then designed to discriminate among the five classes of leukocytes. Results described in [345] are reported in Table 3.17, expressed in terms of values of Accuracy, Sensitivity and Precision on the test set only for the implemented ANN classifier, because the value of Accuracy for the Decision Tree reported in [345] is almost equal to 70 %, that is much lower than the Accuracy obtained by the ANN classifier.

The reported results show that the proposed approach was suitable for a white blood cell classification; it has to be noted that performance was very good, even though the developed image processing procedure was quite simple.

In a subsequent work, Convolutional Neural Networks have been proposed to perform leukocyte discrimination by means of two approaches based on Deep Learning.

Firstly, a pre-trained Convolutional Neural Network has been deployed to extract features from the same set of segmented images of the previous work [345] and it was subsequently

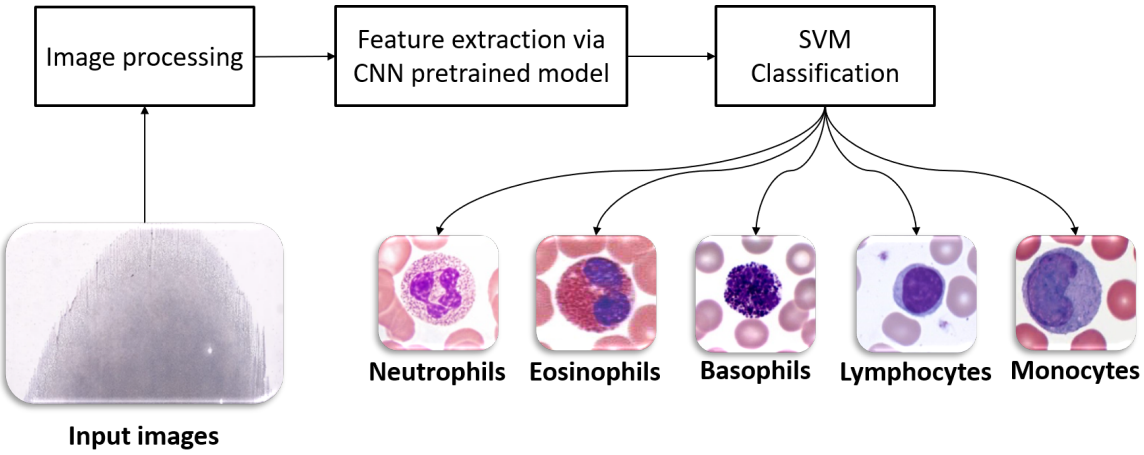


Fig. 3.34 Workflow for leukocyte classification considering CNNs as feature extractors

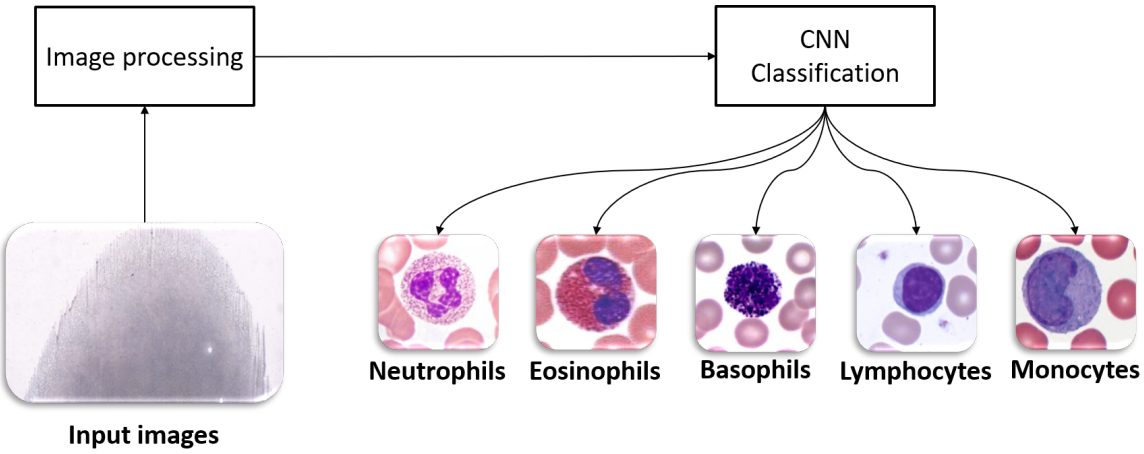


Fig. 3.35 Workflow for a leukocyte classification using a CNN as a classifier

combined with an SVM classifier to perform the desired discrimination. The workflow followed in this approach is represented in Fig. 3.34.

Then, the classification capabilities of the CNN have been investigated to classify leukocytes using the segmented images as inputs. In this work, the considered model is the AlexNet by Krizhevsky *et al.* [178]. Differently from the previous approaches, which included a step for computing features, hand-crafted or automatically computed using the CNN kernels, the workflow represented in Fig. 3.35, where the CNN is used for classification, shows that the feature extraction step is skipped.

As for the previous work, Table 3.18 and Table 3.19 report all the obtained results expressed in terms of Accuracy, Sensitivity and Precision on the test set.

Table 3.18 Values of Accuracy, Sensitivity and Specificity for leukocyte classification evaluated via a 1-vs-all approach using an SVM classifier.

	Accuracy	**Sensitivity**	**Specificity**
Neutrophils vs all	*97.89 %*	*98.78 %*	*95.76 %*
Lymphocytes vs all	*98.59 %*	*97.93 %*	*98.83 %*
Monocytes vs all	*98.67 %*	*48.15 %*	*99.76 %*
Eosinophils vs all	*99.38 %*	*66.67 %*	*99.61 %*
Basophils vs all	*100 %*	*100 %*	*100 %*

Table 3.19 Accuracy, Sensitivity and Specificity for leukocyte classification evaluated via a 1-vs-all approach using a CNN classifier

	Accuracy	**Sensitivity**	**Specificity**
Neutrophils vs all	*97.73 %*	*97.12 %*	*99.20 %*
Lymphocytes vs all	*97.73 %*	*99.70 %*	*97.03 %*
Monocytes vs all	*98.67 %*	*48.15 %*	*99.76 %*
Eosinophils vs all	*99.61 %*	*66.67 %*	*99.84 %*
Basophils vs all	*100 %*	*100 %*	*100 %*

The results reported show that both the proposed approaches based on CNNs seem to be very promising; in fact, the reported values of Accuracy are always higher than 95 %. At the same time, the Sensitivity and Specificity values for both approaches considered fluctuate depending on the type of cells. However, this result is understandable by considering that the amount of samples for particular cells, such as Neutrophils and Lymphocytes, was greater than the number of the other kinds of cells.

Chapter 4

Case Studies: Decision Support Systems based on Signal Processing

After the description of case studies regarding Decision Support Systems supporting the clinical diagnosis based on imaging systems reported in Chapter 3, this Chapter describes the research works conducted in the field of interest of this thesis during the Ph.D. activities, focusing on studies concerning the signal processing for diagnosing and assessing neurological disorders.

In details, the designed and implemented decision support systems and the technological framework for objectively assessing and monitoring different pathological conditions will be analysed in the clinical areas of neurology, psychology and physiatry. These works will deal with classification methodologies based on non-invasive biomarkers for early diagnosis, but also will investigate how innovative frameworks making use of Virtual Reality (VR) could improve the living conditions for older people.

4.1 Motivations

Along with tumoral diseases, whose pervasiveness and diffusion have been discussed in the previous Chapter, ageing is becoming a serious problem for modern society, especially in industrialised countries. In fact, according to a 2015 United Nations report on world population ageing, the number of people aged sixty and older worldwide is projected to more than double in the next thirty-five years, reaching almost 2.1 billion people [350]. Figure 4.1 shows the exponential growth of people older than sixty-five all over the world, from Sixties to today. Most of this growth will come from developing regions of the world, although the oldest old, people having more than eighty years of age, are the fastest-growing segment

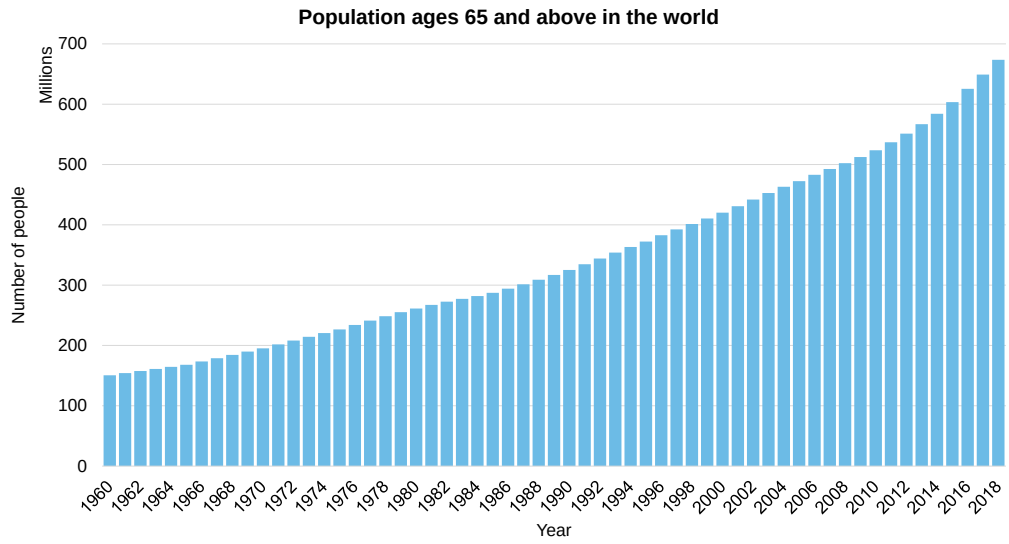


Fig. 4.1 Number of people older than sixty-five all over the world from 1960 to 2018. Data from

of the population in developed regions. Despite these improvements in life expectancy, dementia and related neurodegenerative conditions have arguably become the most dreaded maladies of older people. Furthermore, recent studies evidenced how the two phenomena, ageing and widespread diffusion of cancers, are strictly correlated with each other [351, 352], thus making ageing a severe, and steady, problem to deal with.

Based on these premises, the research works reported in this thesis come from the needs of the healthcare system to support clinical diagnosis and objectively assess the pathological conditions in order to improve the diagnostic accuracy and the therapeutic, or rehabilitation, efficacy for increasing the life expectancy of subjects and, at the same time, improve their wellness state.

The rest of this Chapter is organised as follows: Section 4.2 shows the innovative works for assessing the neurophysiological and cognitive status of different classes of subjects, considering the Alzheimer's Disease, Fibromyalgia and Mild Cognitive Impairment. Section 4.3, instead, reports the contributions about rehabilitation and assessment of physical deficits due to neurodegenerative disorders, focusing on the Parkinson's Disease.

4.2 Neurology and Psychology

As for areas related to, or dealing with, images, the assessment of the cognitive status, or the evaluation of other kinds of pathologies, are of fundamental importance, in order to decide the correct therapy for the patient. In fact, neurodegenerative diseases have different courses on subjects, e.g. based on the pathology itself, the genotypes of the subject and environmental factors.

This section describes all the research works conducted in the fields of neurology and psychology; in all cases, the research contribution aimed at discriminating the presence or absence of pathologies (e.g., Alzheimer's Disease vs Normal Ageing), or assessing the cognitive status due to ageing or traumatic conditions.

4.2.1 Artificial Neural Networks for Discriminating Alzheimer's Disease from Normal Ageing based on EEG Biomarkers

Alzheimer's Disease (AD) is the most prevalent neurodegenerative disorder affecting aged people. In AD, a progressive neurodegeneration leads to dementia, characterized by severe cognitive deficits, behavioural symptoms, and loss of autonomy in the daily life activities [353].

In the past years, the International Working Group (IWG) and the US National Institute on Ageing–Alzheimer's Association (NIA-AA) have proposed an algorithm for the diagnosis of AD based on in vivo biomarkers and clinical phenotypes of disease [354–358]. For example, the cerebrospinal fluid (CSF) offers a window to the brain as the brain's metabolism and pathology is reflected in the CSF that could easily be collected through a lumbar puncture, which is a safe and well-tolerated procedure. However, it is still an invasive procedure having the risk of complications during its execution.

The last IWG guidelines encourage the use of topographic markers for evaluating the pathology, even if they are not diagnostic at all. These markers are quite useful for mapping structural and functional impairment of the brain over time, especially in elderly subjects with initial objective evidence of Mild Cognitive Impairment (MCI) including memory and other cognitive domains but with preserved independence in the daily activities. The topographic markers include maps of brain hypo-metabolism [359], as revealed by fluorodeoxyglucose (FDG) - PET, and maps of brain atrophy and abnormalities of structural and functional brain connectivity, as revealed by structural and functional magnetic resonance imaging (MRI). All those methodologies can capture several processes of disease, but their use is limited because

of low availability of the instruments, cost or invasiveness, especially for serial recordings over time for monitoring purposes.

Keeping the intrinsic limitations of CSF, MRI and PET in mind, several independent research groups tested indexes of resting state eyes-closed electroencephalographic (rsEEG) rhythms as candidate topographic markers of AD [360]. EEG rhythms are the most important features of the collective behaviour of brain neural populations and are very relevant for assessing human cognition. Furthermore, EEG procedures are largely available in any country, well tolerated by patients, not affected by subjects' anxiety or task difficulty, and can be repeated over time without habituation effects [360].

Previous studies in AD patients and elderly subjects with amnesic MCI have shown that rsEEG may have promising markers for a neurophysiological evaluation of disease status as topographic markers. When compared to groups of normal elderly (Nold) subjects, AD groups have been characterized by the high energy of widespread delta and theta rhythms, as well as by the low energy of posterior alpha and/or beta rhythms [361–365].

In literature, the use of rsEEG variables as neurophysiologic topographic markers of AD implies that these variables could classify Nold and AD individuals at least with a moderate classification accuracy of 75 – 80 %.

In previous studies, cortical sources of rsEEG rhythms in MCI, AD, and control groups of subjects were estimated by the freeware low-resolution brain electromagnetic tomography (LORETA). Loreta analysis of limited frequency bands can be used to determine which regions of the brain are activated during different states or mental tasks, helping to determine if the brain is operating in an electrical optimal way. There are multiple implementations of LORETA:

- sLORETA: standardized low resolution brain electromagnetic tomography [366]. It has no localization bias in the presence of measurement and biological noise.
- eLORETA: exact low resolution brain electromagnetic tomography [367]. The first ever 3D, discrete, distributed, linear solution to the inverse problem of EEG/MEG with exact localization (zero localization error).

The aim was to enhance the spatial information content of scalp-recorded EEG data and to unveil topography of EEG abnormalities associated with AD from prodromal to overt clinical stages [360, 368–371]. It was reported that temporal, parietal, and occipital cortical sources of delta and alpha rhythms were altered in AD groups compared with control groups as a function of cognitive deficits and abnormalities in brain integrity [360, 368–371].

In a previous work, the same research group tested the hypothesis that Nold and AD individuals with dementia can be discriminated with moderate accuracy using topographic markers of source current density and functional connectivity of the rsEEG [360]. Results showed a classification accuracy of 75.5 % in the discrimination of 120 AD patients with dementia from 100 matched Nold subjects based on cortical source current density.

4.2.1.1 Research Contribution

In this research work, it was tested whether a multivariate classification with Artificial Neural Networks (ANNs) cold improve the classification accuracy, obtained in [360], of those original rsEEG markers in AD and Nold individuals. The main issue was whether the combined use of cortical source current density and functional connectivity (lagged linear connectivity - LLC) as inputs of a trained ANN would provide more accurate classifications than those obtained with the two classes of spectral EEG markers considered separately. LLC is a linear measure of rsEEG coherence that overcomes the problem of the high phase synchronization and the zero-lag coherence possibly introduced by the procedure of eLORETA source estimation. LLC is also expected to minimize the influence of a third rsEEG source having an influence on the instantaneous coherence between two rsEEG sources not dependent each other (the so-called "common feeding" issue).

In the considered study, the ANNs for the classification of the Nold and AD individuals were set-up according to the optimised approach reported in [372]. Each investigated ANN worked on a different set of features. The experimental design aimed at computing the sensitivity, specificity, and accuracy of the classifications of the Nold and AD individual datasets for the following input data:

1. The 4 most discriminant markers of eLORETA source current density (SCD), namely parietal, temporal, and occipital theta/alpha 1 and the occipital delta/alpha 1;

2. The 4 most discriminant markers of eLORETA source functional connectivity, namely inter-hemispherical occipital delta/alpha 1, intra - hemispherical left occipital-temporal theta/alpha1, right parietal-limbic alpha1, and right occipital - temporal theta/alpha 1;

3. The combination of the mentioned 8 most discriminant markers of eLORETA source current density (SCD) and the functional connectivity (LLC).

The scalp areas related to the features analysed and classified by the ANN are reported in Figure 4.2.

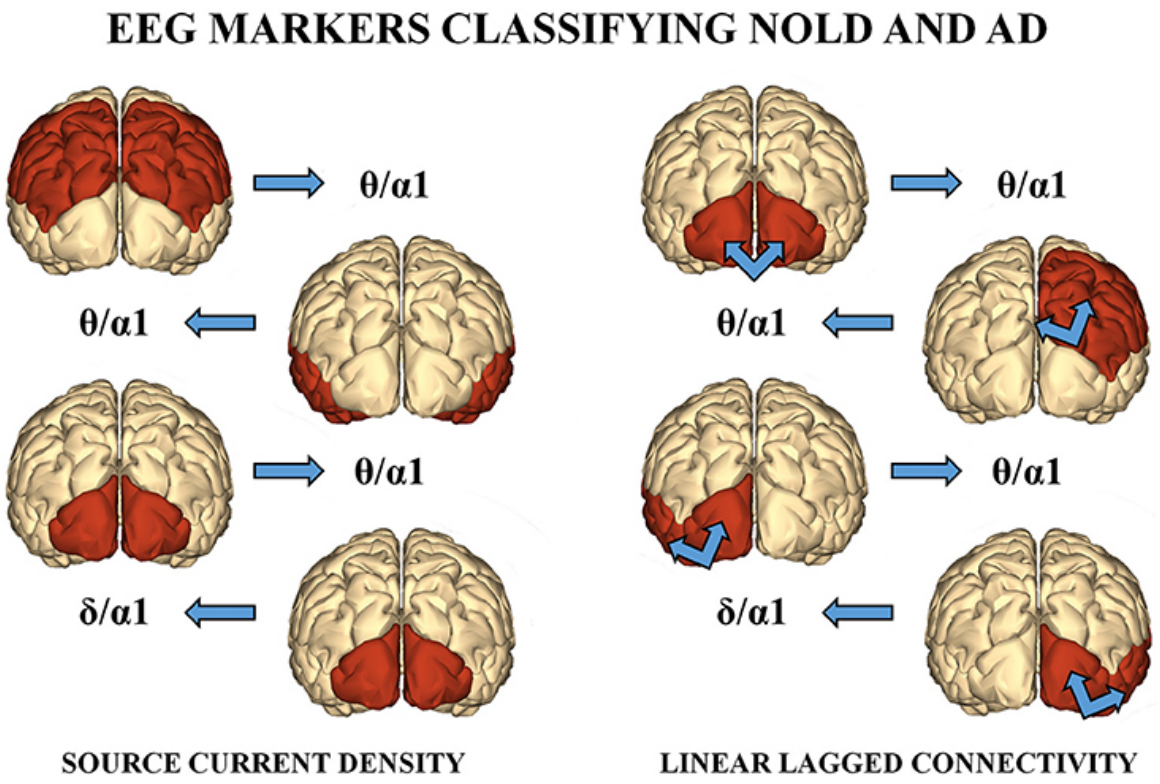


Fig. 4.2 A topographical representation of the following best 4 discriminant markers of the rsEEG SCD for the classification of Nold and AD individuals. These markers are the following (from the top to the bottom): parietal theta/alpha 1, temporal theta/alpha 1, occipital theta/alpha 1, and occipital delta/alpha 1. (Right): A topographical representation of the following best 4 discriminant markers of the rsEEG LLC for the classification between Nold and AD individuals. These markers are the following (from the top to the bottom): inter-hemispherical occipital delta/alpha 2, intra-hemispherical right parietal-limbic alpha 1, intra-hemispherical left occipital-temporal theta/alpha 1, intra-hemispherical right occipital-temporal theta/alpha 1.

Table 4.1 Accuracy, sensitivity (true positive rate), and specificity (true negative rate) of the ANNs proposed, expressed as a percentage (mean ± standard deviation).

	Feature set		
	4 SCD	**4 LLC**	**4 SCD + 4 LLC**
Sensitivity (mean % ± std)	*79.3 ± 10.6*	*74.2 ± 11.4*	*80 ± 10.8*
Specificity (mean % ± std)	*74.3 ± 13.2*	*68.9 ± 14.6*	*72.7 ± 12.9*
Accuracy (mean % ± std)	*77 ± 5*	*71.6 ± 6.5*	*76.7 ± 5.2*

Figure 4.3 shows the peculiar architectures of the ANNs used for the above three sets of features (they were optimised by the single objective genetic algorithm reported in [108, 119]). For a given session of training, validation of test of the ANN, the mean and standard deviation of the classification accuracy were computed performing 500 repetitions of the session, randomly permuting the input dataset.

Among the above sessions of ANN classification of the Nold and AD individuals, the best 4 discriminant markers of the rsEEG source current density reached the following classification rate: sensitivity of 79.3 %, specificity of 74.3%, and accuracy of 77 %. Furthermore, the best 4 discriminant markers of the rsEEG source lagged linear connectivity showed sensitivity of 74.2 %, specificity of 68.9 %, and accuracy of 71.6 %. Finally, the combination of the above best 8 discriminant markers of the rsEEG source current density and linear lagged connectivity exhibited sensitivity of 80 %, specificity of 72.7 %, and accuracy of 76.7 %. Table 4.1 reports these values associated with their standard deviations.

The Shapiro-Wilk normality test ($p < 0.05$) showed that the accuracy values in the 500 iterations of any classification session (i.e., the best 4 discriminant markers of the rsEEG source current density; the best 4 discriminant markers of the rsEEG linear lagged connectivity; the above best 8 discriminant markers) did not follow a Gaussian distribution. From these distributions, the Kruskal-Wallis test disclosed a statistically significant effect ($p < 0.0001$) while Dunn's post-hoc test revealed some interesting statistically significant differences in the classification accuracy between session pairs. Specifically, the classification accuracy was higher for the best 4 discriminant markers of rsEEG source current density than the for best 4 discriminant markers of rsEEG source linear lagged connectivity ($p < 0.0001$). Furthermore, this accuracy was lower for the best four discriminant markers of rsEEG source linear lagged connectivity than the best eight discriminant markers of the rsEEG source current density and linear lagged connectivity ($p < 0.0001$). In contrast, no difference in the classification accuracy was found between the best four discriminant markers of the rsEEG

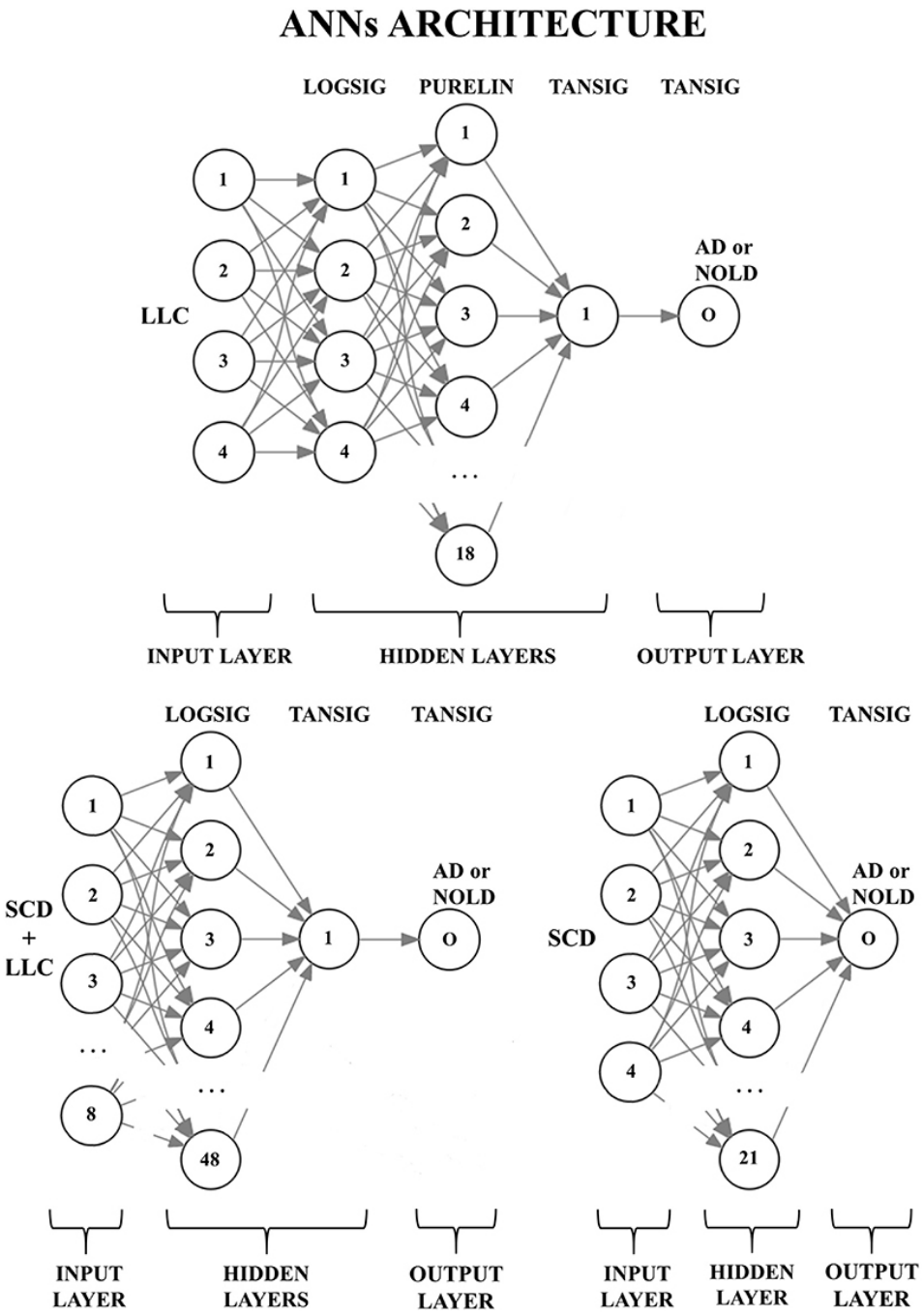


Fig. 4.3 Structures of the three artificial neural networks (ANNs) used to classify Alzheimer's disease patients with dementia (AD) from Normal elderly subjects (Nold). EEG markers are given as inputs in the first layer (input layer); every node of every successive layer (i.e., the hidden layers and the output layer) is characterized by an activation function: A non-linear function to decide, in analogy with biological neurons, the output of the node (0 or 1). The output node (O) provides the classification result (AD or Nold). Legend for the input markers: (top) the four best Lagged Linear Connectivity (LLC) markers; (bottom left) the four best LLC markers together with the four best Source Current Density (SCD) markers; (bottom right) the four best SCD markers. Legend for the activation functions: log-sigmoid (logsig), linear (purelin), and tan-sigmoid (tansig).

source current density and the best eight discriminant markers of the rsEEG source current density and linear lagged connectivity ($p > 0.05$).

These results suggest that a non-linear (ANN) multivariate classification cross-validated the results obtained in the previous work using a linear univariate classifier [360]. Although the linear rsEEG markers of cortical current density and connectivity probe different relevant neurophysiological mechanisms underpinning cortical arousal and vigilance in AD patients, they provide quite redundant information for classification purposes. In future AD studies, inputs to ANNs should combine the markers considered here with other linear (i.e., directed transfer function, phase lag index) and non-linear (i.e., chaos, entropy, synchronization likelihood) rsEEG markers to improve the moderate classification accuracy these markers can achieve (about 75 - 80 %).

4.2.2 Chronic Pain Modulation by Motor Cortex Activation

Fibromyalgia (FM) is a condition of chronic pain whose etiopathogenetic mechanisms are not yet known [373]. The most typical symptom of FM disease is the widespread skeletal muscle pain, with associated fatigue, alteration of mood, sleep disturbance, cognitive dysfunction and poor quality of life [374, 375]. Several experimental studies have found an analgesic effect on pain induced by non-invasive brain stimulation techniques such as repetitive transcranial magnetic stimulation (rTMS) and transcranial direct current stimulation (tDCS) on the motor areas [376–380]. The activation of the primary motor cortex seems to interact with the cortical regions responsible for pain processing and have a modulation function on the tM1-thalamic inhibitory networks (thalamus is the major source of cortical inputs shaping sensation, action and cognition) [381]. Recent evidence indicates an altered functional organization of the primary motor cortex in subjects suffering from chronic pain [382].

Researchers suggest that motor activity leads to an improvement in the quality of life of patients [383, 384] so exercise is recommended for the treatment of fibromyalgia symptoms. Moreover, FM patients have a peculiar limitation to the movement that can manifest itself with dysfunctions in muscle coordination, difficulty in postural control and reduced speed of motor performance [385]. However, patients suffering from chronic pain are unlikely to exercise because they fear their painful condition may worsen [386]. The exploration of the functional basis of motor cortical areas may be an interesting field to investigate in FM disease.

4.2.2.1 Research Contribution

The research work conducted in this context aimed to explore the complex mechanisms of interaction between motor activity and pain, which have not been yet clearly understood. In order to evaluate the interaction between these two conditions, cortical responses have to be measured and analysed from both cognitive and functional point of views. The co-recording of EEG and functional Near-Infrared Spectroscopy (fNIRS) has been demonstrated to be a very promising technique to explore both the electrical and metabolic activities during multimodal stimulations condition [387]. In a preliminary study by Gentile *et al.*, the concomitant recording of fNIRS and laser evoked potentials (LEPS) allowed to explore the complex mutual interference between motor cortex activation and the processing of painful stimuli in FM patients and healthy subjects [388]. The choice of the multimodal method of EEG-fNIRS simultaneous recording was aimed at exploring the electrophysiological and functional mechanisms underlying the voluntary activation of cortical areas involved in movement and in pain processing. The advantage of co-registration lies also in being able to obtain functional and electrical data at low cost and with good tolerance to motion artefacts [389]. Moreover, the light emission in the near-infrared does not contaminate the electro-physiological signal and vice versa [390].

The principal aim of this research work was to investigate the motor cortical metabolism and changes of LEPs parameters in FM patients and healthy subjects, testing whether there were possible changes induced in motor cortex activation by laser stimulation and modifications in LEPs during movement tasks. The FM patients showed reduced modulation of cortical motor activity during movement as a probable effect of chronic inhibition. The LEPs amplitude decreased during movement task both in patients and controls, though the FM group showed greater internal variability.

The specific aims of the described work were:

- To compare the changes of haemoglobin activity recorded in the motor cortical regions during slow and fast finger tapping task between patients and controls;
- To compare LEPs changes during slow and fast motor activity between patients and controls;
- To verify the effects of laser stimulation of the moving hand and contralateral non-moving hand on haemoglobin activity
- To correlate FNIRs/LEPs changes with clinical data in the FM group.

Thirty-eight patients with fibromyalgia diagnosis and twenty-one healthy subjects served as participants. All the subjects were right-handed, as confirmed by Edinburgh Handedness Inventory [391]. The experimental procedures of the study were approved by the Ethics Committee of Bari Policlinic General Hospital. All the participants signed a written informed consent before inclusion in the study. The exclusion criteria for the recruitment of the study were: less than eight years of education, any peripheral or central nervous system (CNS) diseases, including spinal cord diseases and radiculopathies, psychiatric diseases, diabetes, active and/or positive history of thyroid insufficiency, renal failure, auto-immune diseases, inflammatory arthritis, systemic connective tissue disease, present or previous history of cancer, as well as use of drugs acting on the CNS or chronic opioid therapy. The FM patients were admitted to the study after their first visit at the Applied Neurophysiology and Pain Unit of Bari University, and before taking the suggested treatment.

Participants lied on a comfortable chair in a relaxed state. Before the beginning of the experiment the researcher explained the experimental protocol to each subject. Subjects were invited to perform a finger tapping task, pressing a push-button panel with the right-hand thumb in 2 modalities, one in a slow way at a defined rhythm, and one pressing as fast as possible. The experimental procedure was based on nine sessions for each subject, as reported in Fig. 4.4.

Preliminarily we recorded 2 minutes of resting state, during which the participant was requested to stay relaxed with open eyes, fixing on a point on the computer monitor. The subsequent experimental conditions were randomised, and each pre-task baseline was one minute in duration. In the laser stimulation condition the participant received laser stimuli on the right- or left-hand dorsum. To keep the participant's attention active, the experimenter asked him to count the perceived laser stimuli. Participants were asked to concentrate on the motor task while keeping the rest of their body motionless. The slow finger tapping (SFT) task consisted of pressing a button with the right thumb every 5 seconds following the indications of the experimenter. The fast finger tapping (FFT) task consisted of clicking a button as quickly as possible. We used the controlled-slow-speed and the maximal-fast-speed the subjects could reach to evaluate the net effect of the movement or of the maximal motor performance on the cortical metabolism.

Both SFT and FFT procedures were repeated during laser stimulation of the right hand (moving hand) and the left–non-moving—hand (subjects performed motor task with the right hand while stimulated on the left one). The laser stimulation of the inactive-left hand served to evaluate the net effect of movement as distractor from painful stimulation.

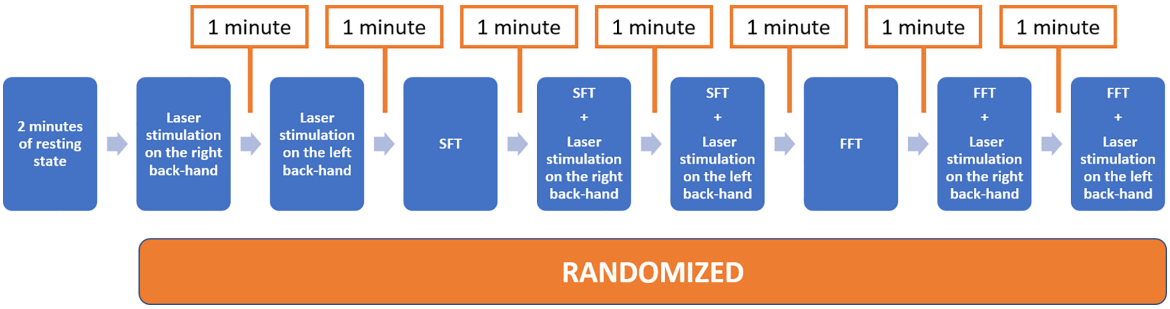


Fig. 4.4 Experimental Design: Randomised sequence of experimental conditions after two minutes of resting-state

The speed of the finger tapping tasks was calculated as the number of times per second in which the subject clicked the button on the panel. The interstimulus interval between the experimental conditions was fixed at 60 seconds.

The experiment was performed with a co-recording fNIRS-EEG by a compatible cap and a black over-cap to mitigate a possible interference generated by ambient light on the fNIRS signal. We used a continuous wave NIRS system (NIRSport 8X8, Nirx Medical Technologies LLC, Berlin, Germany). The fNIRS instrument included LED sources and photosensitive detectors. Each source employed two LEDs emitting a near-infrared light at 760 nm and 850 nm. The fNIRS signal sampling rate was 7.81 Hz. The arrangement of sources and detectors resulted in a total of 20 fNIRS measurement channels, 10 for each side of hemisphere (Fig. 4.5). Probes were placed on the motor areas. The inter-optode distance was fixed at 30 mm since in was demonstrated to be the optimal to measure the haemodynamic activity variations over the cerebral surface [392]. Each recording was preceded by a calibration procedure to verify that a good fNIRS signal acquisition was guaranteed. During the calibration procedure the NIRSport instrument determines the signal amplification for each source-detector combination.

EEG data, instead, were recorded and amplified using Micromed System Plus (Mogliano Veneto, Italy) at a sampling frequency of 256 Hz. We used a montage with 61 scalp electrodes positioned according to 10 - 20 International System with reference to the nasion and the ground electrode at the Fpz. Two additional electrodes located above the eyebrows served for electro-oculogram recording. The impedance was kept below 5 $k\Omega$. During the EEG recording, we used digital filters in the 0.1 - 70 Hz range and a 50 Hz notch filter to allow signal inspection.

Regarding laser stimulation, nociceptive stimuli consisted of laser pulses delivered by a CO2 laser (wavelength, 10.6 mm; beam diameter, 2 mm, Neurolas Electronic Engineering Florence, Italy). The interval between each laser stimulus was fixed at 10 seconds. Patients

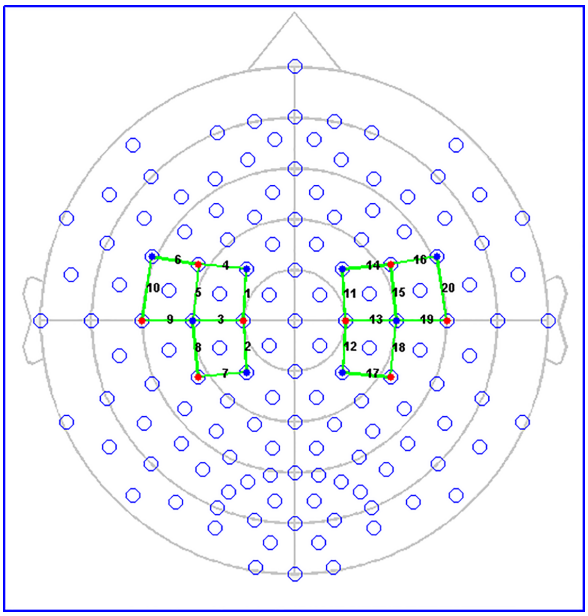


Fig. 4.5 Channels and optodes configuration. Filled red circles indicate sources. Filled blue circles represent detectors. The numbered green lines show the recording channels.

and controls were stimulated on the back of the hand by laser stimuli of 30 msec duration. The laser intensity was adjusted based on the patient's specific pain perception threshold [393, 394].

The fNIRS signal processing included the baseline subtraction, removal of discontinuities and motion artefacts. Raw data were digitally filtered in the band-pass 0.005 - 0.2 Hz to remove low oscillations, like respiratory and cardiac frequencies from fNIRS signal. Optical intensity measurements were converted to oxyhaemoglobin ($\Delta HbO2$) and deoxyhaemoglobin (ΔHb) concentration by the modified Beer-Lambert law [395, 396]. The unit of haemoglobin concentration is measured in mmol per liter (mmol/liter). The mean values of the haemoglobin concentration were subtracted to calculate the changes in $\Delta HbO2$ and ΔHb during the experimental tasks.

To analyse the EEG signal, an open-source Matlab toolbox, namely Letswave 6 (André Mouraux, Brussels, Belgium; www.letswave.org). The pre-processing signal method consisted of frequency filtering, bad electrodes interpolation, segmentation in epochs, artefact rejection, independent component analysis (ICA) decomposing method for ocular artifacts. The Butterworth IIIR filter was applied for bandpass filtering in the 0.01 - 30 Hz range. Bad channels were removed with subsequent interpolation. The motor artefacts were visually inspected and removed. We applied ICA method to remove ocular and motor artefacts from the EEG signal. EEG epochs were averaged in the 100 msec preceding and 1000

msec following laser stimuli. In this study, the N1, an early component detected on the contralateral temporal regions at the stimulation side (T3 or T4 channel), and the N2 and P2 vertex waves (late component) recorded on the Cz electrode were considered [394, 397]. The waves amplitude was computed from the baseline. Latencies were measured from the 0 time to the maximal amplitude of each wave.

Finally, statistical data analysis was performed. For all statistical tests, a p-value lower than 0.05 was considered statistically significant. A two-way analysis of variance (ANOVA) corrected for age was used for the comparison of finger tapping speed between groups.

A statistical parametric mapping was performed for fNIRS data [398]. The F-contrast between groups plotting the F-values for all the channels during the baseline condition, FFT and SFT conditions, for both $\Delta HbO2$ and ΔHb. Results are reported in Fig. 4.6 where F-contrast comparisons between the FM and control groups are represented for the different experimental conditions. The significant changes in oxyhaemoglobin levels between groups were on the channel 10 for $\Delta HbO2$ for the resting state condition. Concerning FFT, the higher difference between the groups is located on channels 4, 6 and 10 for $\Delta HbO2$. During FFT with concomitant laser stimulation on the left-hand condition, the higher difference between the groups was located on channel 4, 6 and 10 for $\Delta HbO2$ and on channel 6 for ΔHb. Considering SFT condition, the higher difference between the groups was located on channel 6 for ΔHb. Finally, the higher difference between the groups was located on channel 6 for ΔHb during the SFT condition.

Group-level average LEPs in the experimental conditions with laser stimulation on the right and left hands are shown in Figure 4.7 and Figure 4.8, respectively. For most of the investigated LEPs parameters both in the patient group and in the control group there were no statistically significant changes between the different experimental conditions. However, the N1 and N2P2 amplitude was significantly smaller in patients than controls when the stimulation was on the right hand. Moreover, we observed a significant difference in N1 latency between groups for experimental condition of FFT task during laser stimulation on the right hand. We also observed no significant changes in LEPs parameters when the stimulation was on the left hand independent from experimental condition.

The main results of this study partly confirmed previous findings reported in [388]. Patients suffering from FM had a reduced motor performance as tested by finger tapping task, and a reduced tone of cortical motor areas, especially evident during fast movement. Concurrent phasic pain stimulation had limited effect on motor cortex metabolism in both groups, nor the motor activity changed the laser evoked responses in a relevant way. The reduced tone of motor areas activation was independent of FM duration and severity.

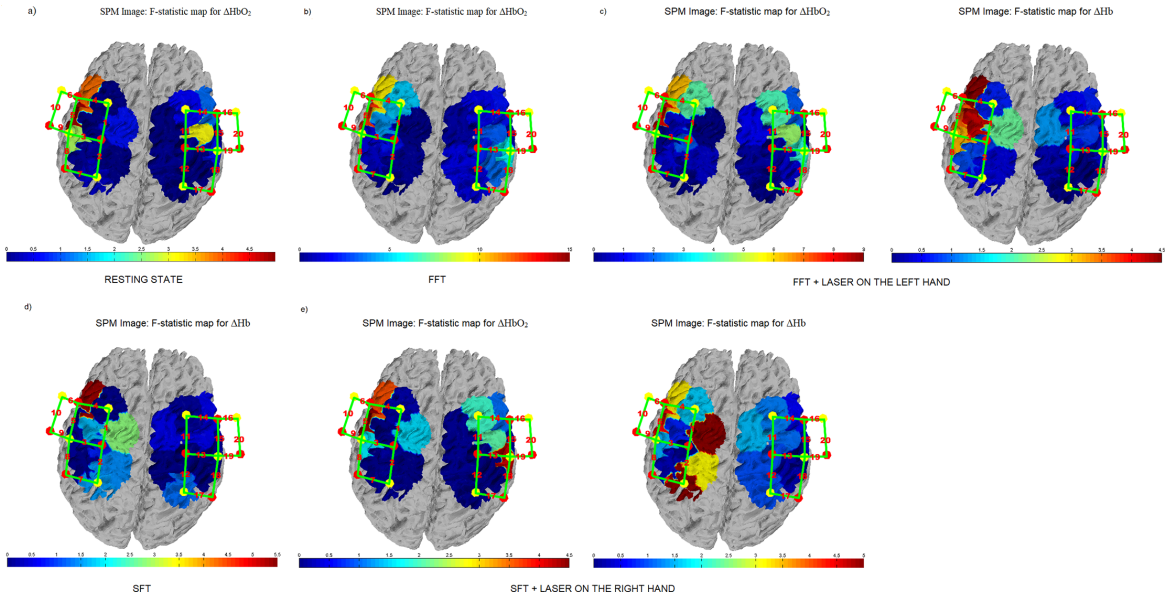


Fig. 4.6 F-statistic values of $\Delta HbO2$ and ΔHb during different conditions. FM and Control groups activation maps using canonical HRF model. The higher difference between control subjects and patients' activations is represented with the red colour. (a) F-statistic values of $\Delta HbO2$ during the resting state condition; (b) F-statistic values of $\Delta HbO2$ during the FFT condition; (c) F-statistic values of $\Delta HbO2$ and ΔHb during the FFT + LASER ON THE LEFT-HAND condition; (d) F-statistic values of ΔHb during the SFT condition; (e) F-statistic values of $\Delta HbO2$ and ΔHb during the SFT + LASER ON THE RIGHT-HAND condition.

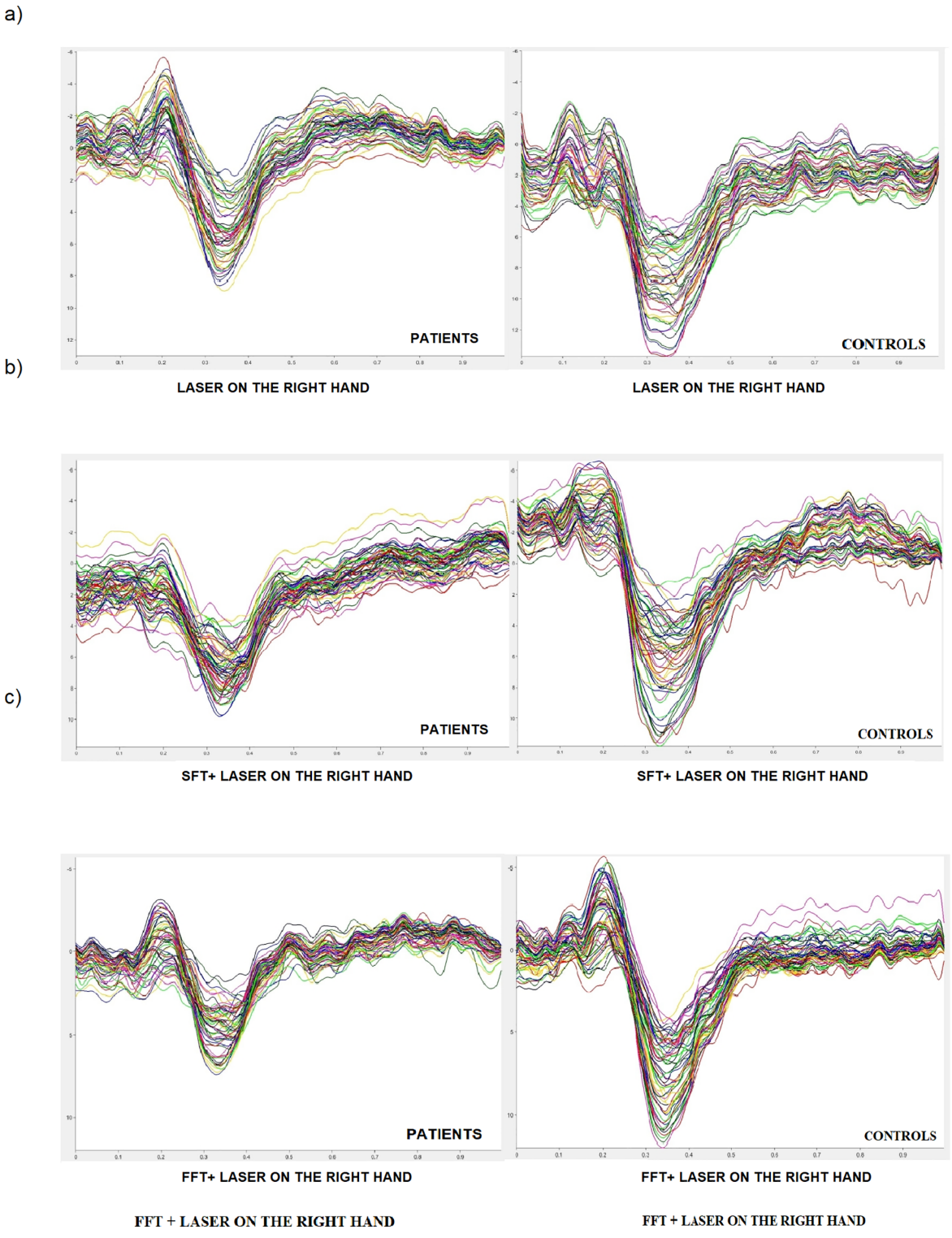
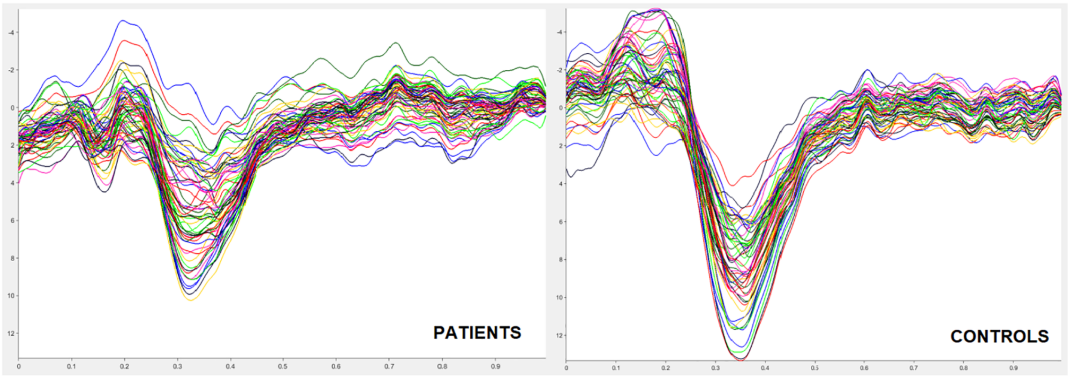
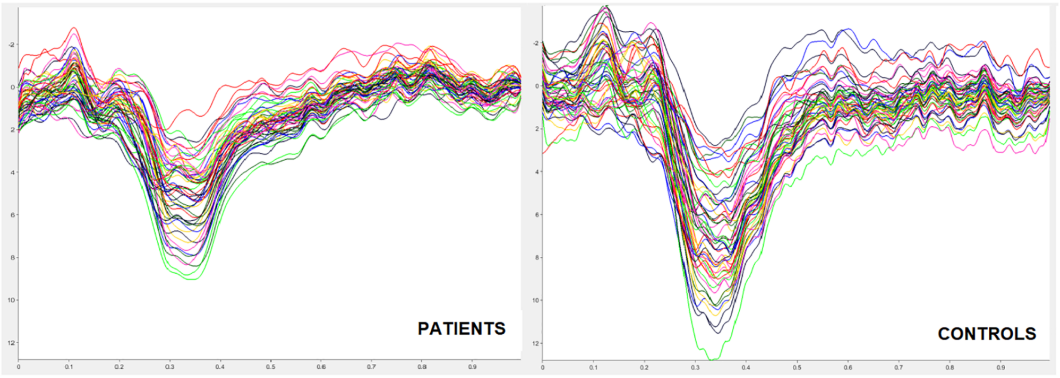


Fig. 4.7 Grand average of LEPs by right hand stimulation in patients and controls. (a) laser on the right hand, (b) SFT task during concomitant stimulation on the right hand, (c) FFT task during concomitant stimulation on the right hand both in patients and controls.

a)



b)



c)

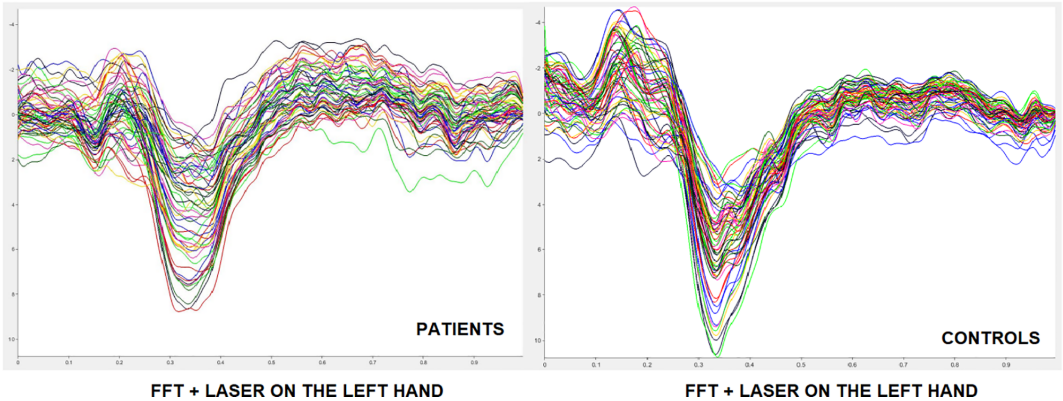


Fig. 4.8 Grand average of LEPs by left hand stimulation in patients and controls. (a) laser on the left hand, (b) SFT task during concomitant stimulation on the left hand, (c) FFT task during concomitant stimulation on the left hand) both in patients and controls.

Reduced motor performance and motor cortical areas activation in FM patients. The slow motor performance expressed by FM patients during finger tapping was present in all the experimental conditions that requested a rapid movement and independently of the laser stimulation on the active or on the inactive hand. We hypothesize that FM patients had a lower speed of finger tapping task than controls due to the interaction of several factors linked to pain condition. The low motor performance of FM patients could be due to fear of movement or cognitive problems with impaired motor programming [399]. It is possible that the reduced speed of information processing which often characterizes patients with chronic pain can also affect the control and speed of motor responses [400]. The motor impairment could be a constitutional tract in FM, as it seemed independent from disease severity and duration.

As expected, the spatial distribution of brain activity during the movement of the right hand involved the left prefrontal regions, corresponding to the primary and supplementary motor cortex. In the resting state, patients and controls showed a significantly different activation in channel 10, with a trend toward a greater level of concentrations of $\Delta HbO2$ in healthy subjects. Furthermore, results indicated that there were significant differences in motor cortical activation between patients and controls during the fast movement condition on channels 4, 6, 10. We did not observe a compensatory activity of right hemisphere in FM patients, as generally occurs in unilateral motor cortex dysfunction. Also, no-relevant correlation between the motor speed and haemodynamic responses was found, either in patients or in controls. The results of haemodynamic responses suggest that FM patients could have a dysfunction in supplementary and primary motor cortex modulation. In this regard, a possible altered cortical motor function could characterize the chronic pain syndrome. Patients did not show any modulation of haemoglobin levels during the concurrent laser stimulation, confirming a rigid modality of motor cortical activation. In fact, scientific evidences suggest a complex mechanism of reorganization of the motor cortex in conditions of chronic pain, whose functioning is not yet clear [401].

Effects of movement on laser evoked responses. The obtained results indicated that the amplitude of the LEPs components was different between FM patients and healthy subjects, independently of the different experimental conditions. According to previous studies [394, 402], patients with chronic pain can present alteration in expression of nociceptive responses. Fibromyalgia is characterised by a complex interaction of peripheral and central neuronal factors with a dysfunction of small fibres coexisting with central amplification of pain. These phenomena could lead to variable group results, depending upon the prevailing phenotypical expression. In the present results, FM patients showed in basal, i.e. conditions without

motor activity, with smaller LEP responses as compared to controls. In general, movement seemed to affect LEPs in a not relevant way either in patients or in controls. Healthy subjects exhibited an increase of N1 latency during the execution of the fast finger tapping task. Probably this result is due to a possible movement–related somatosensory interference on cortical areas receiving multimodal somatosensory stimuli [403]. The features of N1 wave were unchanged during the other experimental conditions, suggesting that this interference could emerge only during fast movement. This phenomenon was absent in FM patients, as the reduced tone of motor cortex activation and the low motor performance could exert slight interference on concurrent cortical pain processing networks. If the lack of modulation effect of the finger tapping task on the amplitude of the laser cortical responses could be reasonable in FM for the low motor efficiency, the same phenomenon occurring in healthy subjects deserves further comments. The N2P2 reduction occurred in the phase prior to the motor execution, when the laser stimulus was delivered on the hand that was supposed to move. In this case the process of movement preparation generated an inhibition of the vertex LEPs that was independent from a pure cognitive distraction effect.

4.2.3 Virtual Reality to Improve Wayfinding in Ageing People

The progressive extension of human survival in the occidental society generates new problems in the organization of daily living activities and optimal environmental contexts, where the residual capabilities may be facilitated and even promoted in the presence of cognitive dysfunction. In normal ageing, visual recognition of daily life stimuli is impaired with respect to young people, depending upon the intrinsic characteristics of the object and the context in which the old person is embedded [404–406]. Spatial recognition is a fundamental element to achieve a good orientation in different environments, like a house, hospitals, and health facilities, in order to facilitate activities in daily living. Spatial disorientation and reducing wayfinding abilities is an early and invalidating symptom of dementia. Architectural wayfinding for old people and dementia were also studied with regard to nursing homes, where structural design, as well as furniture, light, and colors, may be adapted to dementia - friendly environment models [407]. However, the prevention of progressive loss of functional capability may be reached through the adaptation of the environment to the sensory and cognitive dysfunctions, caused by ageing and dementia, in order to prolong the permanence at home and avoid institutionalization. Moreover, wayfinding would be improved also in hospitals, where aged people are often disoriented and prone to incongruent behaviors. Some easy modifications of ambient color and luminance may be obtained by

home automation, which may improve wayfinding by enabling the remote control of site lights and luminance [408].

In order to design the best conditions to be applied for site recognition improvement, Virtual Reality (VR) techniques may efficaciously reproduce home, hospital or nursing house ambients, enabling the extraction of cognitive response to environment features changes. In fact, there are recent evidences on the reliability of cognitive-related responses obtained by oddball paradigms realized in VR frames [409–412], which seems to be a solid method to explore the capacity to recognize a target place during virtual wayfinding [411]. In theory, VR might test the real environment's impact on cognitive capacities in aging and dementia, as well as any change easily produced by home automation control to improve wayfinding even in unknown places. So far, the main aims of the described research study have been:

1. The validation of a VR model consisting in the reproduction of a real house environment modified by home automation as a reliable stimulus to obtain a P3b response similar to that resulted from standard visual oddball paradigm [412].

2. The preliminary testing of this novel model in a cohort of young and old healthy subjects, in order to prove possible age-related changes of target places recognition in response to ambient condition changes obtained by home automation techniques. This has been done in view of a possible application to mild cognitive impairment and dementia, to make changes of home or hospital environment easy and rapid for a better cognitive impact and facilitation of daily living activity.

4.2.3.1 Research Contribution

In order to immerse the user in the VR scenes, an Oculus Rift DK2 was used, which is a virtual reality head-mounted display headset developed by Oculus VR. For the purpose of this project, when the user wore the device, he saw the virtual scene through which the visual stimulation occurred. Binocular vision is the way in which we see two views of the world simultaneously; the view from each eye is slightly different and our brain combines them into a single three-dimensional stereoscopic image, an experience known as stereopsis. The Oculus Rift presents two images, one for each eye, generated by two virtual cameras separated by a short distance (Fig. 4.9). The Oculus Rift DK2 employs an OLED panel for each eye, with a resolution of 960x1080 and a refresh rate of 75 Hz (it globally refreshes, rather than forming the image line by line). The panels have low persistence, meaning that each frame is displayed only for 2 ms. This combination of higher refresh rate, global refresh, and low persistence let the user experience none of the motion blurring or judders that are

Fig. 4.9 The Oculus Rift used in the experiments.

experienced on a regular monitor. It uses high-quality lenses to allow for a wide field of view. The separation of the lenses is adjustable by a dial on the bottom of the device, in order to accommodate a wide range of inter-pupillary distances. In order to work, the Rift was connected by a cable to a PC equipped with a powerful GPU, at least equivalent to an NVIDIA GeForce GTX 970 or AMD R9 290, and a CPU at least equivalent to an Intel i5-4590.

The P3b paradigm was designed in order to test the best features of a target site to be recognised in a virtual ambient reproducing a real house. The bathroom door was selected as the target stimulus, for its presence in all houses and health facilities, where an elderly subject could have difficulty in moving toward this room essential for daily living.

In the virtual environment that simulates a home, all rooms that were different from the bathroom were identified as frequent stimuli (F) while the rare stimulus was assigned to the bathroom room itself. To differentiate the target stimulus from the frequent ones, and to understand which features are better for target room doors recognition, bathroom doors were semi-opened, in order to permit the recognition of the typical furniture. In addition, different colour and light conditions, easily obtained in the real ambient from a remote home automation control, were applied. Bathroom doors were illuminated in white, like the other non-target doors (W), or coloured with a green (G) or red spotlight (Fig. 4.10). Subjects were informed that they would perform a virtual walking through an apartment, looking for the bathroom door, and were requested to press a button as soon as the bathroom door appeared in the field of view. To make the experiment easier, navigation inside the environment was automatically controlled and had a fixed duration; in this way, rare and frequent stimuli were

Fig. 4.10 Different ways for representing the target stimulus

proposed in succession and in a controlled number. In order to examine the best position for the bathroom to be recognised within the virtual house, three different Virtual Environments (VE) were depicted, where the bathroom was designed in the aisle (A), living room (L) and bedroom (B), respectively. Each VE constituted a single block, where each block included 30 rare stimuli (10 for each color spotlight) and 120 frequent ones (all other rooms doors). Thanks to the anchors inserted in the VE, the system computed random virtual paths that, alternatively, started from the bathroom, ending up in another place of the house and vice versa, in order to have a rare stimulus followed by a certain number of frequent ones. During the automated navigation in VE, whenever an open door entered the field of view of the virtual camera, a trigger was sent to the EEG recording system according to the door type; in particular, bathroom door throws a rare trigger, with different values depending on the spotlight colours, while the doors of the remaining three rooms launched the same frequent trigger value.

During the experiment, a virtual agent, equipped with a virtual camera that reproduced its visualization on the user's head-mounted display, followed the route. In order to allow the health technicians to supervise the experiment evolution, an operating mode of the application was developed. In this modality, the operator saw the same virtual environment visualization of the patient and other information related to the acquisition protocol. A monitor, used exclusively by the operator, was in fact employed. In the operator display, trigger information showed which type of stimulus was visible in the field of view and, accordingly, which trigger code was sent to the EEG recording system. Other information provided to the operator were the number of stimuli sequence blocks, and the next destination, e.g. "bathroom" or "other place" (Fig. 4.11).

Fig. 4.11 The stimulation apparatus, connected with the EEG amplifiers.

A reliable P3b response was obtained in 22 cases. In single subjects, the amplitudes and latency values of the P3b obtained in VR condition did not exceed a 10 % difference from those obtained in the fixed - position condition by the standard visual oddball paradigm.

P3b Latency P3b latency was not significantly changed as an effect of the group and the different visual stimulations (Table 4.2).

P3b amplitude The MANOVA analysis showed a significant effect of group, stimulus type and for the interaction group x stimulus (MANOVA test: effect of group (DF = 1), F (Roy radix) = 2.009, DF = 57, p = 0.007; stimuli (DF = 3), F = 2.46, DF = 57, p = 0.001; groups x stimuli (DF = 4), F = 1.94, DF = 57, p = 0.009). The effect of the Virtual Environment was also significant as regards to P3b amplitude of target stimuli, but the interaction groups x VE was not significant though it closely approached significance (MANOVA group (DF = 1), F (Roy Radix) = 2.46, DF = 57, p = 0.016; VE (DF = 2), F = 2.080, DF = 57, p = 0.036; groups x VR (DF = 3), F = 1.83, DF = 57, p = 0.068 n.s.). In the Young group, the P3b amplitude did not appear clearly different in different VE conditions, while in the Old group the P3b appeared more diffused over the parietal and central derivations during the vision of target stimuli into the Aisle and Bedroom Virtual Environments (Figs. 4.12, 4.13).

However, the Bonferroni test did not show any significant differences among the different virtual context in any group. In the Young group, all target stimuli determined a significant

Table 4.2 ANOVA test: group F = 0.56 (DF = 1) n.s.; VE (DF = 2), F = 0.65 n.s.; stimuli (DF 2), F = 0.58, n.s.; group x stimuli (DF = 2), F = 0.67 n.s.; group x VE (DF = 2), F = 0.06 n.s.; group x stimulus x VE (DF = 4), F = 0.52 n.s

	Old Group		**Young Group**	
VE stimulus	*Mean (ms)*	*SD*	*Mean (ms)*	*SD*
		Aisle		
White	430.04	30.32	409.83	29.03
Red	411.46	33.52	366.86	29.03
Green	382.81	30.32	400.72	29.03
		Living Room		
White	391.02	31.80	393.23	29.03
Red	424.22	31.80	396.48	29.03
Green	391.02	31.80	402.70	30.32
		Bedroom		
White	356.58	38.01	360.03	29.03
Red	413.09	35.55	381.51	29.03
Green	362.85	33.52	351.56	33.52

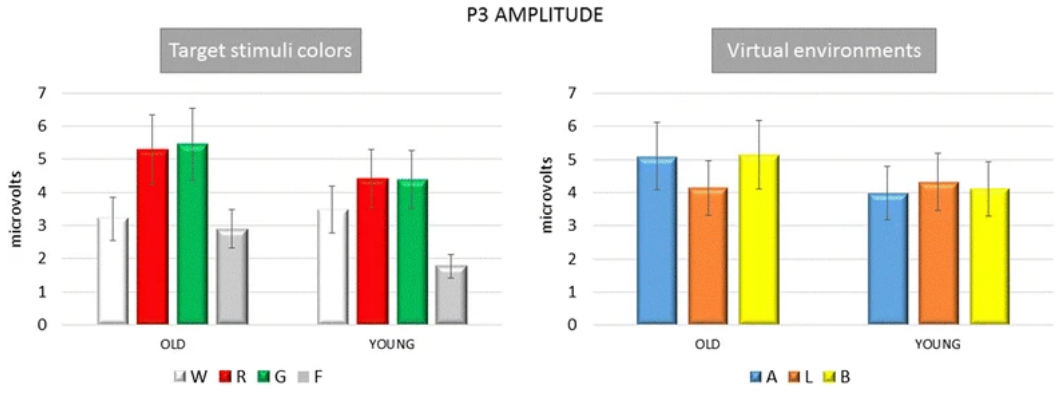


Fig. 4.12 Mean and standard errors of P3b amplitude on Pz EEG channels in the Young and Old groups under different stimulation conditions. (Right) The P3b amplitude averaged across target stimuli is reported A = aisle; L = living room; B = bedroom. (Left) The P3b amplitude obtained by target stimuli are reported in W white, R red, and G green conditions, as well as the amplitude of the positive response in the 300–600 ms time interval after F - frequent stimuli.

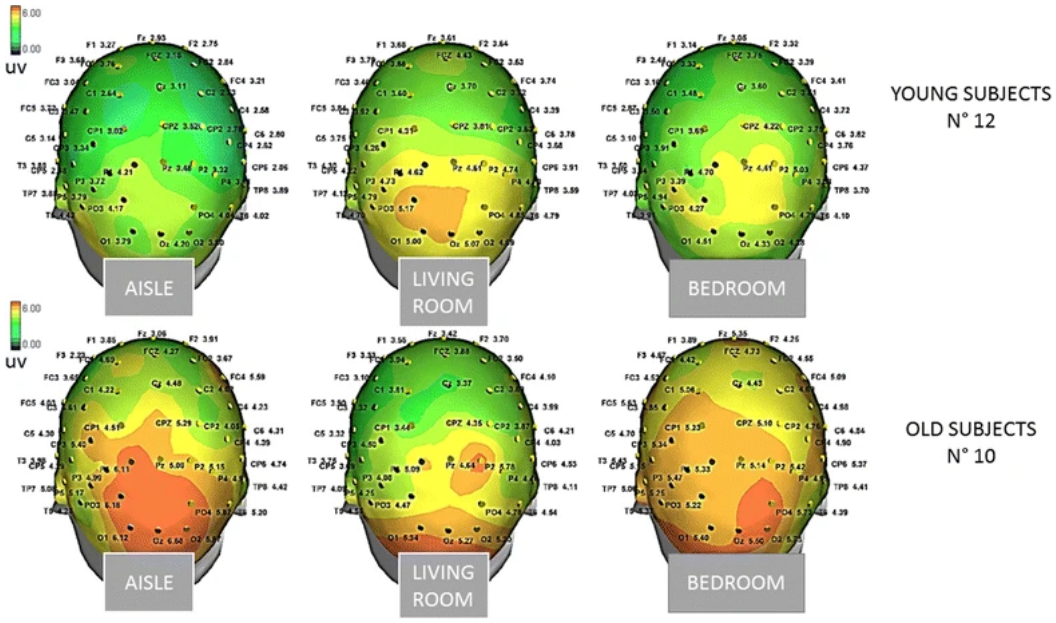


Fig. 4.13 The grand average of P3b response amplitude by target stimuli in the different virtual environments are represented by a scalp map model provided by ASA software.

increase in P3b amplitude compared to the Frequent stimuli condition, whatever was the color of the target door (Figs. 4.12, 4.14, 4.15, 4.16).

The Bonferroni test showed that the P3b was significantly increased on the parietal, occipital and central electrodes in W, R and G conditions, compared to F (Fig. 4.17). The Bonferroni test showed also that in the elderly group the P3b obtained by the green and red colours were significantly different from the frequent stimulus, on the parietal, occipital, and central derivations, while the White stimulus did not evoke a significantly larger P3b with respect to frequent stimulus (Figs. 4.12, 4.14, 4.15, 4.16, 4.17). In addition, the P3b evoked by the White door was significantly different from that obtained in the G and R conditions (W vs. G $p < 0.05$ on P3, O1, P7, T6, C7 derivations; W vs. R $p < 0.01$ on P3, O1, P7, T6, C7, CP3, < 0.01 on Pz, Cp1, Cpz derivations).

The results of this study may confirm in a preliminary way the reliability of a VR model for extracting a P3b component and evaluate the possible age-related cognitive impact of simulated architectural context changes, which are easily applicable to a real setting.

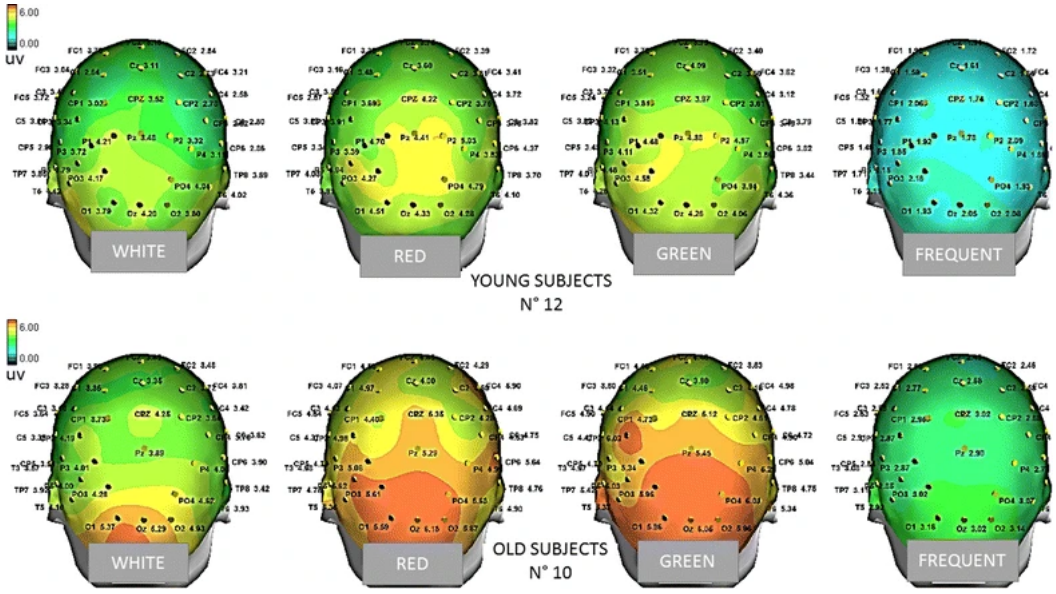


Fig. 4.14 The grand average of the P3b response amplitude of the target and frequent stimuli are represented by a scalp map model provided by ASA software.

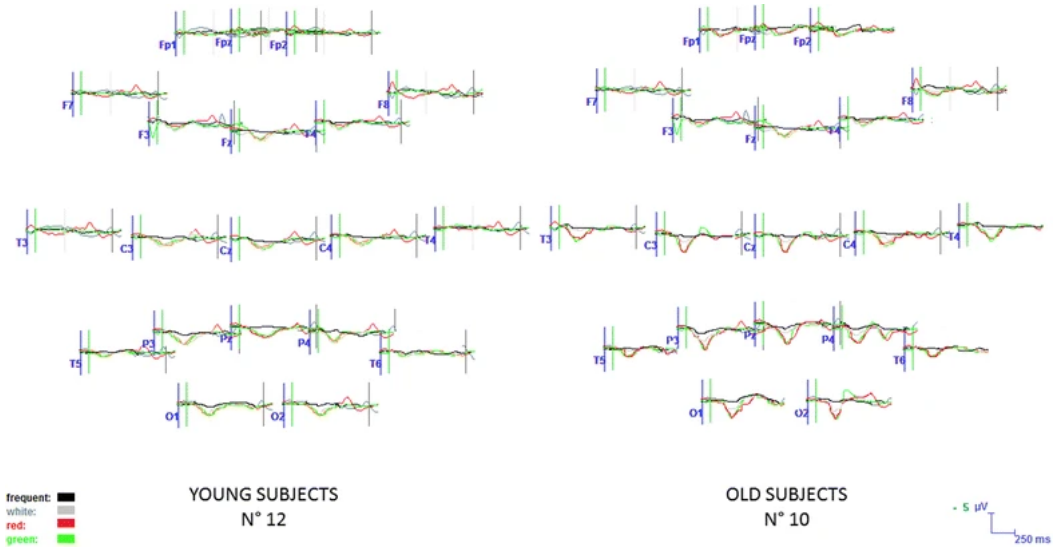


Fig. 4.15 The grand average of the P3b response obtained by the target and frequent stimuli are represented for 20 representative channels in the two groups. In the old subjects, the response by the target stimulus in the White condition, represented in gray, is hardly distinguished from the response to frequent stimuli (represented in black), while the P3b evoked by target stimuli in Red and Green conditions (represented with these colors) is clearly different from the response to the standard one. In the young subjects, the P3b evoked by all target stimuli is clearly recognizable from the response evoked by the standard one.

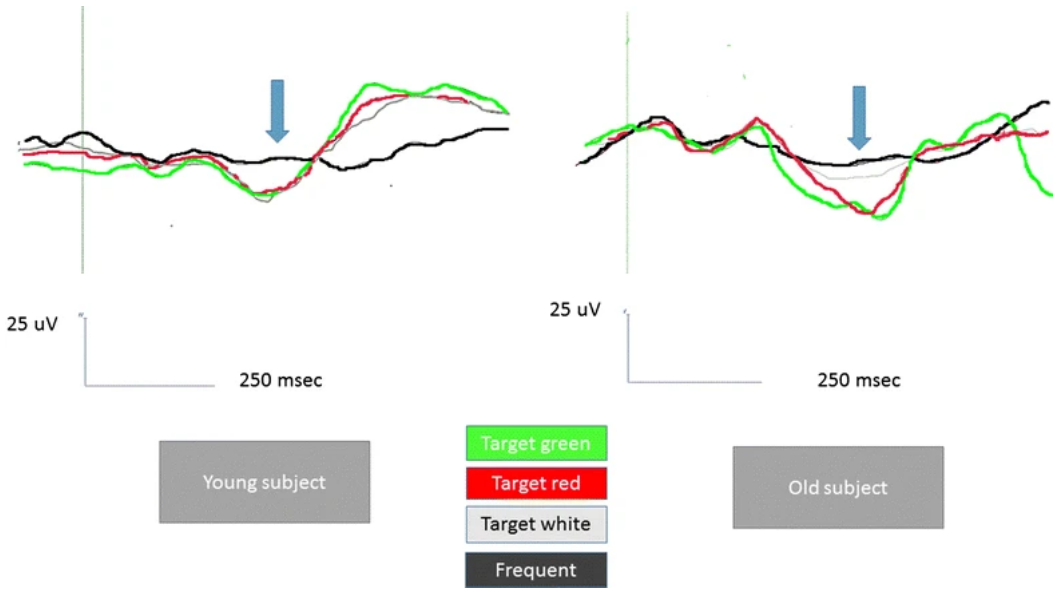


Fig. 4.16 A detailed example of the P3b responses in 2 representative subjects (Pz channel). In the old subject, male, 75 years old, the response by the target stimulus in the White condition, is reduced in amplitude with respect to the response in red and green conditions (represented with these colors). In the young subject, male, 30 years old, the target P3b waves have similar amplitude, independently of the colour of the stimulus.

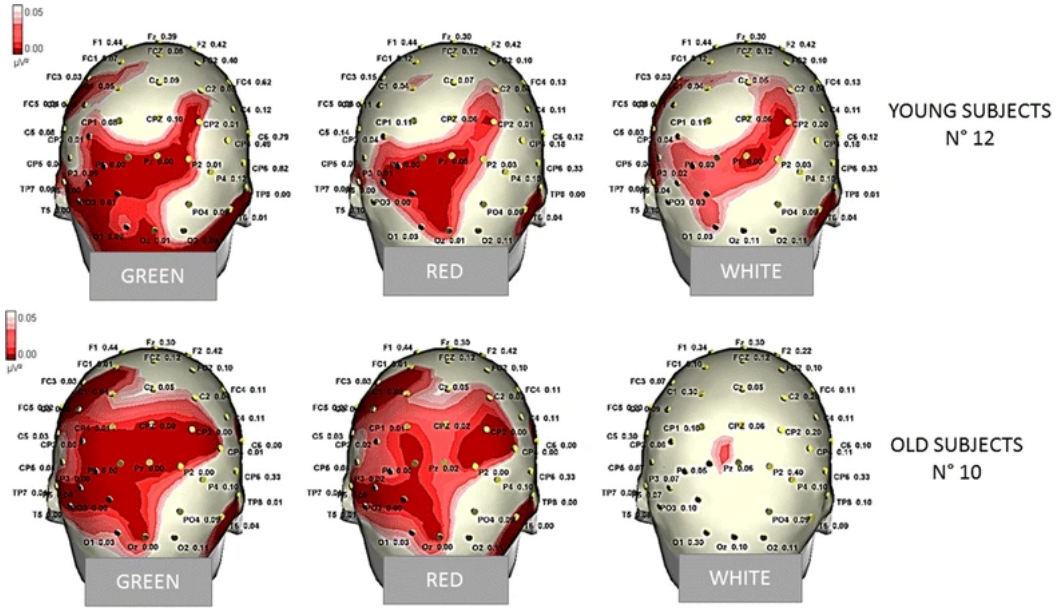


Fig. 4.17 Statistical Probability Maps representing the results of the Bonferroni test comparing the P3b amplitude by the different target stimuli with the response obtained by standard stimuli in single groups. The red shades represent the significant p-values.

4.2.4 Virtual Reality for Training Functional Living Skills on Persons with Major Neurocognitive Disorder

The innovative use of technology-based interventions for people with early or prodromal dementia is a growing area of study and offers many possibilities for improving the independence and the quality of life of people with acquired neurocognitive disorders. In the Desktop Visual Display Systems (DVDS) the person interacts with a monitor displaying 3D objects and environments, along with auditory and visual stimuli, making such environments even more similar to the real one. Using such a kind of system is simple, requires limited instrumentation and short training for new users. Only a few studies focusing on functional living skills used Virtual Training (VT) with patients affected by dementia [413–415]. Functional living skills include complex activities of daily life, such as cooking, phoning, cleaning, washing laundry, shopping, taking medicines, and using transportation. When these skills deteriorate, the patient loses his/her self-sufficiency and, consequently, his/her self-esteem and well-being. These skills require a great neuropsychological organisation and, in patients with Major NeuroCognitive Disorder (M-NCD), progressively deteriorate as a result of the cognitive decline. Therefore, the interventions for slowing down this decline are of fundamental importance.

In this context, for assessment and rehabilitation purposes, serious games in Augmented (AR) or Virtual Reality (VR) could be a good alternative to the traditional approach to interact with patients, collect statistics, recognise and classify patterns, and design unsupervised machine learning algorithms for clustering their cognitive skills and performance [186–188]. The results of earlier studies on Virtual Reality applied to the empowerment or evaluation of functional living skills are encouraging, and it appears that VT has the same effectiveness of rehabilitation approaches carried out in the natural environment. The study by Hoffmann et al. [413] showed improvements in a shopping task, with decreased errors and times; a follow-up after three weeks showed that people maintained their skills; the participants in the study also appreciated the proposed interactive modality. Van Schaick et al. [414] investigated both facilitators and barriers for a walk in a city centre, highlighting a performance improvement in this task. Foloppe et al. [415] used VR with a patient with Alzheimer's disease to relearn cooking activities, suggesting that VR can produce improvements in the same way as real-life relearning.

4.2.4.1 Research Contribution

This study aimed at verifying the feasibility of a non-immersive Virtual Training (VT) on daily living skills for patients with Major NeuroCognitive Disorder (M-NCD) and to verify the possible generalisation of improved skills repeating the same tasks in the natural environment, by administering the in vivo tests before (T1) and after (T3) the VT.

A digital system was set up with a server connecting a database to the apps installed on a touch TV. The database was developed in PostgreSQL with an interface implemented as a C# Dynamic-Link Library (DLL); the system contains a suite of apps created with Unity 3D, a cross-platform real-time game engine, often employed for developing VR scenarios for rehabilitation purposes [416]. Four apps were developed, focusing on four functional skills, and namely:

1. to provide information (or information): 30 questions in verbal and written form, including general knowledge, personal, family, spatial and temporal orientation, and with multiple-choice answers appearing on the screen in written form;

2. taking medicines (or medicines): at appropriate times, the scene presents five medicine boxes placed on a kitchen table (Fig. 4.18D); a set of instructions explains when each drug should be taken; the implemented task consists in choosing the appropriate box by touching one of the shown medicine boxes, for 10 times, as a response to 10 verbal questions, presented randomly during each session;

3. preparing the suitcase for a weekend away from home (or suitcase): a single scene, with shelves containing clothes to be placed into a suitcase on a sofa (Fig. 4.18C);

4. shopping at the supermarket following a shopping list (or supermarket): the shopping list includes five products and remains available on the screen. The first scene is a kitchen, with the shopping list, money and a wallet; the second scene includes a supermarket shelf with different products and a shopping cart (Fig. 4.18A). The third scene includes the cash counter to pay for products (Fig. 4.18B); the application also requires the user to pay the right (or enough) amount due to the cashiers, and check the possible change returned.

In addition, a fifth app, not yet included in the suite of games, has been developed:

5. Preparing a juice for two persons (or juice): a single scene with a kitchen (Fig. 4.19) showing several kinds of fruit on a shelf. When the scene starts, the user is asked to

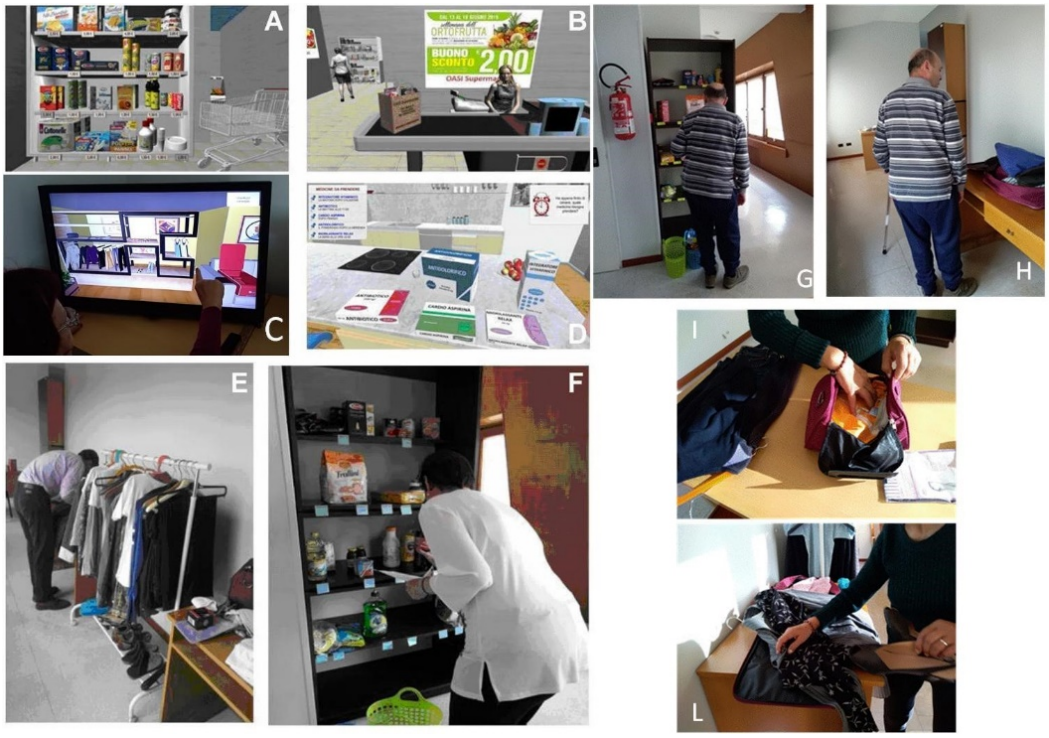


Fig. 4.18 App's scenes (A-B-C-D) and examples of the in vivo test (E-F-G-H-I-L).

choose two types of fruit, randomly selected, and must complete all the steps required for making the juice, serve it and clean the dishes.

Supermarket, suitcase and juice offered the possibility of a video demonstration before starting the task. The apps were developed based on some principles and procedures of Applied Behaviour Analysis, including:

- verbal reinforcement after the correct response (i.e., good!, ok!, congratulations!, well done!);
- correction after the wrong response, consisting of question repetition and use of the least-to-most prompting (up to a maximum of three prompts), i.e., the prompting system starts with the least intrusive prompt, then suggest to the next with major assistance;
- task analysis and total task chaining for the supermarket and suitcase apps, where total task chaining is completing the entire task following the steps which is composed of in the right order.

The first instruction, concerning how to perform the task, was given simultaneously in a verbal and written manner. The application of the procedures previously reported is

Fig. 4.19 Scene from Preparing a juice app.

fundamental for encouraging and stimulating people to perform the rehabilitative sessions in the desired way over time.

The first aim of this study was to verify the feasibility of a non-immersive virtual training (VT) on functional living skills (providing information, taking medicines, preparing a suitcase, and shopping to the supermarket) applied to patients with mild to moderate M-NCD. With reference to the VT sessions administered to EG, results showed improvements in the number of correct responses and a decrease in the number of missed responses, as well as prompts and total execution times. In the in vivo test, Experimental Group (EG) also showed decreased execution times and increased correct responses, and statistically significant differences were found in comparison with the Control Group (CG). This last result seems to suggest the need for specific training to improve the functional skills of everyday life; indeed, the CG, despite having benefited from a group cognitive stimulation training, did not improve performances in the in vivo tests at T3 compared to T1.

The second aim was to verify the eventual generalisation in the natural environment of the skills improved during the VT. Results reported in Table 4.3 showed improvements in three of the four skills performed in the natural environment. This generalisation was spontaneous since it took place without specific naturalistic training. This result encourages the continuation of the study since it will be possible to recover many different functional living skills, by developing various and appropriate apps, and then administering the VT. This new technological way for recovering and maintaining daily living skills also appears to

Table 4.3 Statistical significance of the differences between T1 and T3 in vivo test.

	EG	**CG**	**p=**
Information, N of correct answers, median (interquartile range)	1.5 (1/2.75)	-1 (-1/1)	ns
Information, total time execution in seconds, median (interquartile range)	-54 (-58.75 /-40.25)	-18 (-62/-4)	ns
Suitcase, N of correct answers, median (interquartile range)	4 (1.75/4.75)	0 (-2/0)	0.008
Suitcase, total time execution in seconds, median (interquartile range)	-0.5 (-2075/28.75)	-92 (-212/-13)	ns
Medicines, N of correct answers, median (interquartile range)	2 (2/3)	0 (0/0)	0.032
Medicines, total time execution in seconds, median (interquartile range)	-57 (-71/-41)	37 (26-58)	0.008
Supermarket, N of correct answers, median (interquartile range)	2.5 (1.25/3.75)	0 (-1/1)	0.05
Supermarket, total time execution in seconds, median (interquartile range)	-84.5 (-186.5/11.5)	-55 (-92/-2)	ns

be able to reduce rehabilitation costs of the healthcare service and might be used as a remote tele-rehabilitation tool, with decreased rehabilitation costs and saving time for patients and their caregivers.

4.3 Physiatry

Parkinson's Disease (PD) is one of the most widely spread neurodegenerative disorders. In detail, it is a degenerative brain disorder characterized by a loss of midbrain dopamine (DA) neurons [417]. The main clinical PD symptoms involving body movements are: tremor, rigidity, bradykinesia, and gait abnormalities. Commonly physicians' evaluations are based on historical information from the patient, regarding the motor function during activities of daily living and clinical observations, using a clinical rating scale such as the Unified PD Rating Scale (UPDRS) [418] and Hoehn and Yahr staging scale [419]. The UPDRS scale is a numeric scale widely used to assess PD severity. Even if nowadays there are several scientific results supporting its validity, subjectivity and low efficiency are inevitable, as most of the diagnostic criteria use descriptive symptoms, which cannot provide a quantified diagnostic basis. In particular, the main problems regard the evaluation of the severity of specific symptoms such as freezing of gait [420–422], dysarthria [423], tremor [424–427],

bradykinesia [428–430] and dyskinesia [431–435].Therefore, the development of computer-assisted diagnosis and computer expert system is very important.

Another interesting research field focuses on the analysis of different common life tasks such as handwriting. Handwriting is a highly over-learned fine and complex manual skill involving an intricate blend of cognitive, sensory and perceptual-motor components [436]. For these reasons, the presence of abnormalities in the handwriting process is a well-known and well-recognized manifestation of a wide variety of neuromotor diseases. There are two main difficulties related to handwriting and affecting Parkinson's Disease (PD) patients: (i) the difficulty in controlling the amplitude of the movement, i.e., decreased letter size (micrography) and failing in maintaining stroke width of the characters as writing progresses [437], and (ii) the irregular and bradykinetic movements, i.e., increased movement time, decreased velocities and accelerations, and irregular velocity and acceleration trends over time [438]. For these reasons, in literature, there are several works investigating the possibility of a differentiation between PD patients and healthy subjects by means of computer-aided handwriting analysis tools.

4.3.1 Machine Learning for Assessing the Progression of Parkinson's Disease

The main goal of this study was to provide a toolbox to support clinicians in the objective assessment of the typical PD motor issues and alterations. The assessment has been done by means of an overall integration of different information sources (i.e., the integration of different features coming from different tasks). Analyses were performed focusing on the independent analysis of three main source of information: two common tasks usually used by the physician to evaluate the motor impairment of the patients and a novel approach based on handwriting analysis. As a result, an objective description of the patients status can be made and classified, so that the patients can be monitored during the disease progress. The main novel contributions with respect to the state of the art is the design, the development and the evaluation with both healthy subjects and PD patients of two systems: a vision-based system able to capture specific movements of different main UPDRS scale exercises and a handwriting analysis tool able to extract biometric signals related to pen movements and to muscular activity.

Furthermore, several classifiers able to assess and rate the movement impairment of PD patients were developed and compared using a specific set of features extracted from the previous system set-up.

Fig. 4.20 The left image shows a healthy subject wearing the two passive finger markers. The images reported on the right show the foot of a subject doing the foot tapping exercise while he is wearing a passive marker on the toes.

4.3.1.1 Research Contribution

This research work shows two parallel analyses: a first one focused on the analysis of the motor tasks according to UPDRS scale and a second one base on handwriting analysis.

Regarding the assessment of the UPDRS scale tasks, a system able to reproduce and record two tasks was designed and developed; the tasks were: Finger Tapping and Foot Tapping. Two separate vision-based systems able to acquire the movement of the thumb, the index finger and the toes were also designed and developed. Both acquisition systems were based on passive markers made of reflective material and the Microsoft Kinect One RGBD camera. A brief and detailed description of two acquisition systems follows (Fig. 4.20):

- Finger Tapping exercise set-up (Fig. 4.20 left): this test considers the examination of both hands separately. The tested subject is seated in front of the camera and is instructed to tap the index finger on the thumb ten times as quickly and as widely as possible. During the task the subject wears two thimbles made of a reflective material on both the index finger and the thumb.

- Foot Tapping exercise set-up (Fig. 4.20 right): the feet are tested separately. The tested subject sits in a straight-backed chair in front of the camera and has both feet on the floor. He is then instructed to place the heel on the ground in a comfortable position and then tap the toes ten times as widely and as fast as possible. A system of stripes with a reflective material are positioned on the toes.

For the handwriting analysis, instead, the system set-up for data acquisition includes two main sensors: (i) the Myo Gesture Control Armband that allows one to synchronously acquire eight different sEMG sources of the forearm, and (ii) the WACOM Cintiq 13” HD,

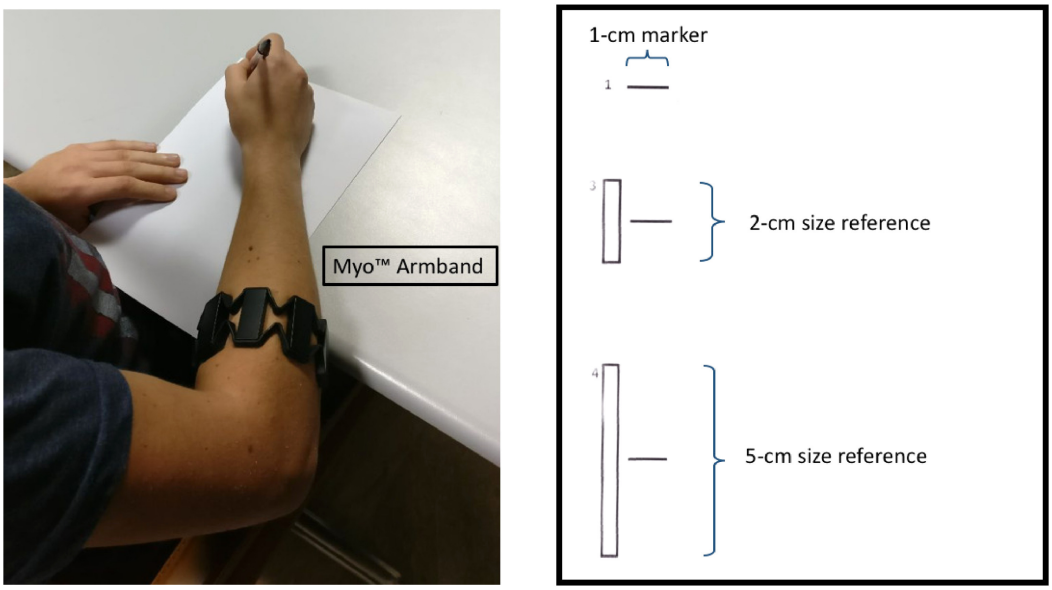


Fig. 4.21 The system setup used for the experimental tests to validate the proposed approach (left). The paper sheet template, replicated on the tablet with two (vertical) visual marker size references of 2 cm and 5 cm, respectively, for writing WPs 2 and 3, and three (horizontal) 1 cm markers used for spatial mapping needed to extract the execution average linear speed feature for all three writing tasks.

a graphics tablet providing visual feedback for acquiring pen tip planar coordinates and pressure, and the tilt of the pen with respect to the writing surface.

For the experiments, whose setup is represented in Fig. 4.21, we used three writing patterns (WPs) leading to as many writing tasks, these are: a five-turn spiral drawn in anticlockwise direction (WP 1), a sequence of eight Latin letters "l" with a size of 2.5 cm (WP 2) and with a size of 5 cm (WP 3). Since the last two WPs were size-constrained, a visual marker was provided as reference. In the experiment, we asked each subject to perform the three writing tasks four times each for a total of twelve tasks: the first time for familiarisation purposes, whereas the other three outcomes were acquired and stored for subsequent feature extraction and processing. An example of the writing tasks executions is reported in Fig. 4.22 for both an healthy subject and a PD patient.

The subject was also asked to rest between two subsequent handwriting tasks for at least three seconds. The beginning of the task signal acquisition was triggered by a positive pen pressure applied on the graphic tablet. The processing of the acquired raw signals led to the extraction of several features.

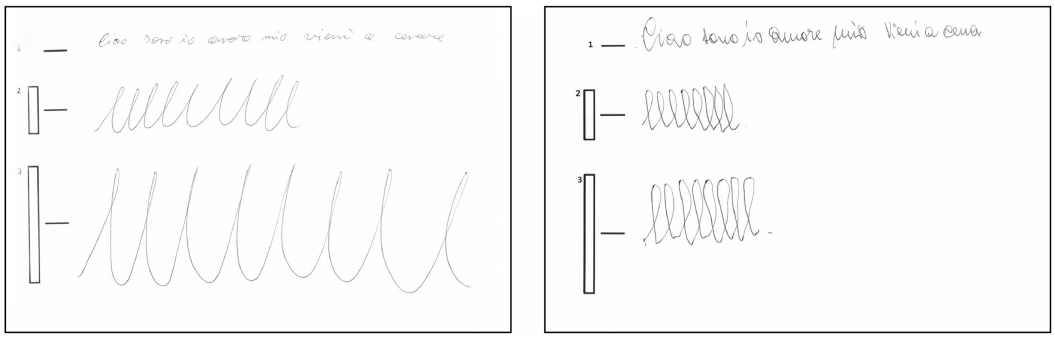


Fig. 4.22 Two samples of a repetition of all three writing tasks respectively performed by a healthy subject (left) and by a PD patient (right).

Also for the handwriting analysis, starting from the biometric signals acquired during the handwriting tasks, several features were extracted. As a result, three different feature datasets were created:

- the first dataset including only the features extracted from writing pattern 1 (41 features);
- the second dataset including only the features extracted from writing pattern 2 (43 features);
- the third dataset including only the features extracted from writing pattern 3 (43 features);

The first part of the study investigated on the capabilities of discrimination by means of two different machine learning approaches, one for each study, in order to classify PD patients from Healthy subjects. For the motor tasks, three different binary support vector machine (SVM) classifiers were trained, based on the following three set of features: (i) all the finger tapping features; (ii) all the foot tapping features; (iii) both finger and foot tapping features. The handwriting features, instead, have been used to train a classifier based on an Artificial Neural Network (ANN) whose optimal topology was found using a Single-Objective Genetic Algorithm (GA) [108].

In the second part of the study the "Mild PD patients vs Moderate PD patients" classification was performed. As for the previous approach, two different classification approaches, one for each reported study, were investigated.

Regarding the classification of Healthy subjects vs PD patients, for the UPDRS based tasks, the following results were obtained:

- Finger tapping features set: the best classifier was the Gaussian SVM with an accuracy of 71.0 %, a sensitivity of 75.7 % and a specificity of 65.5 %.
- Foot tapping features set: the best classifier was the Gaussian SVM with an accuracy of 85.5 %, a sensitivity of 91.0 % and a specificity of 79.0 %.
- Both finger and foot tapping feature sets: the best classifier was the Quadratic SVM with an accuracy of 87.1 %, a sensitivity of 87.8 % and a specificity of 86.0 %.

For the handwriting datasets, instead:

- *Dataset with all 41 features extracted from writing pattern 1*: the optimal topology for the ANN featured 2 hidden layers (respectively composed of 186 and 15 neurons), and 2 neurons for the output layer. The activation function found by the GA was logsig for the hidden layers. For the output layer, the softmax function was preliminarily selected as activation function. Accuracy: 90.76 % (std = 0.0764), specificity: 0.8530 (std = 0.1553), sensitivity: 0.9389 (std = 0.0720).
- *Dataset with all 43 features extracted from writing pattern 2*: the optimal topology for the ANN featured 2 hidden layers (respectively composed of 44 and 10 neurons), and 2 neurons for the output layer. The activation function found by the GA was logsig for the hidden layers. For the output layer, the softmax function was preliminarily selected as activation function. Accuracy: 92.98 % (std = 0.0523), specificity: 0.8970 (std = 0.1212), sensitivity: 0.9486 (std = 0.0587).
- *Dataset with all 43 features extracted from writing pattern 3*: the optimal topology for the ANN featured 3 hidden layers (respectively composed of 232, 82 and 7 neurons), and 2 neurons for the output layer. The activation function found by the GA was logsig for the hidden layers. For the output layer, the softmax function was preliminarily selected as activation function. Accuracy: 95.95 % (std = 0.0479), specificity: 0.9425 (std = 0.0831), sensitivity: 0.9691 (std = 0.0575).

Concerning with the Mild PD patients vs Moderate PD patients classification, for the UPDRS based tasks the following results were obtained:

- Finger tapping features set: Gaussian SVM classifiers were not able to adequately discriminate between the two classes.
- Foot tapping features set: the best classifier was the Gaussian SVM with accuracy of 81.0 %, sensitivity of 84.0 % and specificity of 78.0 %.

- Both finger and foot tapping feature sets: the best classifier was the Gaussian SVM with accuracy of 78.0 %, sensitivity of 89.0 % and specificity of 64.0 %.

For the handwriting datasets, instead, the results were:

- *Dataset with all 41 features extracted from writing pattern 1*: the optimal topology for the ANN featured 3 hidden layers (respectively composed of 59, 65 and 2 neurons), and 2 neurons for the output layer. The activation function found by the GA was logsig for the hidden layers. For the output layer, the softmax function was preliminarily selected as activation function. Accuracy: 94.34 % (std = 0.0626), specificity: 0.9595 (std = 0.0763), sensitivity: 0.9220 (std = 0.1158).

- *Dataset with all 43 features extracted from writing pattern 2*: the optimal topology for the ANN featured 3 hidden layers (respectively composed of 133, 18 and 1 neurons), and 2 neurons for the output layer. The activation function found by the GA was logsig for the hidden layers. For the output layer, the softmax function was preliminarily selected as activation function. Accuracy: 87.26 % (std = 0.0850), specificity: 0.8720 (std = 0.1206), sensitivity: 0.8733 (std = 0.1544).

- *Dataset with all 43 features extracted from writing pattern 3*: the optimal topology for the ANN featured 3 hidden layers (respectively composed of 65, 36 and 7 neurons), and 2 neurons for the output layer. The activation function found by the GA was logsig for the hidden layers. For the output layer, the softmax function was preliminarily selected as activation function. Accuracy: 91.86 % (std = 0.0830), specificity: 0.9169 (std = 0.1167), sensitivity: 0.9220 (std = 0.1286).

The results of the classification between mild and moderate PD patients indicated that the foot tapping features are more representative than the finger tapping ones. In fact, the SVM classifier based on the foot tapping features reached the best score in terms of accuracy and specificity. The results obtained from the handwriting analysis, instead, showed that healthy subjects can be differentiated from PD patients and that mild PD patients can be differentiated from moderate PD patients both with high classification accuracy (over 90 %).

According to the results obtained, both approaches could lead to a comprehensive tool suite for evaluating and grading Parkinson's disease which is, at the same time, able to synthesise different aspects of subject symptoms by means of the analysis of different sources of information.

Chapter 5

Conclusion

The studies conducted in this Ph.D. work aimed to design, develop and evaluate intelligent systems to support clinical diagnosis through a quantitative and objective assessment of pathologies, ranging from diagnosis in radiology to rehabilitation support using innovative devices.

After the introduction highlighting the importance of having an objective evaluation of pathological conditions with the support of automatic systems, the methodological approaches and technologies concerning with Computer Aided Diagnosis systems have been introduced in Chapter 2. The innovative methodological approach based on Deep Learning algorithms for both classification and segmentation has also been described.

All studies have been conducted in the context of precision medicine, an ongoing scenario allowing doctors and researchers to perform diagnosis more accurately than in the past, also allowing clinicians to prescribe drugs, therapies and rehabilitation treatments based on the specific genetic makeup of the subjects, or based on the specific mutation of the pathology affecting the patients. As already reported in Chapter 1, this new approach is in contrast to a one-size-fits-all, in which disease treatment and prevention strategies are developed for the average person, with less consideration for the differences between individuals.

In fact, all the frameworks designed and implemented during this Ph.D. aimed to detect and extract pattern of features which allowed to classify, or segment, biomedical signals for supporting a clinical diagnosis otherwise obtainable with more invasive procedures. Thus, such biomarkers, if integrated with the -omic data, in addition to make the clinical decision, they also could specify the pathology based on specific patient characteristic, guiding the possible treatment, assessment and monitoring over time.

The other side of the coin, however, is not only related to clinical aspects but includes the healthcare costs which will increase with precision medicine, including management,

technology and development costs. Developing targeted drugs is likely to be expensive and not accessible to everybody. In addition, much research is still needed to reach the performance required for commercialising such drugs, but also to introduce the automatic decision support systems in the clinical practice.

In order to overcome the previous limitations and, at the same time, continue the research towards precision medicine, developing Computer Aided Diagnosis systems for making quantitative assessments or objective evaluations of investigated diseases could help physicians in moving toward the precision medicine. In fact, obtaining the same (or very close) levels of accuracy in discriminating pathologies as the genetic approach, i.e., classifying the mutation of a tumour based on image analysis, will certainly reduce the costs of genetic tests, as well as reduce the time for formulating a diagnosis.

Research in the field of Radiomics, an emerging translational field of research aiming to extract mineable high-dimensional data from clinical images, is going toward precision medicine. However, this should not be limited to diagnosis, but future research in this sense should include precision medicine features also for prognosis, therapy and rehabilitation.

My Publications

In the following, it is reported a complete list of the works published during the Ph.D.

1. Eleonora Gentile, Antonio Brunetti, Katia Ricci, Marianna Delussi, Vitoantonio Bevilacqua and Marina de Tommaso. Mutual interaction between motor cortex activation and pain in fibromyalgia: EEG-fNIRS study. *PLOS ONE* 15(1): e0228158. doi:10.1371/journal.pone.0228158

2. Gianpaolo Francesco Trotta, Roberta Pellicciari, Antonio Boccaccio, Antonio Brunetti, Giacomo Donato Cascarano, Vito Modesto Manghisi, Michele Fiorentino, Antonio Emmanuele Uva, Giovanni Defazio and Vitoantonio Bevilacqua. A neural network-based software to recognise blepharospasm symptoms and to measure eye closure time. *Computers in Biology and Medicine*, Volume 112 (2019), 103376. doi:10.1016/j.compbiomed.2019.103376

3. Leonarda Carnimeo, Gianpaolo Francesco Trotta, Antonio Brunetti, Giacomo Donato Cascarano, Domenico Buongiorno, Claudio Loconsole, Eugenio Di Sciascio and Vitoantonio Bevilacqua. A Proposal of a Healthcare Network based on Big Data Analytics for Parkinson's diseases. *Journal of Engineering*, Volume 2019, Issue 6., pp. 4603 – 4611. doi:10.1049/joe.2018.5142

4. Claudio Loconsole, Giacomo Donato Cascarano, Antonio Brunetti, Gianpaolo Francesco Trotta, Giacomo Losavio, Vitoantonio Bevilacqua, and Eugenio Di Sciascio. A model-free technique based on computer vision and sEMG for classification in Parkinson's disease by using computer-assisted handwriting analysis. *Pattern Recognition Letters*, Volume 121, 2019, Pages: 28-36, doi:10.1016/j.patrec.2018.04.006

5. Antonio Brunetti, Leonarda Carnimeo, Gianpaolo Francesco Trotta, and Vitoantonio Bevilacqua. Computer-Assisted Frameworks for Classification of Liver, Breast and Blood Neoplasias via Neural Networks: a Survey based on Medical Images. *Neurocomputing*, Volume 335, 2019, Pages: 274-298, doi:10.1016/j.neucom.2018.06.080

6. Domenico Buongiorno, Ilaria Bortone, Giacomo Donato Cascarano, Gianpaolo Francesco Trotta, Antonio Brunetti, and Vitoantonio Bevilacqua. A low-cost vision system based on the analysis of motor features for recognition and severity rating of Parkinson's Disease. *BMC Medical Informatics and Decision Making*, Volume 19, 243 (2019), doi:10.1186/s12911-019-0987-5

7. Vitoantonio Bevilacqua, Antonio Brunetti, Andrea Guerriero, Gianpaolo Francesco Trotta, Michele Telegrafo, and Marco Moschetta. A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images. *Cognitive Systems Research* 53 (2019) pp: 3–19 doi:10.1016/j.cogsys.2018.04.011

8. Vitoantonio Bevilacqua, Antonio Brunetti, Giacomo Donato Cascarano, Andrea Guerriero, Francesco Pesce, Marco Moschetta, and Loreto Gesualdo. A comparison between two semantic deep learning frameworks for the autosomal dominant polycystic kidney disease segmentation based on magnetic resonance images. *BMC Medical Informatics and Decision Making*, Volume 19, 244 (2019), doi:10.1186/s12911-019-0988-4

9. Giacomo Donato Cascarano, Claudio Loconsole, Antonio Brunetti, Antonio Lattarulo, Domenico Buongiorno, Giacomo Losavio, Eugenio Di Sciascio, and Vitoantonio Bevilacqua. Biometric handwriting analysis to support Parkinson's Disease assessment and grading. *BMC Medical Informatics and Decision Making*, Volume 19, 252 (2019), doi:10.1186/s12911-019-0989-3

10. Antonio Brunetti, Domenico Buongiorno, Gianpaolo Francesco Trotta, and Vitoantonio Bevilacqua. Computer Vision and Deep Learning Techniques for Pedestrian Detection and Tracking: A Survey. *Neurocomputing*, Volume 300, 26 July 2018, Pages: 17-33, doi: 10.1016/j.neucom.2018.01.092

11. Sara Invitto, Antonio Calcagnì, Arianna Mignozzi, Rosanna Scardino, Giulia Piraino, Daniele Turchi, Irio De Feudis, Antonio Brunetti, Vitoantonio Bevilacqua, and Marina de Tommaso. Face Recognition, Musical Appraisal, and Emotional Crossmodal Bias. F*rontiers in Behavioural Sciences*, Volume 11 (144), 2 August 2017, ISSN 1662-5153, doi: 10.3389/fnbeh.2017.00144

12. Vitoantonio Bevilacqua, Antonio Brunetti, Gianpaolo Francesco Trotta, Leonarda Carnimeo, Francescomaria Marino, Vito Alberotanza, Arnaldo Scardapane. A Deep Learning Approach for Hepatocellular Carcinoma Grading. *International Journal of Computer Vision and Image Processing* (IJCVIP), Volume 7 (2), Pages 1-18, doi:10.4018/IJCVIP.2017040101

13. Vitoantonio Bevilacqua, Nicola Pietroleonardo, Vito Triggiani, Antonio Brunetti, Anna Maria Di Palma, Michele Rossini, and Loreto Gesualdo. An innovative neural network framework to classify blood vessels and tubules based on Haralick features evaluated in histological images of kidney biopsy. *Neurocomputing*, Volume 228, 8 March 2017, Pages 143-153, ISSN 0925-2312, doi:10.1016/j.neucom.2016.09.091

14. Marina de Tommaso, Katia Ricci, Marianna Delussi, Anna Montemurno, Eleonora Vecchio, Antonio Brunetti, and Vitoantonio Bevilacqua. Testing a novel method for improving wayfinding by means of a P3b Virtual Reality Visual Paradigm in normal aging. *SpringerPlus*, (2016), volume 5 (1), art no.1297. doi:10.1186/s40064-016-2978-7

15. Antonio Ivano Triggiani, Vitoantonio Bevilacqua, Antonio Brunetti, Roberta Lizio, Giacomo Tattoli, Fabio Cassano, Andrea Soricelli, Raffaele Ferri, Flavio Nobili, Loreto Gesualdo, Maria Rosaria Barulli, Rosanna Tortelli, Valentina Cardinali, Antonio Giannini, Pantaleo Spagnolo, Silvia Armenise, Grazia Buenza, Gaetano Scianatico, Giancarlo Logroscino, Giordano Lacidogna, Francesco Orzi, Carla Buttinelli, Franco Giubilei, Claudio Del Percio, Giovanni B. Frisoni, Claudio Babiloni. Classification of healthy subjects and Alzheimer's disease patients with dementia from cortical sources of resting state EEG rhythms: A study using artificial neural networks. *Frontiers in Neuroscience*, Volume 10 (604), January 2017, ISSN 1662-453X, doi:10.3389/fnins.2016.00604

16. Antonio Brunetti, Giacomo Donato Cascarano, Irio De Feudis, Marco Moschetta, Loreto Gesualdo, and Vitoantonio Bevilacqua. Detection and Segmentation of Kidneys from Magnetic Resonance Images in Patients with Autosomal Dominant Polycystic Kidney Disease. In: Huang DS., Jo KH., Huang ZK. (eds) *Intelligent Computing Theories and Application. ICIC 2019. Lecture Notes in Computer Science*, vol 11644. Springer, Cham. doi: 10.1007/978-3-030-26969-2_60

17. Simonetta Panerai, Valentina Catania, Francesco Rundo, Vitoantonio Bevilacqua, Antonio Brunetti, Claudio De Meo, Donatella Gelardi, Claudio Babiloni, and Raffaele Ferri. Feasibility of a non-immersive virtual reality training on functional living skills applied to person with major neurocognitive disorder. In: Huang DS., Huang ZK., Hussain A. (eds) *Intelligent Computing Methodologies. ICIC 2019. Lecture Notes in Computer Science*, vol 11645. Springer, Cham. doi. 10.1007/978-3-030-26766-7_63

18. Giacomo Donato Cascarano, Francesco Saverio Debitonto, Ruggero Lemma, Antonio Brunetti, Domenico Buongiorno, Irio De Feudis, Andrea Guerriero, Michele Rossini, Francesco Pesce, Loreto Gesualdo, Vitoantonio Bevilacqua. An Innovative Neural Network Framework For Glomerulus Classification Based On Morphological And Texture Features Evaluated In Histological Images Of Kidney Biopsy. In: Huang DS., Huang ZK., Hussain A. (eds) *Intelligent Computing Methodologies. ICIC 2019. Lecture Notes in Computer Science*, vol 11645. Springer, Cham. doi:10.1007/978-3-030-26766-7_66

19. Domenico Buongiorno, Giacomo Donato Cascarano, Antonio Brunetti, Irio De Feudis, Vitoantonio Bevilacqua. A Survey on Deep Learning in Electromyographic Signal Analysis. In: Huang DS., Huang ZK., Hussain A. (eds) *Intelligent Computing Methodologies. ICIC 2019. Lecture Notes in Computer Science*, vol 11645. Springer, Cham. doi: 10.1007/978-3-030-26766-7_68

20. Giacomo Donato Cascarano, Antonio Brunetti, Domenico Buongiorno, Gianpaolo Francesco Trotta, Claudio Loconsole, Ilaria Bortone, and Vitoantonio Bevilacqua. A Multi-modal Tool Suite for Parkinson Evaluation and Grading. In: Esposito A., Faundez-Zanuy M., Morabito F., Pasero E. (eds) N*eural Approaches to Dynamics of Signal Exchanges. Smart Innovation, Systems and Technologies*, vol 151. Springer, Singapore. doi: 10.1007/978-981-13-8950-4_24

21. Gianpaolo Francesco Trotta, Sergio Mazzola, Giuseppe Gelardi, Antonio Brunetti, Nicola Marino, and Vitoantonio Bevilacqua. Reconstruction, Optimisation and Quality Check of Microsoft HoloLens-acquired 3D Points Clouds. In: Esposito A., Faundez-Zanuy M., Morabito F., Pasero E. (eds) *Neural Approaches to Dynamics of Signal Exchanges. Smart Innovation, Systems and Technologies*, vol 151. Springer, Singapore. doi: 10.1007/978-981-13-8950-4_9

22. Marina de Tommaso, Eleonora Gentile, Katia Ricci, Anna Montemurno, Marianna Delussi, Eleonora Vecchio, Giancarlo Logroscino, Antonio Brunetti, and Vitoantonio Bevilacqua. Bioelectrical Correlates of Emotional Changes Induced by Environmental Sound and Colour: from Virtual Reality to Real Life. In: Masia L., Micera S., Akay M., Pons J. (eds) *Converging Clinical and Engineering Research on Neurorehabilitation III. Proceedings of the 4th International Conference on NeuroRehabilitation* (ICNR2018) ICNR 2018. Biosystems & Biorobotics, vol 21. Springer, Cham, 2019. doi: 10.1007/978-3-030-01845-0_197

23. Claudio Del Percio, Vitoantonio Bevilacqua, Antonio Brunetti, Roberta Lizio, Andrea Soricelli, Roberto Ferri, Flavio Nobili, Loreto Gesualdo, Giancarlo Logroscino, Marina De Tommaso, Antonio Ivano Triggiani, Giovanni Battista Frisoni, and Claudio Babiloni. Classification of Healthy Subjects and Alzheimer's Disease Patients with Dementia from Cortical Sources of Resting State EEG Rhythms: Comparing Different Approaches. In: Masia L., Micera S., Akay M., Pons J. (eds) C*onverging Clinical and Engineering Research on Neurorehabilitation III. Proceedings of the 4th International Conference on NeuroRehabilitation* (ICNR2018) ICNR 2018. Biosystems & Biorobotics, vol 21. Springer, Cham, 2019. doi:10.1007/978-3-030-01845-0_196

24. Vitoantonio Bevilacqua, Claudio Loconsole, Antonio Brunetti, Giacomo Donato Cascarano, Antonio Lattarulo, Giacomo Losavio, and Eugenio Di Sciascio. A model-free computer-assisted handwriting analysis exploiting optimal topology ANNs on biometric signals in Parkinson's disease research. In: Huang, D., Jo, K., Zhang, X. (eds.) *Intelligent Computing Theories and Application. ICIC 2018. Lecture Notes in Computer Science*, vol. 10955, pp. 650-655. Springer, Cham, 2018. doi:10.1007/978-3-319-95933-7_74

25. Vitoantonio Bevilacqua, Antonio Brunetti, Giacomo Donato Cascarano, Flavio Palmieri, Andrea Guerriero, and Marco Moschetta. A Deep Learning Approach for the Automatic Detection and Segmentation in Autosomal Dominant Polycystic Kidney Disease based on Magnetic Resonance Images. In: Huang, D., Jo, K., Zhang, X. (eds.) *Intelligent Computing Theories and Application. ICIC 2018. Lecture Notes in Computer Science*, vol. 10955, pp. 643-649. Springer, Cham, 2018. doi:10.1007/978-3-319-95933-7_73

26. Vitoantonio Bevilacqua, Antonio Brunetti, Gianpaolo Francesco Trotta, Domenico De Marco, Marco Giuseppe Quercia, Domenico Buongiorno, Alessia D'Introno, Francesco Girardi, and Attilio Guarini. A Novel Deep Learning Approach in Haematology for Classification of Leucocytes. In: Esposito A., Faundez-Zanuy M., Morabito F., Pasero E. (eds) *Quantifying and Processing Biomedical and Behavioral Signals. WIRN 2017. Smart Innovation, Systems and Technologies*, vol 103. Springer, Cham., 2018 doi:10.1007/978-3-319-95095-2_25

27. Ilaria Bortone, Gianpaolo Francesco Trotta, Giacomo Donato Cascarano, Paola Regina, Antonio Brunetti, Irio De Feudis, Domenico Buongiorno, Claudio Loconsole, and Vitoantonio Bevilacqua. A Supervised Approach to Classify the Status of Bone Mineral Density in Post-Menopausal Women through Static and Dynamic Baropodometry. In: *Proceedings of the 2018 International Joint Conference on Neural Networks* (IJCNN). IEEE, 2018 doi:10.1109/IJCNN.2018.8489205

28. Claudio Loconsole, Giacomo Donato Cascarano, Antonio Lattarulo, Antonio Brunetti, Gianpaolo Francesco Trotta, Domenico Buongiorno, Ilaria Bortone, Irio De Feudis, Giacomo Losavio, Vitoantonio Bevilacqua, and Eugenio Di Sciascio A comparison between ANN and SVM classifiers for Parkinson's disease by using a model-free computer-assisted handwriting analysis based on biometric signals. In: *Proceedings of the 2018 International Joint Conference on Neural Networks* (IJCNN). IEEE, 2018 doi:10.1109/IJCNN.2018.8489293

29. Vitoantonio Bevilacqua, Gianpaolo Francesco Trotta, Antonio Brunetti, Nicholas Caporusso, Claudio Loconsole, Giacomo Donato Cascarano, Francesco Catino, Pantaleo Cozzoli, Giancarlo Delfine, Adriano Mastronardi, Andrea Di Candia, Giuseppina Lelli, and Pietro Fiore. A Comprehensive Approach for Physical Rehabilitation Assessment in Multiple Sclerosis Patients Based on Gait Analysis. In: Duffy V., Lightner N. (eds) *Advances in Human Factors and Ergonomics in Healthcare and Medical Devices. AHFE 2017. Advances in Intelligent Systems and Computing*, vol 590, pp. 119-128. Springer, Cham, 2018. doi:10.1007/978-3-319-60483-1_13

30. Vitoantonio Bevilacqua, Gianpaolo Francesco Trotta, Claudio Loconsole, Antonio Brunetti, Nicholas Caporusso, Giuseppe Maria Bellantuono, Irio De Feudis, Donato Patruno, Domenico De Marco, Andrea Venneri, Maria Grazia Di Vietro, Giacomo Losavio, and Sabina Ilaria Tatò. A RGB-D sensor based tool for assessment and rating of movement disorders. In: Duffy V., Lightner N. (eds) *Advances in Human Factors and Ergonomics in Healthcare and Medical Devices. AHFE 2017. Advances in Intelligent Systems and Computing*, vol 590, pp. 119-128. Springer, Cham, 2018. doi:10.1007/978-3-319-60483-1_12

31. Nicholas Caporusso, Luigi Biasi, Giovanni Cinquepalmi, Gianpaolo Francesco Trotta, Antonio Brunetti, and Vitoantonio Bevilacqua. Enabling Touch-based Communication in Wearable Devices for People with Sensory and Multisensory Impairments. In: Ahram T., Falcão C. (eds) *Advances in Human Factors in Wearable Technologies and Game Design. AHFE 2017. Advances in Intelligent Systems and Computing*, vol 608, pp. 149-159. Springer, Cham, 2018. doi:10.1007/978-3-319-60639-2_15

32. Nicholas Caporusso, Luigi Biasi, Giovanni Cinquepalmi, Gianpaolo Francesco Trotta, Antonio Brunetti, and Vitoantonio Bevilacqua. A Wearable Device Supporting Multiple Touch- and Gesture-based Languages for the Deaf-blind. In: Ahram T., Falcão C. (eds) *Advances in Human Factors in Wearable Technologies and Game Design. AHFE 2017. Advances in Intelligent Systems and Computing*, vol 608, pp. 32-41. Springer, Cham, 2018. doi:10.1007/978-3-319-60639-2_4

33. Vitoantonio Bevilacqua, Antonio Brunetti, Gianpaolo Trotta, Giovanni Dimauro, Katarina Elez, Vito Alberotanza, and Arnaldo Scardapane. A Novel Approach for Hepatocellular Carcinoma Detection and Classification Based on Triphasic CT Protocol. In: *2017 IEEE Congress on Evolutionary Computation, CEC 2017* – Proceedings, pp. 1856-1863. IEEE, 2017. doi:10.1109/CEC.2017.7969527

34. Sara Invitto, Giulia Piraino, Arianna Mignozzi, Simona Capone, Giovanni Montagna, Pietro Aleardo Siciliano, Andrea Mazzatenta, Gianbattista Rocco, Irio De Feudis, Gianpaolo Francesco Trotta, Antonio Brunetti, and Vitoantonio Bevilacqua. Smell and Meaning: an OERP study. In: In: Esposito A., Faudez-Zanuy M., Morabito F., Pasero E. (eds) *Multidisciplinary Approaches to Neural Computing. Smart Innovation, Systems and Technologies*, vol 69, pp. 289-300. Springer, Cham, 2017. doi:10.1007/978-3-319-56904-8_28

35. Vitoantonio Bevilacqua, Marco Riezzo, Antonio Brunetti, Francesco Russo, Benedetta D'Attoma, and Giuseppe Riezzo. Analysis and optimization of the 13C octanoic acid breath test. In: *2017 Proceedings of the International Joint Conference on Neural Networks* (IJCNN 2017), pp. 4528-4533. IEEE; 2017. doi:10.1109/IJCNN.2017.7966430

36. Sara Invitto, Arianna Mignozzi, Giulia Piraino, Gianbattista Rocco, Irio De Feudis, Antonio Brunetti, and Vitoantonio Bevilacqua. Artificial Neural Network Analysis and ERP in Intimate Partner Violence. In: Esposito A., Faudez-Zanuy M., Morabito F., Pasero E. (eds) *Multidisciplinary Approaches to Neural Computing. Smart Innovation, Systems and Technologies*, vol 69, pp. 247-254. Springer, Cham, 2017. doi:10.1007/978-3-319-56904-8_24

37. Ilaria Bortone, Gianpaolo Francesco Trotta, Antonio Brunetti, Giacomo Donato Cascarano, Claudio Loconsole, Nadia Agnello, Alberto Argentiero, Giuseppe Nicolardi, Antonio Frisoli, and Vitoantonio Bevilacqua. A Novel Approach in Combination of 3D Gait Analysis Data for Aiding Clinical Decision-Making in Patients with Parkinson's Disease. In: Huang DS., Jo KH., Figueroa-García J. (eds) *Intelligent Computing Theories and Application. ICIC 2017. Lecture Notes in Computer Science*, vol 10362, pp. 493-503. Springer, Cham, 2017. doi:10.1007/978-3-319-63312-1_43

38. Claudio Loconsole, Gianpaolo Francesco Trotta, Antonio Brunetti, Joseph Trotta, Angelo Schiavone, Sabina Ilaria Tatò, Giacomo Losavio, and Vitoantonio Bevilacqua. Computer Vision and EMG-Based Handwriting Analysis for Classification in Parkinson's Disease. In: Huang DS., Jo KH., Figueroa-García J. (eds) *Intelligent Computing Theories and Application. ICIC 2017. Lecture Notes in Computer Science*, vol 10362, pp. 504-514. Springer, Cham, 2017. doi:10.1007/978-3-319-63312-1_44

39. Vitoantonio Bevilacqua, Daniele Altini, Martino Bruni, Marco Riezzo, Antonio Brunetti, Claudio Loconsole, Andrea Guerriero, Gianpaolo Francesco Trotta, Rocco Fasano, Marica Di Pirchio, Cristina Tartaglia, Elena Ventrella, Michele Telegrafo and Marco Moschetta. A Supervised Breast Lesion Images Classification from Tomosynthesis Technique. In: Huang DS., Jo KH., Figueroa-García J. (eds) *Intelligent Computing Theories and Application. ICIC 2017. Lecture Notes in Computer Science*, vol 10362, pp. 483-489. Springer, Cham, 2017. doi:10.1007/978-3-319-63312-1_42

40. Vitoantonio Bevilacqua, Sergio Simeone, Antonio Brunetti, Claudio Loconsole, Gianpaolo Francesco Trotta, Salvatore Tramacere, Antonio Argentieri, Francesco Ragni, Giuseppe Criscenti, Andrea Fornaro, Rosalina Mastronardi, Serena Cassetta and Giuseppe D'Ippolito. A Computer Aided Ophthalmic Diagnosis System Based on Tomographic Features. In: Huang DS., Hussain A., Han K., Gromiha M. (eds) *Intelligent Computing Methodologies. ICIC 2017. Lecture Notes in Computer Science*, vol 10363, pp. 508-609. Springer, Cham, 2017. doi:10.1007/978-3-319-63315-2_52

41. Vitoantonio Bevilacqua, Antonio Emmanuele Uva, Michele Fiorentino, Gianpaolo Francesco Trotta, Maurizio Dimatteo, Enrico Nasca, Attilio Nicola Lino Nocera, Giacomo Donato Cascarano, Antonio Brunetti, Nicholas Caporusso, Roberta Pellicciari, and Giovanni Defazio. A comprehensive method for assessing the blepharospasm cases severity. In: Santosh K., Hangarge M., Bevilacqua V., Negi A. (eds) *Recent Trends in Image Processing and Pattern Recognition. RTIP2R 2016. Communications in Computer and Information Science*, vol 709, pp. 369-381. Springer, Singapore, 2017. doi:10.1007/978-981-10-4859-3_33

42. Vitoantonio Bevilacqua, Leonarda Carnimeo, Antonio Brunetti, Andrea De Pace, Pietro Galeandro, Gianpaolo Francesco Trotta, Nicholas Caporusso, Francescomaria Marino, Vito Alberotanza, and Arnaldo Scardapane. Synthesis of a Neural Network Classifier for Hepatocellular Carcinoma Grading based on triphasic CT images. In: Santosh K., Hangarge M., Bevilacqua V., Negi A. (eds) *Recent Trends in Image Processing and Pattern Recognition. RTIP2R 2016. Communications in Computer and Information Science*, vol 709, pp. 356-368. Springer, Singapore, 2017. doi:10.1007/978-981-10-4859-3_32

References

[1] Eric S Holmboe and Steven J Durning. Assessing clinical reasoning: moving from in vitro to in vivo. *Diagnosis*, 1(1):111–117, 2014.

[2] Hardeep Singh and Mark L Graber. Improving diagnosis in health care–the next imperative for patient safety. *The New England journal of medicine*, 373(26):2493, 2015.

[3] Martin A Makary and Michael Daniel. Medical error—the third leading cause of death in the us. *Bmj*, 353:i2139, 2016.

[4] Molla S Donaldson, Janet M Corrigan, Linda T Kohn, et al. *To err is human: building a safer health system*, volume 6. National Academies Press, 2000.

[5] Mike J Hallworth. The '70% claim': what is the evidence base? *Annals of Clinical Biochemistry*, 48(6):487–488, 2011. doi: 10.1258/acb.2011.011177.

[6] Reza Mirnezami, Jeremy Nicholson, and Ara Darzi. Preparing for precision medicine. *New England Journal of Medicine*, 366(6):489–491, 2012.

[7] Francis S. Collins and Harold Varmus. A new initiative on precision medicine. *New England Journal of Medicine*, 372(9):793–795, 2015. doi: 10.1056/NEJMp1500523.

[8] Hugo JWL Aerts. The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA oncology*, 2(12):1636–1642, 2016.

[9] Pohchoo Seow, Jeannie Hsiu Ding Wong, Azlina Ahmad-Annuar, Abhishek Mahajan, Nor Aniza Abdullah, and Norlisah Ramli. Quantitative magnetic resonance imaging and radiogenomic biomarkers for glioma characterisation: a systematic review. *The British journal of radiology*, 91(1092):20170930, 2018.

[10] Virendra Kumar, Yuhua Gu, Satrajit Basu, Anders Berglund, Steven A Eschrich, Matthew B Schabath, Kenneth Forster, Hugo JWL Aerts, Andre Dekker, David Fenstermacher, et al. Radiomics: the process and the challenges. *Magnetic resonance imaging*, 30(9):1234–1248, 2012.

[11] Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, Sara Carvalho, Ruud GPM Van Stiphout, Patrick Granton, Catharina ML Zegers, Robert Gillies, Ronald Boellard, André Dekker, et al. Radiomics: extracting more information from

medical images using advanced feature analysis. *European journal of cancer*, 48(4): 441–446, 2012.

[12] Carlo N De Cecco, Maria Ciolina, Damiano Caruso, Marco Rengo, Balaji Ganeshan, Felix G Meinel, Daniela Musio, Francesca De Felice, Vincenzo Tombolini, and Andrea Laghi. Performance of diffusion-weighted imaging, perfusion imaging, and texture analysis in predicting tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3t mr: initial experience. *Abdominal Radiology*, 41(9): 1728–1735, 2016.

[13] Karoline Skogen, Anselm Schulz, Johann Baptist Dormagen, Balaji Ganeshan, Eirik Helseth, and Andrès Server. Diagnostic performance of texture analysis on mri in grading cerebral gliomas. *European journal of radiology*, 85(4):824–829, 2016.

[14] Andrew D Smith, Mark R Gray, Sara Martin Del Campo, Darya Shlapak, Balaji Ganeshan, Xu Zhang, and William E Carson III. Predicting overall survival in patients with metastatic melanoma on antiangiogenic therapy and recist stable disease on initial posttherapy images using ct texture analysis. *American Journal of Roentgenology*, 205(3):W283–W293, 2015.

[15] Koichi Hayano, Fang Tian, Avinash R Kambadakone, Sam S Yoon, Dan G Duda, Balaji Ganeshan, and Dushyant V Sahani. Texture analysis of non-contrast enhanced ct for assessing angiogenesis and survival of soft tissue sarcoma. *Journal of computer assisted tomography*, 39(4):607, 2015.

[16] Carlo N De Cecco, Balaji Ganeshan, Maria Ciolina, Marco Rengo, Felix G Meinel, Daniela Musio, Francesca De Felice, Nicola Raffetto, Vincenzo Tombolini, and Andrea Laghi. Texture analysis as imaging biomarker of tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3-t magnetic resonance. *Investigative radiology*, 50(4):239–245, 2015.

[17] Haowei Zhang, Caleb M Graham, Okan Elci, Michael E Griswold, Xu Zhang, Majid A Khan, Karen Pitman, Jimmy J Caudell, Robert D Hamilton, Balaji Ganeshan, et al. Locally advanced squamous cell carcinoma of the head and neck: Ct texture and histogram analysis allow independent prediction of overall survival in patients treated with induction chemotherapy. *Radiology*, 269(3):801–809, 2013.

[18] Karoline Skogen, Balaji Ganeshan, Catriona Good, Giles Critchley, and Ken Miles. Measurements of heterogeneity in gliomas on computed tomography relationship to tumour grade. *Journal of neuro-oncology*, 111(2):213–219, 2013.

[19] David A Mankoff. A definition of molecular imaging. *Journal of Nuclear Medicine*, 48(6):18N–21N, 2007.

[20] Christian J Herold, Jonathan S Lewin, Andreas G Wibmer, James H Thrall, Gabriel P Krestin, Adrian K Dixon, Stefan O Schoenberg, Rena J Geckle, Ada Muellner, and

Hedvig Hricak. Imaging in the age of precision medicine: summary of the proceedings of the 10th biannual symposium of the international society for strategic studies in radiology. *Radiology*, 279(1):226–238, 2015.

[21] Christian J. Herold, Jonathan S. Lewin, Andreas G. Wibmer, James H. Thrall, Gabriel P. Krestin, Adrian K. Dixon, Stefan O. Schoenberg, Rena J. Geckle, Ada Muellner, and Hedvig Hricak. Imaging in the age of precision medicine: Summary of the proceedings of the 10th biannual symposium of the international society for strategic studies in radiology. *Radiology*, 279(1):226–238, 2016. doi: 10.1148/radiol.2015150709.

[22] Maciej A Mazurowski. Radiogenomics: what it is and why it is important. *Journal of the American College of Radiology*, 12(8):862–866, 2015.

[23] Robin W Jansen, Paul van Amstel, Roland M Martens, Irsan E Kooi, Pieter Wesseling, Adrianus J de Langen, Catharina W Menke-Van der Houven, et al. Non-invasive tumor genotyping using radiogenomic biomarkers, a systematic review and oncology-wide pathway analysis. *Oncotarget*, 9(28):20134, 2018.

[24] Javier M Romero, H Bart Brouwers, JingJing Lu, Josser E Delgado Almandoz, Hillary Kelly, Jeremy Heit, Joshua Goldstein, Jonathan Rosand, and R Gilberto Gonzalez. Prospective validation of the computed tomographic angiography spot sign score for intracerebral hemorrhage. *Stroke*, 44(11):3097–3102, 2013.

[25] Thomas Mirsen. Acute treatment of hypertensive intracerebral hemorrhage. *Current treatment options in neurology*, 12(6):504–517, 2010.

[26] Karl F Schmitt and David A Wofford. Financial analysis projects clear returns from electronic medical records: Demonstrating the economic benefits of an electronic medical record is possible with the input of staff who can identify the technology's benefits. *Healthcare Financial Management*, 56(1):52–58, 2002.

[27] Anant Madabhushi and George Lee. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical Image Analysis*, 33:170 – 175, 2016. doi: 10.1016/j.media.2016.06.037.

[28] Xin Qi, Daihou Wang, Ivan Rodero, Javier Diaz-Montes, Rebekah H Gensure, Fuyong Xing, Hua Zhong, Lauri Goodell, Manish Parashar, David J Foran, et al. Content-based histopathology image retrieval using cometcloud. *BMC bioinformatics*, 15(1):287, 2014.

[29] L. A. D. Cooper, A. B. Carter, A. B. Farris, F. Wang, J. Kong, D. A. Gutman, P. Widener, T. C. Pan, S. R. Cholleti, A. Sharma, T. M. Kurc, D. J. Brat, and J. H. Saltz. Digital pathology: Data-intensive frontier in medical imaging. *Proceedings of the IEEE*, 100(4):991–1003, April 2012. doi: 10.1109/JPROC.2011.2182074.

[30] Yukako Yagi and John R Gilbertson. Digital imaging in pathology: the case for standardization, 2005.

[31] Nataliya Titova and K Ray Chaudhuri. Personalized medicine in parkinson's disease: time to be precise. *Movement Disorders*, 32(8):1147, 2017.

[32] Haydeh Payami. The emerging science of precision medicine and pharmacogenomics for parkinson's disease. *Movement disorders*, 32(8):1139–1146, 2017.

[33] R Cascella, C Strafella, M Ragazzo, S Zampatti, P Borgiani, S Gambardella, A Pirazzoli, G Novelli, and E Giardina. Direct pcr: a new pharmacogenetic approach for the inexpensive testing of hla-b* 57: 01. *The pharmacogenomics journal*, 15(2):196, 2015.

[34] John S Satterlee, Andrea Beckel-Mitchener, A Roger Little, Dena Procaccini, Joni L Rutter, and Amy C Lossie. Neuroepigenomics: resources, obstacles, and opportunities. *Neuroepigenetics*, 1:2–13, 2015.

[35] Abhijeet Pataskar and Vijay K Tiwari. Computational challenges in modeling gene regulatory events. *Transcription*, 7(5):188–195, 2016.

[36] E Giardina and CF Caltagirone. The irccs network of neuroscience and neurorehabilitation: the italian platform for care and research about neurodegenerative disorders. In *EUROPEAN JOURNAL OF NEUROLOGY*, volume 25, pages 209–209. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2018.

[37] Russell R Pate, Michael Pratt, Steven N Blair, William L Haskell, Caroline A Macera, Claude Bouchard, David Buchner, Walter Ettinger, Gregory W Heath, Abby C King, et al. Physical activity and public health: a recommendation from the centers for disease control and prevention and the american college of sports medicine. *Jama*, 273(5):402–407, 1995.

[38] Rory A Cooper and Louis A Quatrano. Research on physical activity and health among people with disabilities: a consensus statement. *Journal of Rehabilitation Research & Development*, 36(2), 1999.

[39] J Larry Durstine, Patricia Painter, Barry A Franklin, Don Morgan, Kenneth H Pitetti, and Scott O Roberts. Physical activity for the chronically ill and disabled. *Sports Medicine*, 30(3):207–219, 2000.

[40] Gregory W Heath and Peter H Fentem. Physical activity among persons with disabilities–a public health perspective. *Exercise and sport sciences reviews*, 25: 195–234, 1997.

[41] JAMES H Rimmer, DAVID Braddock, and KENNETH H Pitetti. Research on physical activity and disability: an emerging national priority. *Medicine and science in sports and exercise*, 28(11):1366–1372, 1996.

[42] James H Rimmer. Health promotion for people with disabilities: the emerging paradigm shift from disability prevention to prevention of secondary conditions. *Physical therapy*, 79(5):495–502, 1999.

[43] Roy J Shephard. Benefits of sport and physical activity for the disabled: implications for the individual and for society. *Scandinavian journal of rehabilitation medicine*, 23 (2):51–59, 1991.

[44] Yinong Chong, Richard J Klein, Christine Plepys, and Richard Troiano. Operational definitions for year 2000 objectives: priority area 1: physical activity and fitness. *Healthy people statistical notes*, (18), 1998.

[45] Hidde P Van der Ploeg, Allard J Van der Beek, Luc HV Van der Woude, and Willem van Mechelen. Physical activity for people with a disability. *Sports medicine*, 34(10): 639–649, 2004.

[46] Bram van Ginneken, Cornelia M. Schaefer-Prokop, and Mathias Prokop. Computer-aided diagnosis: How to move from the laboratory to the clinic. *Radiology*, 261(3): 719–732, 2011. doi: 10.1148/radiol.11091710.

[47] Clifford S Goodman and Roy Ahn. Methodological approaches of health technology assessment. *International Journal of Medical Informatics*, 56(1):97 – 105, 1999. ISSN 1386-5056. doi: 10.1016/S1386-5056(99)00049-0.

[48] Edward Hance Shortliffe and Bruce G Buchanan. *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*. Addison-Wesley Publishing Company, 1985.

[49] Donald Waterman. A guide to expert systems. 1986.

[50] Shu-Hsien Liao. Expert system methodologies and applications—a decade review from 1995 to 2004. *Expert systems with applications*, 28(1):93–103, 2005.

[51] Arati Gurung, Carolyn G Scrafford, James M Tielsch, Orin S Levine, and William Checkley. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. *Respiratory medicine*, 105(9):1396–1403, 2011.

[52] Fabian J Theis and Anke Meyer-Bäse. *Biomedical signal analysis: Contemporary methods and applications*. MIT Press, 2010.

[53] Ziad Obermeyer and Ezekiel J Emanuel. Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, 375(13):1216, 2016.

[54] Wenqing Sun, Bin Zheng, and Wei Qian. Computer aided lung cancer diagnosis with deep learning algorithms. In *Medical imaging 2016: computer-aided diagnosis*, volume 9785, page 97850Z. International Society for Optics and Photonics, 2016.

[55] Maryellen L Giger. Machine learning in medical imaging. *Journal of the American College of Radiology*, 15(3):512–520, 2018.

[56] Antonio Brunetti, Leonarda Carnimeo, Gianpaolo Francesco Trotta, and Vitoantonio Bevilacqua. Computer-assisted frameworks for classification of liver, breast and blood neoplasias via neural networks: A survey based on medical images. *Neurocomputing*, 335:274–298, 2019. doi: 10.1016/j.neucom.2018.06.080.

[57] Harry Weinrauch and Albert W Hetherington. Computers in medicine and biology. *Journal of the American Medical Association*, 169(3):240–245, 1959.

[58] Robert S Ledley. Digital electronic computers in biomedical science. *Science*, 130 (3384):1225–1234, 1959.

[59] Steven G Vandenberg. Medical diagnosis by computer: Recent attempts and outlook for the future. *Behavioral Science*, 5(2):170, 1960.

[60] Robert S Ledley and Lee B Lusted. Reasoning foundations of medical diagnosis. *Science*, 130(3366):9–21, 1959.

[61] Sholom M Weiss, Casimir A Kulikowski, Saul Amarel, and Aran Safir. A model-based method for computer-aided medical decision-making. *Artificial intelligence*, 11(1-2): 145–172, 1978.

[62] William B Schwartz. Medicine and the computer: the promise and problems of change. In *Use and Impact of Computers in Clinical Medicine*, pages 321–335. Springer, 1970.

[63] Howard L Bleich. Computer-based consultation: electrolyte and acid-base disorders. *The American journal of medicine*, 53(3):285–291, 1972.

[64] Robert A Rosati, J Frederick McNeer, C Frank Starmer, Brant S Mittler, James J Morris, and Andrew G Wallace. A new information system for medical practice. *Archives of Internal Medicine*, 135(8):1017–1024, 1975.

[65] G Anthony Gorry, Jerome P Kassirer, Alvin Essig, and William B Schwartz. Decision analysis as the basis for computer-aided management of acute renal failure. *The American journal of medicine*, 55(4):473–484, 1973.

[66] Peter Szolovits, Ramesh S Patil, and William B Schwartz. Artificial intelligence in medical diagnosis. *Annals of internal medicine*, 108(1):80–87, 1988.

[67] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.

[68] Mark A Musen, Blackford Middleton, and Robert A Greenes. Clinical decision-support systems. In *Biomedical informatics*, pages 643–674. Springer, 2014.

[69] Kunio Doi. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4-5): 198–211, 2007.

[70] Gwilym S Lodwick, Cosmo L Haun, Walton E Smith, Roy F Keller, and Eddie D Robertson. Computer diagnosis of primary bone tumors: A preliminary report. *Radiology*, 80(2):273–275, 1963.

[71] Phillip H Meyers, Charles M Nice Jr, Hal C Becker, Wilson J Nettleton Jr, James W Sweeney, and George R Meckstroth. Automated computer analysis of radiographic images. *Radiology*, 83(6):1029–1034, 1964.

[72] Fred Winsberg, Milton Elkin, Josiah Macy Jr, Victoria Bordaz, and William Weymouth. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. *Radiology*, 89(2):211–215, 1967.

[73] James C Mohr. American medical malpractice litigation in historical perspective. *Jama*, 283(13):1731–1737, 2000.

[74] David M Studdert, Michelle M Mello, Troyen A Brennan, et al. Medical malpractice. *New England Journal of Medicine*, 350(3), 2004.

[75] Brian V Nahed, Maya A Babu, Timothy R Smith, and Robert F Heary. Malpractice liability and defensive medicine: a national survey of neurosurgeons. *PloS one*, 7(6): e39237, 2012.

[76] SM Astley and Fiona Jane Gilbert. Computer-aided detection in mammography. *Clinical radiology*, 59(5):390–399, 2004.

[77] LGBA Quekel, R Goei, AGH Kessels, and JMA van Engelshoven. Detection of lung cancer on the chest radiograph: impact of previous films, clinical information, double reading, and dual reading. *Journal of clinical epidemiology*, 54(11):1146–1150, 2001.

[78] Kunio Doi. Current status and future potential of computer-aided diagnosis in medical imaging. *The British journal of radiology*, 78(suppl_1):s3–s19, 2005.

[79] Maciej A Mazurowski, Piotr A Habas, Jacek M Zurada, Joseph Y Lo, Jay A Baker, and Georgia D Tourassi. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural networks*, 21(2):427–436, 2008.

[80] Xingquan Zhu and Xindong Wu. Class noise vs. attribute noise: A quantitative study. *Artificial intelligence review*, 22(3):177–210, 2004.

[81] Robert Nisbet, John Elder, and Gary Miner. Chapter 4 - data understanding and preparation. In Robert Nisbet, John Elder, and Gary Miner, editors, *Handbook of Statistical Analysis and Data Mining Applications*, pages 49 – 75. Academic Press, Boston, 2009. ISBN 978-0-12-374765-5. doi: 10.1016/B978-0-12-374765-5.00004-8.

[82] Davide Chicco. Ten quick tips for machine learning in computational biology. *BioData mining*, 10(1):35, 2017.

[83] Vitoantonio Bevilacqua, A. Aulenta, E. Carioggia, Giuseppe Mastronardi, Filippo Menolascina, G. Simeone, Angelo Paradiso, Antonio Scarpa, and Diego Taurino. Metallic artifacts removal in breast CT images for treatment planning in radiotherapy by means of supervised and unsupervised neural network algorithms. In De-Shuang Huang, Laurent Heutte, and Marco Loog, editors, *Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues, Third International Conference on Intelligent Computing, ICIC 2007, Qingdao, China, August 21-24, 2007, Proceedings*, volume 4681 of *Lecture Notes in Computer Science*, pages 1355–1363. Springer, 2007. ISBN 978-3-540-74170-1. doi: 10.1007/978-3-540-74171-8_138.

[84] Muna O Al-Hatmi and Jabar H Yousif. A review of image enhancement systems and a case study of salt &pepper noise removing. *International Journal of Computation and Applied Sciences IJOCAAS*, 3(2):217–223, 2017.

[85] B Suneetha and A JhansiRani. A survey on image processing techniques for brain tumor detection using magnetic resonance imaging. In *Innovations in Green Energy and Healthcare Technologies (IGEHT), 2017 International Conference on*, pages 1–6. IEEE, 2017.

[86] G Niranjana and M Ponnavaikko. A review on image processing methods in detecting lung cancer using ct images. In *Technical Advancements in Computers and Communications (ICTACC), 2017 International Conference on*, pages 18–25. IEEE, 2017.

[87] King-Sun Fu and JK Mui. A survey on image segmentation. *Pattern recognition*, 13 (1):3–16, 1981.

[88] Hui Zhang, Jason E Fritts, and Sally A Goldman. Image segmentation evaluation: A survey of unsupervised methods. *computer vision and image understanding*, 110(2): 260–280, 2008.

[89] Ruikai Zhang, Yali Zheng, Tony Wing Chung Mak, Ruoxi Yu, Sunny H Wong, James YW Lau, and Carmen CY Poon. Automatic detection and classification of colorectal polyps by transferring low-level cnn features from nonmedical domain. *IEEE journal of biomedical and health informatics*, 21(1):41–47, 2016.

[90] Venkatesh N Murthy, Subhransu Maji, and R Manmatha. Automatic image annotation using deep learning representations. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pages 603–606. ACM, 2015.

[91] Robert M Haralick, Karthikeyan Shanmugam, et al. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6):610–621, 1973.

[92] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

[93] Nicholas A Hamilton, Radosav S Pantelic, Kelly Hanson, and Rohan D Teasdale. Fast automated cell phenotype image classification. *BMC bioinformatics*, 8(1):110, 2007.

[94] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. *Computer vision–ECCV 2006*, pages 404–417, 2006.

[95] Elaine Johanna Limkin, Sylvain Reuzé, Alexandre Carré, Roger Sun, Antoine Schernberg, Anthony Alexis, Eric Deutsch, Charles Ferté, and Charlotte Robert. The complexity of tumor shape, spiculatedness, correlates with tumor radiomic shape features. *Scientific reports*, 9(1):1–12, 2019.

[96] Robert J Gillies, Paul E Kinahan, and Hedvig Hricak. Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2016.

[97] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[98] Laurens Van Der Maaten, Eric Postma, and Jaap Van den Herik. Dimensionality reduction: a comparative. *J Mach Learn Res*, 10:66–71, 2009.

[99] Bo Li, Chao Wang, and De-Shuang Huang. Supervised feature extraction based on orthogonal discriminant projection. *Neurocomputing*, 73(1-3):191–196, 2009.

[100] De-Shuang Huang and Jian-Xun Mi. A new constrained independent component analysis method. *IEEE transactions on neural networks*, 18(5):1532–1535, 2007.

[101] Zhan-Li Sun, De-Shuang Huang, Chun-Hou Zheng, and Li Shang. Optimal selection of time lags for tdsep based on genetic algorithm. *Neurocomputing*, 69(7-9):884–887, 2006.

[102] Chun-Hou Zheng, De-Shuang Huang, Zhan-Li Sun, Michael R Lyu, and Tat-Ming Lok. Nonnegative independent component analysis based on minimizing mutual information technique. *Neurocomputing*, 69(7-9):878–883, 2006.

[103] Chun-Hou Zheng, De-Shuang Huang, Kang Li, George Irwin, and Zhan-Li Sun. Misep method for postnonlinear blind source separation. *Neural computation*, 19(9): 2557–2578, 2007.

[104] Chun-Hou Zheng, De-Shuang Huang, and Li Shang. Feature selection in independent component subspace for microarray data classification. *Neurocomputing*, 69(16-18): 2407–2410, 2006.

[105] Imola K Fodor. A survey of dimension reduction techniques. Technical report, Lawrence Livermore National Lab., CA (US), 2002.

[106] De-Shuang Huang and Wen Jiang. A general cpl-ads methodology for fixing dynamic parameters in dual environments. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(5):1489–1500, 2012.

[107] Vitoantonio Bevilacqua, Maurizio Triggiani, Maurizio Dimatteo, Giuseppe Bellantuono, Antonio, Leonarda Carnimeo, Francescomaria Marino, Michele Telegrafo, and Marco Moschetta. Computer assisted detection of breast lesions in magnetic resonance images. In De-Shuang Huang, Vitoantonio Bevilacqua, and Prashan Premaratne, editors, *Intelligent Computing Theories and Application - 12th International Conference, ICIC 2016, Lanzhou, China, August 2-5, 2016, Proceedings, Part I*, volume 9771 of *Lecture Notes in Computer Science*, pages 306–316. Springer, 2016. ISBN 978-3-319-42290-9. doi: 10.1007/978-3-319-42291-6_30.

[108] Vitoantonio Bevilacqua, Antonio Brunetti, Maurizio Triggiani, Domenico Magaletti, Michele Telegrafo, and Marco Moschetta. An optimized feed-forward artificial neural network topology to support radiologists in breast lesions classification. In Friedrich et al. [372], pages 1385–1392. ISBN 978-1-4503-4323-7. doi: 10.1145/2908961.2931733.

[109] Shubhi Sharma and Pritee Khanna. Computer-aided diagnosis of malignant mammograms using zernike moments and svm. *Journal of Digital Imaging*, 28(1):77–90, Feb 2015. ISSN 1618-727X. doi: 10.1007/s10278-014-9719-7.

[110] J Dheeba, N Albert Singh, and S Tamil Selvi. Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *Journal of biomedical informatics*, 49:45–52, 2014.

[111] Ismail Saritas. Prediction of breast cancer using artificial neural networks. *Journal of Medical Systems*, 36(5):2901–2907, 2012.

[112] Vitoantonio Bevilacqua, Paolo Pannarale, Mirko Abbrescia, Claudia Cava, Angelo Paradiso, and Stefania Tommasi. Comparison of data-merging methods with svm attribute selection and classification in breast cancer gene expression. *BMC bioinformatics*, 13 (7):S9, 2012.

[113] Hui-Ling Chen, Bo Yang, Jie Liu, and Da-You Liu. A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. *Expert Systems with Applications*, 38(7):9014–9022, 2011.

[114] Jianmin Jiang, P Trundle, and Jinchang Ren. Medical image analysis with artificial neural networks. *Computerized Medical Imaging and Graphics*, 34(8):617–631, 2010.

[115] RR Janghel, Anupam Shukla, Ritu Tiwari, and Rahul Kala. Breast cancer diagnosis using artificial neural network models. In *Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference on*, pages 89–94. IEEE, 2010.

[116] Dustin Newell, Ke Nie, Jeon-Hor Chen, Chieh-Chih Hsu, Hon J. Yu, Orhan Nalcioglu, and Min-Ying Su. Selection of diagnostic features on breast mri to differentiate between malignant and benign lesions using computer-aided diagnosis: differences in lesions presenting as mass and non-mass-like enhancement. *European Radiology*, 20 (4):771–781, Apr 2010. ISSN 1432-1084. doi: 10.1007/s00330-009-1616-y.

[117] Y Rejani and S Thamarai Selvi. Early detection of breast cancer using svm classifier technique. *arXiv preprint arXiv:0912.2314*, 2009.

[118] Mehmet Fatih Akay. Support vector machines combined with feature selection for breast cancer diagnosis. *Expert systems with applications*, 36(2):3240–3247, 2009.

[119] Vitoantonio Bevilacqua, Giuseppe Mastronardi, Filippo Menolascina, Paolo Pannarale, and Antonio Pedone. A novel multi-objective genetic algorithm approach to artificial neural network topology optimisation: The breast cancer classification problem. In *Proceedings of the International Joint Conference on Neural Networks, IJCNN 2006, part of the IEEE World Congress on Computational Intelligence, WCCI 2006, Vancouver, BC, Canada, 16-21 July 2006*, pages 1958–1965. IEEE, 2006. ISBN 0-7803-9490-9. doi: 10.1109/IJCNN.2006.246940.

[120] Hussein A Abbass. An evolutionary artificial neural networks approach for breast cancer diagnosis. *Artificial intelligence in Medicine*, 25(3):265–281, 2002.

[121] De-Shuang Huang. Systematic theory of neural networks for pattern recognition. *Publishing House of Electronic Industry of China, Beijing*, 201, 1996.

[122] De-shuang Huang. Radial basis probabilistic neural networks: Model and application. *International Journal of Pattern Recognition and Artificial Intelligence*, 13(07):1083–1101, 1999.

[123] De-Shuang Huang and Ji-Xiang Du. A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks. *IEEE Transactions on Neural Networks*, 19(12):2099–2115, 2008.

[124] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[125] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.

[126] IA Basheer and M Hajmeer. Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods*, 43(1):3–31, 2000.

[127] Eric B Baum and David Haussler. What size net gives valid generalization? In *Advances in neural information processing systems*, pages 81–90, 1989.

[128] Farid U Dowla and Leah L Rogers. *Solving problems in environmental engineering and geosciences with artificial neural networks*. Mit Press, 1995.

[129] Simon Haykin and Neural Network. A comprehensive foundation. *Neural Networks*, 2(2004):41, 2004.

[130] Timothy Masters. *Practical neural network recipes in C++*. Morgan Kaufmann, 1993.

[131] Carl G Looney. Advances in feedforward neural networks: demystifying knowledge acquiring black boxes. *IEEE Transactions on Knowledge and Data Engineering*, 8(2): 211–226, 1996.

[132] Kevin Swingler. *Applying neural networks: a practical guide*. Morgan Kaufmann, 1996.

[133] Sylvain Arlot, Alain Celisse, et al. A survey of cross-validation procedures for model selection. *Statistics surveys*, 4:40–79, 2010.

[134] David A Winkler and Tu C Le. Performance of deep and shallow neural networks, the universal approximation theorem, activity cliffs, and qsar. *Molecular informatics*, 36 (1-2), 2017.

[135] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[136] PG Benardos and G-C Vosniakos. Optimizing feedforward artificial neural network architecture. *Engineering Applications of Artificial Intelligence*, 20(3):365–382, 2007.

[137] Jan Kukačka, Vladimir Golkov, and Daniel Cremers. Regularization for deep learning: A taxonomy. 2018.

[138] De-Shuang Huang, Xing-Ming Zhao, Guang-Bin Huang, and Yiu-Ming Cheung. Classifying protein sequences using hydropathy blocks. *Pattern recognition*, 39(12): 2293–2300, 2006.

[139] Kun-Hong Liu and De-Shuang Huang. Cancer classification using rotation forest. *Computers in biology and medicine*, 38(5):601–610, 2008.

[140] Fei Han and De-Shuang Huang. Improved extreme learning machine for function approximation by encoding a priori information. *Neurocomputing*, 69(16-18):2369–2373, 2006.

[141] Josef Sivic, Bryan C Russell, Alexei A Efros, Andrew Zisserman, and William T Freeman. Discovering objects and their location in images. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pages 370–377. IEEE, 2005.

[142] Herman Kamper, Aren Jansen, and Sharon Goldwater. Fully unsupervised small-vocabulary speech recognition using a segmental bayesian model. In *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[143] Honglak Lee, Peter Pham, Yan Largman, and Andrew Y Ng. Unsupervised feature learning for audio classification using convolutional deep belief networks. In *Advances in neural information processing systems*, pages 1096–1104, 2009.

[144] Meherwar Fatima, Maruf Pasha, et al. Survey of machine learning algorithms for disease diagnostic. *Journal of Intelligent Learning Systems and Applications*, 9(01):1, 2017.

[145] Khalid Raza and Nripendra Kumar Singh. A tour of unsupervised deep learning for medical image analysis. *arXiv preprint arXiv:1812.07715*, 2018.

[146] Hoo-Chang Shin, Matthew R Orton, David J Collins, Simon J Doran, and Martin O Leach. Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1930–1943, 2012.

[147] Kiran Vaidhya, Subramaniam Thirunavukkarasu, Varghese Alex, and Ganapathy Krishnamurthi. Multi-modal brain tumor segmentation using stacked denoising autoencoders. In *BrainLes 2015*, pages 181–194. Springer, 2015.

[148] S Sivakumar and C Chandrasekar. Lung nodule segmentation through unsupervised clustering models. *Procedia engineering*, 38:3064–3073, 2012.

[149] Devinder Kumar, Alexander Wong, and David A Clausi. Lung nodule classification using deep features in ct images. In *2015 12th Conference on Computer and Robot Vision*, pages 133–138. IEEE, 2015.

[150] Michiel Kallenberg, Kersten Petersen, Mads Nielsen, Andrew Y Ng, Pengfei Diao, Christian Igel, Celine M Vachon, Katharina Holland, Rikke Rass Winkel, Nico Karssemeijer, et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE transactions on medical imaging*, 35(5): 1322–1331, 2016.

[151] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[152] Nitin Kumar. Thresholding in salient object detection: a survey. *Multimedia Tools and Applications*, pages 1–32, 2017.

[153] Vitoantonio Bevilacqua, Alessandro Piazzolla, and Paolo Stofella. Atlas-based segmentation of organs at risk in radiotherapy in head mris by means of a novel active contour framework. In De-Shuang Huang, Xiang Zhang, Carlos A. Reyes García,

and Lei Zhang, editors, *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence, 6th International Conference on Intelligent Computing, ICIC 2010, Changsha, China, August 18-21, 2010. Proceedings*, volume 6216 of *Lecture Notes in Computer Science*, pages 350–359. Springer, 2010. ISBN 978-3-642-14931-3. doi: 10.1007/978-3-642-14932-0_44.

[154] Li Shang, De-Shuang Huang, Chun-Hou Zheng, and Zhan-Li Sun. Noise removal using a novel non-negative sparse coding shrinkage technique. *Neurocomputing*, 69 (7-9):874–877, 2006.

[155] Paul Aljabar, Rolf A Heckemann, Alexander Hammers, Joseph V Hajnal, and Daniel Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*, 46(3):726–738, 2009.

[156] Jan Sijbers and Andrei Postnov. Reduction of ring artefacts in high resolution micro-ct reconstructions. *Physics in Medicine & Biology*, 49(14):N247, 2004.

[157] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[158] Saleha Masood, Muhammad Sharif, Afifa Masood, Mussarat Yasmin, and Mudassar Raza. A survey on medical image segmentation. *Current Medical Imaging*, 11(1): 3–14, 2015.

[159] Pablo Mesejo, Óscar Ibáñez, Óscar Cordón, and Stefano Cagnoni. A survey on image segmentation using metaheuristic-based deformable models: state of the art and critical analysis. *Applied Soft Computing*, 44:1 – 29, 2016. ISSN 1568-4946. doi: 10.1016/j.asoc.2016.03.004.

[160] Michael D Abràmoff, PJ Magelhaes, and SJ Ram. Image processing with. 2004.

[161] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553): 436–444, 2015.

[162] Hayit Greenspan, Bram van Ginneken, and Ronald M Summers. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, 35(5):1153–1159, 2016.

[163] Vitoantonio Bevilacqua, Nicola Pietroleonardo, Vito Triggiani, Antonio Brunetti, Anna Maria Di Palma, Michele Rossini, and Loreto Gesualdo. An innovative neural network framework to classify blood vessels and tubules based on haralick features evaluated in histological images of kidney biopsy. *Neurocomputing*, 228:143–153, mar 2017. ISSN 18728286. doi: 10.1016/j.neucom.2016.09.091.

[164] Vitoantonio Bevilacqua, Antonio Brunetti, Giacomo Donato Cascarano, Flavio Palmieri, Andrea Guerriero, and Marco Moschetta. A deep learning approach for the automatic detection and segmentation in autosomal dominant polycystic kidney disease based on magnetic resonance images. In De-Shuang Huang, Kang-Hyun Jo, and Xiaolong Zhang, editors, *Intelligent Computing Theories and Application - 14th International Conference, ICIC 2018, Wuhan, China, August 15-18, 2018, Proceedings, Part II*, volume 10955 of *Lecture Notes in Computer Science*, pages 643–649, Cham, 2018. Springer. ISBN 978-3-319-95932-0. doi: 10.1007/978-3-319-95933-7_73.

[165] Amitav Banerjee, UB Chitnis, SL Jadhav, JS Bhawalkar, and S Chaudhury. Hypothesis testing, type i and type ii errors. *Industrial psychiatry journal*, 18(2):127, 2009.

[166] Geraint Lewis, Jessica Sheringham, Jamie Lopez Bernal, and Tim Crayford. *Mastering public health: a postgraduate guide to examinations and revalidation*. CRC Press, 2014.

[167] Seong Ho Park, Jin Mo Goo, and Chan-Hee Jo. Receiver operating characteristic (roc) curve: practical review for radiologists. *Korean journal of radiology*, 5(1):11–18, 2004.

[168] Lee B Lusted. Decision-making studies in patient management. *New England Journal of Medicine*, 284(8):416–424, 1971.

[169] David J Goodenough, Kurt Rossmann, and Lee B Lusted. Radiographic applications of receiver operating characteristic (roc) curves. *Radiology*, 110(1):89–95, 1974.

[170] Charles E Metz. Basic principles of roc analysis. In *Seminars in nuclear medicine*, volume 8, pages 283–298. Elsevier, 1978.

[171] James A Hanley and Barbara J McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

[172] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678. ACM, 2014.

[173] Dong Yu, Adam Eversole, Mike Seltzer, Kaisheng Yao, Zhiheng Huang, Brian Guenter, Oleksii Kuchaiev, Yu Zhang, Frank Seide, Huaming Wang, et al. An introduction to computational networks and the computational network toolkit. *Microsoft Technical Report MSR-TR-2014–112*, 2014.

[174] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.

[175] The Theano Development Team, Rami Al-Rfou, Guillaume Alain, Amjad Almahairi, Christof Angermueller, Dzmitry Bahdanau, Nicolas Ballas, Frédéric Bastien, Justin Bayer, Anatoly Belikov, et al. Theano: A python framework for fast computation of mathematical expressions. *arXiv preprint arXiv:1605.02688*, 2016.

[176] Ronan Collobert, Laurens Van Der Maaten, and Armand Joulin. Torchnet: an open-source platform for (deep) learning research. In *Proceedings of the 33rd International Conference on Machine Learning (ICML-2016)*, pages 19–24, 2016.

[177] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1717–1724. IEEE, 2014.

[178] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[179] Andrea Vedaldi and Karel Lenc. Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 689–692. ACM, 2015.

[180] Jeremy Karnowski. Alexnet + svm, 2015. URL https://jeremykarnowski.files.wordpress.com/2015/07/alexnet2.png. Online; accessed INSERT DATE.

[181] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[182] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11. Springer, 2018.

[183] Karl Fritscher, Patrik Raudaschl, Paolo Zaffino, Maria Francesca Spadea, Gregory C Sharp, and Rainer Schubert. Deep neural networks for fast segmentation of 3d medical images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 158–165. Springer, 2016.

[184] Holger R Roth, Chen Shen, Hirohisa Oda, Masahiro Oda, Yuichiro Hayashi, Kazunari Misawa, and Kensaku Mori. Deep learning and its application to medical image segmentation. *arXiv preprint arXiv:1803.08691*, 2018.

[185] Grigore C Burdea and Philippe Coiffet. *Virtual reality technology*. John Wiley & Sons, 2003.

[186] Vitoantonio Bevilacqua, Antonio Brunetti, Giuseppe Trigiante, Gianpaolo Francesco Trotta, Michele Fiorentino, Vito Manghisi, and Antonio Emmanuele Uva. Design and

development of a forearm rehabilitation system based on an augmented reality serious game. In *Italian Workshop on Artificial Life and Evolutionary Computation*, pages 127–136. Springer, 2015.

[187] Vitoantonio Bevilacqua, Antonio Brunetti, Davide de Biase, Giacomo Tattoli, Rosario Santoro, Gianpaolo Francesco Trotta, Fabio Cassano, Michele Pantaleo, Giuseppe Mastronardi, Fabio Ivona, et al. A p300 clustering of mild cognitive impairment patients stimulated in an immersive virtual reality scenario. In *International Conference on Intelligent Computing*, pages 226–236. Springer, 2015.

[188] Marina de Tommaso, Katia Ricci, Marianna Delussi, Anna Montemurno, Eleonora Vecchio, Antonio Brunetti, and Vitoantonio Bevilacqua. Testing a novel method for improving wayfinding by means of a p3b virtual reality visual paradigm in normal aging. *SpringerPlus*, 5(1):1297, 2016.

[189] Giuseppe Riva, Brenda K Wiederhold, and Fabrizia Mantovani. Neuroscience of virtual reality: From virtual exposure to embodied medicine. *Cyberpsychology, Behavior, and Social Networking*, 22(1):82–96, 2019.

[190] Giuseppe Riva. Virtual reality in psychotherapy. *Cyberpsychology & behavior*, 8(3): 220–230, 2005.

[191] Roos Pot-Kolder, Wim Veling, Willem-Paul Brinkman, and Mark van der Gaag. Virtual reality and psychotic disorders. In *Virtual Reality for Psychological and Neurocognitive Interventions*, pages 289–305. Springer, 2019.

[192] Paul MG Emmelkamp. Technological innovations in clinical assessment and psychotherapy. *Psychotherapy and psychosomatics*, 74(6):336–343, 2005.

[193] Giuseppe Riva, Andrea Gaggioli, and Fabrizia Mantovani. Are robots present? from motor simulation to "being there". *Cyberpsychology & Behavior*, 11(6):631–636, 2008.

[194] Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior*, 10(1):45–56, 2007.

[195] Maria T Schultheis and Albert A Rizzo. The application of virtual reality technology in rehabilitation. *Rehabilitation psychology*, 46(3):296, 2001.

[196] Barbara O Rothbaum, Larry F Hodges, David Ready, Ken Graap, and Renato D Alarcon. Virtual reality exposure therapy for vietnam veterans with posttraumatic stress disorder. *The Journal of clinical psychiatry*, 2001.

[197] Gabriele Optale, Silvia Marin, Massimiliano Pastore, Alberto Nasta, and Carlo Pianon. Male sexual dysfunctions and multimedia immersion therapy (follow-up). *CyberPsychology & Behavior*, 6(3):289–294, 2003.

[198] Paul MG Emmelkamp, Mary Bruynzeel, Leonie Drost, and Charles AP G van der Mast. Virtual reality treatment in acrophobia: a comparison with exposure in vivo. *CyberPsychology & Behavior*, 4(3):335–339, 2001.

[199] Soliane Scapin, Maria Elena Echevarria-Guanilo, Paulo Roberto Boeira Fuculo Junior, Natalia Goncalves, Patricia Kuerten Rocha, and Rebeca Coimbra. Virtual reality in the treatment of burn patients: A systematic review. *Burns*, 44(6):1403–1416, 2018.

[200] Daniela Villani, Francesco Riva, and Giuseppe Riva. New technologies for relaxation: The role of presence. *International Journal of Stress Management*, 14(3):260, 2007.

[201] Giuseppe Riva, Monica Bacchetta, Gianluca Cesa, Sara Conti, and Enrico Molinari. Six-month follow-up of in-patient experiential cognitive therapy for binge eating disorders. *Cyberpsychology & behavior*, 6(3):251–258, 2003.

[202] Giuseppe Riva. From virtual to real body: virtual reality as embodied technology. *J. Cyber Ther. Rehabil*, 1:7–22, 2008.

[203] Craig D Murray and Michael S Gordon. Changes in bodily awareness induced by immersive virtual reality. *CyberPsychology & Behavior*, 4(3):365–371, 2001.

[204] Marta Ferrer-García and José Gutiérrez-Maldonado. The use of virtual reality in the study, assessment, and treatment of body image in eating disorders and nonclinical samples: a review of the literature. *Body image*, 9(1):1–11, 2012.

[205] Francesca Morganti. Virtual interaction in cognitive neuropsychology., 2004.

[206] Marina de Tommaso, Eleonora Gentile, Katia Ricci, Anna Montemurno, Marianna Delussi, Eleonora Vecchio, Giancarlo Logroscino, Antonio Brunetti, and Vitoantonio Bevilacqua. Bioelectrical correlates of emotional changes induced by environmental sound and colour: From virtual reality to real life. In *International Conference on NeuroRehabilitation*, pages 982–985. Springer, 2018.

[207] Alessandra Gorini, Andrea Gaggioli, and Giuseppe Riva. Virtual worlds, real healing. *Science*, 318(5856):1549–1549, 2007.

[208] Alessandra Gorini, Andrea Gaggioli, and Giuseppe Riva. A second life for ehealth: prospects for the use of 3-d virtual worlds in clinical psychology. *Journal of medical Internet research*, 10(3):e21, 2008.

[209] Michael J Tarr and William H Warren. Virtual reality in behavioral neuroscience and beyond. *nature neuroscience*, 5(11s):1089, 2002.

[210] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L Siegel, Lindsey A Torre, and Ahmedin Jemal. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6):394–424, 2018.

[211] J Ferlay, M Colombet, I Soerjomataram, C Mathers, DM Parkin, M Piñeros, A Znaor, and F Bray. Estimating the global cancer incidence and mortality in 2018: Globocan sources and methods. *International journal of cancer*, 144(8):1941–1953, 2019.

[212] Jerrold T Bushberg. *The essential physics of medical imaging*. Lippincott Williams & Wilkins, 2002.

[213] Steve Webb. *The physics of medical imaging*. CRC Press, 1988.

[214] Clare MC Tempany, Kelly H Zou, Stuart G Silverman, Douglas L Brown, Alfred B Kurtz, and Barbara J McNeil. Staging of advanced ovarian cancer: comparison of imaging modalities—report from the radiological diagnostic oncology group. *Radiology*, 215(3):761–767, 2000.

[215] Susan G Orel and Mitchell D Schnall. Mr imaging of the breast for the detection, diagnosis, and staging of breast cancer. *Radiology*, 220(1):13–30, 2001.

[216] Victoria Kut, William Spies, Stewart Spies, William Gooding, and Athanassios Argiris. Staging and monitoring of small cell lung cancer using [18f] fluoro-2-deoxy-d-glucose-positron emission tomography (fdg-pet). *American journal of clinical oncology*, 30 (1):45–50, 2007.

[217] A Vestito, Fabio Felice Mangieri, G Gatta, Marco Moschetta, B Turi, and Antonietta Ancona. Breast carcinoma in elderly women. our experience. *Il giornale di chirurgia*, 32(1):411–416, 2011.

[218] Debbie Saslow, Carla Boetes, Wylie Burke, Steven Harms, Martin O Leach, Constance D Lehman, Elizabeth Morris, Etta Pisano, Mitchell Schnall, Stephen Sener, et al. American cancer society guidelines for breast screening with mri as an adjunct to mammography. *CA: a cancer journal for clinicians*, 57(2):75–89, 2007.

[219] Ravi K Samala, Heang-Ping Chan, Lubomir Hadjiiski, Kenny Cha, and Mark A Helvie. Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis. *SPIE medical imaging. International Society for Optics and Photonics*, pages 97850Y–97850Y, 2016.

[220] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[221] Gisella Gennaro, Alicia Toledano, Cosimo Di Maggio, Enrica Baldan, Elisabetta Bezzon, Manuela La Grassa, Luigi Pescarini, Ilaria Polico, Alessandro Proietti, Aida Toffoli, et al. Digital breast tomosynthesis versus digital mammography: a clinical performance study. *European radiology*, 20(7):1545–1553, 2010.

[222] Etta D Pisano and Christopher A Parham. Digital mammography, sestamibi breast scintigraphy, and positron emission tomography breast imaging. *Radiologic Clinics of North America*, 38(4):861–869, 2000.

[223] Heang-Ping Chan, Jun Wei, Berkman Sahiner, Elizabeth A Rafferty, Tao Wu, Marilyn A Roubidoux, Richard H Moore, Daniel B Kopans, Lubomir M Hadjiiski, and Mark A Helvie. Computer-aided detection system for breast masses on digital tomosynthesis mammograms: preliminary experience. *Radiology*, 237(3):1075–1080, 2005.

[224] John M Boone, Alex LC Kwan, Kai Yang, George W Burkett, Karen K Lindfors, and Thomas R Nelson. Computed tomography for imaging the breast. *Journal of mammary gland biology and neoplasia*, 11(2):103–111, 2006.

[225] Tao Wu, Juemin Zhang, Richard Moore, Elizabeth Rafferty, Daniel Kopans, Waleed Meleis, and David Kaeli. Digital tomosynthesis mammography using a parallel maximum-likelihood reconstruction method. In *Proc. SPIE*, volume 5368, pages 0277–786X, 2004.

[226] Katrina E Korhonen, Susan P Weinstein, Elizabeth S McDonald, and Emily F Conant. Strategies to increase cancer detection: Review of true-positive and false-negative results at digital breast tomosynthesis screening. *RadioGraphics*, 36(7):1954–1965, 2016.

[227] Vitoantonio Bevilacqua, Daniele Altini, Martino Bruni, Marco Riezzo, Antonio Brunetti, Claudio Loconsole, Andrea Guerriero, Gianpaolo Francesco Trotta, Rocco Fasano, Marica Di Pirchio, Cristina Tartaglia, Elena Ventrella, Michele Telegrafo, and Marco Moschetta. A supervised breast lesion images classification from tomosynthesis technique. In De-Shuang Huang, Kang-Hyun Jo, and Juan Carlos Figueroa-García, editors, *Intelligent Computing Theories and Application - 13th International Conference, ICIC 2017, Liverpool, UK, August 7-10, 2017, Proceedings, Part II*, volume 10362 of *Lecture Notes in Computer Science*, pages 483–489. Springer, 2017. ISBN 978-3-319-63311-4. doi: 10.1007/978-3-319-63312-1_42.

[228] L-K Soh and Costas Tsatsoulis. Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on geoscience and remote sensing*, 37(2):780–795, 1999.

[229] David A Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of remote sensing*, 28(1):45–62, 2002.

[230] Vitoantonio Bevilacqua, Nicola Pietroleonardo, Vito Triggiani, Loreto Gesualdo, Anna Maria Di Palma, Michele Rossini, Giuseppe Dalfino, and Nico Mastrofilippo. Neural network classification of blood vessels and tubules based on haralick features evaluated in histological images of kidney biopsy. In *Advanced Intelligent Computing Theories and Applications*, pages 759–765. Springer International Publishing, 2015.

[231] Mykola Pechenizkiy, Alexey Tsymbal, and Seppo Puuronen. Pca-based feature transformation for classification: issues in medical diagnostics. In *Proceedings. 17th IEEE Symposium on Computer-Based Medical Systems*, pages 535–540. IEEE, 2004.

[232] Jitendra Virmani, Nilanjan Dey, Vinod Kumar, et al. Pca-pnn and pca-svm based cad systems for breast density classification. In *Applications of intelligent optimization in biology and medicine*, pages 159–180. Springer, 2016.

[233] Dario Garcia-Gasulla, Ferran Parés, Armand Vilalta, Jonatan Moreno, Eduard Ayguadé, Jesús Labarta, Ulises Cortés, and Toyotaro Suzumura. On the behavior of convolutional nets for feature extraction. *Journal of Artificial Intelligence Research*, 61:563–592, 2018.

[234] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[235] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[236] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[237] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[238] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014.

[239] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.

[240] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

[241] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[242] Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. When is "nearest neighbor" meaningful? In *International conference on database theory*, pages 217–235. Springer, 1999.

[243] Pedro Domingos and Michael Pazzani. On the optimality of the simple bayesian classifier under zero-one loss. *Machine learning*, 29(2):103–130, 1997.

[244] Haim Dahan, Shahar Cohen, Lior Rokach, and Oded Maimon. *Proactive data mining with decision trees*. Springer Science & Business Media, 2014.

[245] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of human genetics*, 7(2):179–188, 1936.

[246] Bo Li, Chun-Hou Zheng, and De-Shuang Huang. Locally linear discriminant embedding: An efficient method for face recognition. *Pattern Recognition*, 41(12): 3813–3821, 2008.

[247] Zhibin Liao and Gustavo Carneiro. On the importance of normalisation layers in deep learning with piecewise linear activation units. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–8. IEEE, 2016.

[248] Luke Taylor and Geoff Nitschke. Improving deep learning using generic data augmentation. *arXiv preprint arXiv:1708.06020*, 2017.

[249] SS Kumar and RS Moni. Diagnosis of liver tumor from ct images using fast discrete curvelet transform. *IJCA Special Issue on Computer Aided Soft Computing Techniques for Imaging and Biomedical Applications (CASCT)*, pages 48–57, 2010.

[250] S Gr Mougiakakou, I Valavanis, KS Nikita, A Nikita, and D Kelekis. Characterization of ct liver lesions based on texture features and a multiple neural network classification scheme. In *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, volume 2, pages 1287–1290. IEEE, 2003.

[251] Alessandro Cucchetti, Fabio Piscaglia, Antonia D'Errico Grigioni, Matteo Ravaioli, Matteo Cescon, Matteo Zanello, Gian Luca Grazi, Rita Golfieri, Walter Franco Grigioni, and Antonio Daniele Pinna. Preoperative prediction of hepatocellular carcinoma tumour grade and micro-vascular invasion by means of artificial neural network: a pilot study. *Journal of hepatology*, 52(6):880–888, 2010.

[252] K Mala, V Sadasivam, and S Alagappan. Neural network based texture analysis of liver tumor from computed tomography images. *International Journal of Biological and Medical Sciences*, 21:33–41, 2007.

[253] Yu-Len Huang, Jeon-Hor Chen, and Wu-Chung Shen. Diagnosis of hepatic tumors with texture analysis in nonenhanced computed tomography images. *Academic radiology*, 13(6):713–720, 2006.

[254] S Vijayarani and S Dhayanand. Liver disease prediction using svm and naïve bayes algorithms. *International Journal of Science, Engineering and Technology Research (IJSETR)*, 4(4):816–820, 2015.

[255] Anju Gulia, Rajan Vohra, and Praveen Rani. Liver patient classification using intelligent techniques. *International Journal of Computer Science and Information Technologies*, 5(4):5110–5115, 2014.

[256] Fadl Mutaher Ba-Alwi and Houzifa M Hintaya. Comparative study for analysis the prognostic in hepatitis data: data mining approach. *spinal cord*, 11:12, 2013.

[257] Bekir Karlik. Hepatitis disease diagnosis using backpropagation and the naive bayes classifiers. *IBU Journal of Science and Technology*, 1(1), 2012.

[258] G Sathyadevi. Application of cart algorithm in hepatitis disease diagnosis. In *Recent Trends in Information Technology (ICRTIT), 2011 International Conference on*, pages 1283–1287. IEEE, 2011.

[259] P Rajeswari and G Sophia Reena. Analysis of liver disorder using data mining algorithm. *Global journal of computer science and technology*, 10(14), 2010.

[260] SS Kumar and Devi Devapal. Survey on recent cad system for liver disease diagnosis. In *Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference on*, pages 763–766. IEEE, 2014.

[261] Aaron Adcock, Daniel Rubin, and Gunnar Carlsson. Classification of hepatic lesions using the matching metric. *Computer vision and image understanding*, 121:36–42, 2014.

[262] SS Kumar, RS Moni, and J Rajeesh. Automatic liver and lesion segmentation: a primary step in diagnosis of liver diseases. *Signal, Image and Video Processing*, pages 1–10, 2013.

[263] Peter Dankerl, Alexander Cavallaro, Alexey Tsymbal, Maria Jimena Costa, Michael Suehling, Rolf Janka, Michael Uder, and Matthias Hammon. A retrieval-based computer-aided diagnosis system for the characterization of liver lesions in ct scans. *Academic radiology*, 20(12):1526–1534, 2013.

[264] J Vincey and MMV Jeba. Computer aided diagnosis for liver cancer feature extraction. *The International Journal of Engineering and Science (IJES)*, 2(11):27–30, 2013.

[265] Dorota Duda, Marek Krętowski, and Johanne Bézy-Wendling. Computer-aided diagnosis of liver tumors based on multi-image texture analysis of contrast-enhanced ct. selection of the most appropriate texture features. *Studies in Logic, Grammar and Rhetoric*, 35(1):49–70, 2013.

[266] Yen-Wei Chen, Jie Luo, Chunhua Dong, Xianhua Han, Tomoko Tateyama, Akira Furukawa, and Shuzo Kanasaki. Computer-aided diagnosis and quantification of cirrhotic livers based on morphological analysis and machine learning. *Computational and mathematical methods in medicine*, 2013, 2013.

[267] AK Thakre and AI Dhenge. Ct liver image diagnosis classification system. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(1), 2013.

[268] Priyanjana Sharma, Shagun Malik, Surbhi Sehgal, and Jyotika Pruthi. Computer aided diagnosis based on medical image processing and artificial intelligence methods. *International Journal of Information and Computation Technology*, 3(9):887–92, 2013.

[269] S Gunasundari and M Suganya Ananthi. Comparison and evaluation of methods for liver tumor classification from ct datasets. *International journal of computer applications*, 39(18):46–51, 2012.

[270] Rahmath S Hameed and SS Kumar. Assessment of neural network based classifiers to diagnose focal liver lesions using ct images. *Procedia Engineering*, 38:4048–4056, 2012.

[271] Ömer Kayaalti, Bekir H Aksebzeci, Ibrahim Ö Karahan, Kemal Deniz, Menmet Öztürk, Bülent Yilmaz, Sadik Kara, and Musa H Asyali. Staging of the liver fibrosis from ct images using texture features. In *Health Informatics and Bioinformatics (HIBIT), 2012 7th International Symposium on*, pages 47–52. IEEE, 2012.

[272] Patrick Ferdinand Christ, Florian Ettlinger, Felix Grün, Mohamed Ezzeldin A Elshaera, Jana Lipkova, Sebastian Schlecht, Freba Ahmaddy, Sunil Tatavarty, Marc Bickel, Patrick Bilic, et al. Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks. *arXiv preprint arXiv:1702.05970*, 2017.

[273] Xiao Han. Automatic liver lesion segmentation using a deep convolutional neural network method. *arXiv preprint arXiv:1704.07239*, 2017.

[274] R Vivanti, A Szeskin, N Lev-Cohain, J Sosna, and L Joskowicz. Automatic detection of new tumors and tumor burden evaluation in longitudinal liver ct scan studies. *International Journal of Computer Assisted Radiology and Surgery*, 12(11):1945–1957, 2017.

[275] Yading Yuan. Hierarchical convolutional-deconvolutional neural networks for automatic liver and tumor segmentation. *arXiv preprint arXiv:1710.04540*, 2017.

[276] Avi Ben-Cohen, Idit Diamant, Eyal Klang, Michal Amitai, and Hayit Greenspan. Fully convolutional network for liver segmentation and lesions detection. In *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pages 77–85. Springer, 2016.

[277] Wen Li, Fucang Jia, and Qingmao Hu. Automatic segmentation of liver tumor in ct images with deep convolutional neural networks. *Journal of Computer and Communications*, 3(11):146, 2015.

[278] Alejandro Forner, Josep M Llovet, and Jordi Bruix. Hepatocellular carcinoma. *The Lancet*, 379(9822):1245 – 1255, 2012. ISSN 0140-6736. doi: 10.1016/S0140-6736(11)61347-0.

[279] Maurizio Memeo, AA Ianora Stabile, A Scardapane, P Suppressa, A Cirulli, C Sabba, A Rotondo, and G Angelelli. Hereditary haemorrhagic telangiectasia: study of hepatic vascular alterations with multi-detector row helical ct and reconstruction programs. *La Radiologia Medica*, 109(1-2):125–138, 2005.

[280] Ianora AA Stabile, P Pedote, Am Scardapane, M Memeo, A Rotondo, and G Angelelli. Preoperative staging of gastric carcinoma with multidetector spiral ct. *La Radiologia medica*, 106(5-6):467, 2003.

[281] Vitoantonio Bevilacqua, Antonio Brunetti, Gianpaolo Francesco Trotta, Giovanni Dimauro, Katarina Elez, Vito Alberotanza, and Arnaldo Scardapane. A novel approach for hepatocellular carcinoma detection and classification based on triphasic ct protocol. In *Evolutionary Computation (CEC), 2017 IEEE Congress on*, pages 1856–1863. IEEE, 2017.

[282] Hugh A Edmondson and Paul E Steiner. Primary carcinoma of the liver. a study of 100 cases among 48,900 necropsies. *Cancer*, 7(3):462–503, 1954.

[283] Wen Jiang, De-Shuang Huang, and Shenghong Li. Random walk-based solution to triple level stochastic point location problem. *IEEE transactions on cybernetics*, 46 (6):1438–1451, 2016.

[284] Zhong-Qiu Zhao and De-Shuang Huang. A mended hybrid learning algorithm for radial basis function neural networks to improve generalization capability. *Applied Mathematical Modelling*, 31(7):1271–1281, 2007.

[285] Ji-Xiang Du, De-Shuang Huang, Xiao-Feng Wang, and Xiao Gu. Shape recognition based on neural networks trained by differential evolution algorithm. *Neurocomputing*, 70(4-6):896–903, 2007.

[286] Ji-Xiang Du, De-Shuang Huang, Guo-Jun Zhang, and Zeng-Fu Wang. A novel full structure optimization algorithm for radial basis probabilistic neural networks. *Neurocomputing*, 70(1-3):592–596, 2006.

[287] Fei Han and De-Shuang Huang. A new constrained learning algorithm for function approximation by encoding a priori information into feedforward neural networks. *Neural Computing and Applications*, 17(5-6):433–439, 2008.

[288] Fei Han, Qing-Hua Ling, and De-Shuang Huang. Modified constrained learning algorithms incorporating additional functional constraints into neural networks. *Information Sciences*, 178(3):907–919, 2008.

[289] Jun Zhang, De-Shuang Huang, Tat-Ming Lok, and Michael R Lyu. A novel adaptive sequential niche technique for multimodal function optimization. *Neurocomputing*, 69(16-18):2396–2401, 2006.

[290] Vitoantonio Bevilacqua, Antonio Brunetti, Gianpaolo Francesco Trotta, Leonarda Carnimeo, Francescomaria Marino, Vito Alberotanza, and Arnaldo Scardapane. A deep learning approach for hepatocellular carcinoma grading. *IJCVIP*, 7(2):1–18, 2017. doi: 10.4018/IJCVIP.2017040101.

[291] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.

[292] Jared J. Grantham. Autosomal dominant polycystic kidney disease. *New England Journal of Medicine*, 359(14):1477–1485, oct 2008. ISSN 0028-4793. doi: 10.1056/NEJMcp0804458.

[293] P. C. Harris, K. T. Bae, S. Rossetti, V. E. Torres, J. J. Grantham, A. B. Chapman, L. M. Guay-Woodford, B. F. King, L. H. Wetzel, D. A. Baumgarten, P. J. Kenney, M. Consugar, S. Klahr, W. M. Bennett, C. M. Meyers, Q. Zhang, P. A. Thompson, F. Zhu, and J. P. Miller. Cyst number but not the rate of cystic growth is associated with the mutated gene in autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology*, 17(11):3013–3019, sep 2006. ISSN 1046-6673. doi: 10.1681/ASN.2006080835.

[294] Vicente E. Torres, Arlene B. Chapman, Olivier Devuyst, Ron T. Gansevoort, Jared J. Grantham, Eiji Higashihara, Ronald D. Perrone, Holly B. Krasa, John Ouyang, and Frank S. Czerwiec. Tolvaptan in patients with autosomal dominant polycystic kidney disease. *New England Journal of Medicine*, 367(25):2407–2418, 2012. doi: 10.1056/NEJMoa1205511.

[295] Maria V. Irazabal, Vicente E. Torres, Marie C. Hogan, James Glockner, Bernard F. King, Troy G. Ofstie, Holly B. Krasa, John Ouyang, and Frank S. Czerwiec. Short-term effects of tolvaptan on renal function and volume in patients with autosomal dominant polycystic kidney disease. *Kidney International*, 80(3):295–301, aug 2011. ISSN 15231755. doi: 10.1038/ki.2011.119.

[296] Kyongtae T Bae, Paul K Commean, and Jeongrim Lee. Volumetric measurement of renal cysts and parenchyma using mri: Phantoms and patients with polycystic kidney disease. *Journal of Computer Assisted Tomography*, 24(4):614–619, 2000. ISSN 03638715. doi: 10.1097/00004728-200007000-00019.

[297] B F King, J E Reed, E J Bergstralh, P F Sheedy, and V E Torres. Quantification and longitudinal trends of kidney, renal cyst, and renal parenchyma volumes in autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology : JASN*, 11(8):1505–1511, aug 2000. ISSN 1046-6673.

[298] Jean Nicolas Vauthey, Eddie K. Abdalla, Dorota A. Doherty, Philippe Gertsch, Marc J. Fenstermacher, Evelyne M. Loyer, Jan Lerut, Roland Materne, Xuemei Wang, Arthur

Encarnacion, Delise Herron, Christian Mathey, Giovanni Ferrari, Chuslip Charnsangavej, Kim Anh Do, and Alban Denys. Body surface area and body weight predict total liver volume in western adults. *Liver Transplantation*, 8(3):233–240, mar 2002. ISSN 15276465. doi: 10.1053/jlts.2002.31654.

[299] S. A. Emamian, M. B. Nielsen, J. F. Pedersen, and L. Ytte. Kidney dimensions at sonography: Correlation with age, sex, and habitus in 665 adult volunteers. *American Journal of Roentgenology*, 160(1):83–86, jan 1993. ISSN 0361803X. doi: 10.2214/ajr.160.1.8416654.

[300] Eiji Higashihara, Kikuo Nutahara, Takatsugu Okegawa, Mitsuhiro Tanbo, Hidehiko Hara, Isao Miyazaki, Kuninori Kobayasi, and Toshiaki Nitatori. Kidney volume estimations with ellipsoid equations by magnetic resonance imaging in autosomal dominant polycystic kidney disease. *Nephron*, 129(4):253–262, 2015. ISSN 22353186. doi: 10.1159/000381476.

[301] M. V. Irazabal, L. J. Rangel, E. J. Bergstralh, S. L. Osborn, A. J. Harmon, J. L. Sundsbak, K. T. Bae, A. B. Chapman, J. J. Grantham, M. Mrug, M. C. Hogan, Z. M. El-Zoghby, P. C. Harris, B. J. Erickson, B. F. King, and V. E. Torres. Imaging classification of autosomal dominant polycystic kidney disease: A simple model for selecting patients for clinical trials. *Journal of the American Society of Nephrology*, 26(1):160–172, jan 2015. ISSN 1046-6673. doi: 10.1681/ASN.2013101138.

[302] Kyongtae T. Bae, Cheng Tao, Jinhong Wang, Diana Kaya, Zhiyuan Wu, Junu T. Bae, Arlene B. Chapman, Vicente E. Torres, Jared J. Grantham, Michal Mrug, William M. Bennett, Michael F. Flessner, and Doug P. Landsittel. Novel approach to estimate kidney and cyst volumes using mid-slice magnetic resonance images in polycystic kidney disease. *American Journal of Nephrology*, 38(4):333–341, 2013. ISSN 02508095. doi: 10.1159/000355375.

[303] Jared J. Grantham and Vicente E. Torres. The importance of total kidney volume in evaluating progression of polycystic kidney disease. *Nature Reviews Nephrology*, 12 (11):667–677, nov 2016. ISSN 1759507X. doi: 10.1038/nrneph.2016.135.

[304] Jared J. Grantham, Vicente E. Torres, Arlene B. Chapman, Lisa M. Guay-Woodford, Kyongtae T. Bae, Bernard F. King, Louis H. Wetzel, Deborah A. Baumgarten, Phillip J. Kenney, Peter C. Harris, Saulo Klahr, William M. Bennett, Gladys N. Hirschman, Catherine M. Meyers, Xiaoling Zhang, Fang Zhu, and John P. Miller. Volume progression in polycystic kidney disease. *New England Journal of Medicine*, 354(20): 2122–2130, may 2006. ISSN 0028-4793. doi: 10.1056/NEJMoa054341.

[305] M Biswas, V Kuppili, L Saba, DR Edla, HS Suri, E Cuadrado-Godia, JR Laird, RT Marinhoe, JM Sanches, A Nicolaides, et al. State-of-the-art review on deep learning in medical imaging. *Frontiers in bioscience (Landmark edition)*, 24:392–426, 2019.

[306] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L. Rubin, and Bradley J. Erickson. Deep learning for brain MRI segmentation: State of the art and future directions. *J. Digital Imaging*, 30(4):449–459, 2017. doi: 10.1007/s10278-017-9983-4.

[307] Yann Lecun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553): 436–444, may 2015. ISSN 14764687. doi: 10.1038/nature14539.

[308] Vitoantonio Bevilacqua, Antonio Brunetti, Andrea Guerriero, Gianpaolo Francesco Trotta, Michele Telegrafo, and Marco Moschetta. A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images. *Cognitive Systems Research*, 53:3–19, 2019. doi: 10.1016/j.cogsys.2018.04.011.

[309] Zhen Shen, Wenzheng Bao, and De-Shuang Huang. Recurrent neural network for predicting transcription factor binding sites. *Scientific reports*, 8(1):15270, 2018.

[310] Su-Ping Deng, Shaolong Cao, De-Shuang Huang, and Yu-Ping Wang. Identifying stages of kidney renal cell carcinoma by combining gene expression and dna methylation data. *IEEE/ACM transactions on computational biology and bioinformatics*, 14 (5):1147–1153, 2017.

[311] Hai-Cheng Yi, Zhu-Hong You, De-Shuang Huang, Xiao Li, Tong-Hai Jiang, and Li-Ping Li. A deep learning framework for robust and accurate prediction of ncrna-protein interactions using evolutionary information. *Molecular Therapy-Nucleic Acids*, 11:337–344, 2018.

[312] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, dec 2017. ISSN 13618423. doi: 10.1016/j.media.2017.07.005.

[313] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, apr 2015. ISSN 18792782. doi: 10.1016/j.neunet.2014.09.003.

[314] Riccardo Magistroni, Cristiana Corsi, Teresa Martí, and Roser Torra. A review of the imaging techniques for measuring kidney and cyst volume in establishing autosomal dominant polycystic kidney disease progression. *American journal of nephrology*, 48 (1):67–78, 2018.

[315] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, pages 580–587, Washington, DC, USA, 2014. IEEE Computer Society. ISBN 978-1-4799-5118-5. doi: 10.1109/CVPR.2014.81.

[316] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561, 2015. URL http://arxiv.org/abs/1511.00561.

[317] Gabriel J. Brostow, Julien Fauqueur, and Roberto Cipolla. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*, 30(2): 88–97, 2009. doi: 10.1016/j.patrec.2008.04.005.

[318] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[319] Vitoantonio Bevilacqua, Daniele Altini, Martino Bruni, Marco Riezzo, Antonio Brunetti, Claudio Loconsole, Andrea Guerriero, Gianpaolo Francesco Trotta, Rocco Fasano, Marica Di Pirchio, Cristina Tartaglia, Elena Ventrella, Michele Telegrafo, and Marco Moschetta. A supervised breast lesion images classification from tomosynthesis technique. In De-Shuang Huang, Kang-Hyun Jo, and Juan Carlos Figueroa-García, editors, *Intelligent Computing Theories and Application - 13th International Conference, ICIC 2017, Liverpool, UK, August 7-10, 2017, Proceedings, Part II*, volume 10362 of *Lecture Notes in Computer Science*, pages 483–489, Cham, 2017. Springer. ISBN 978-3-319-63311-4. doi: 10.1007/978-3-319-63312-1_42.

[320] Sebastien C. Wong, Adam Gatt, Victor Stamatescu, and Mark D. McDonnell. Understanding data augmentation for classification: When to warp? In *2016 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2016, Gold Coast, Australia, November 30 - December 2, 2016*, pages 1–6, 2016. doi: 10.1109/DICTA.2016.7797091.

[321] Yan Xu, Ran Jia, Lili Mou, Ge Li, Yunchuan Chen, Yangyang Lu, and Zhi Jin. Improved relation classification by deep recurrent neural networks with data augmentation. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 1461–1470, 2016.

[322] Antonio Brunetti, Domenico Buongiorno, Gianpaolo Francesco Trotta, and Vitoantonio Bevilacqua. Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. *Neurocomputing*, 300:17–33, jul 2018. ISSN 18728286. doi: 10.1016/j.neucom.2018.01.092.

[323] Jaya S Kulchandani and Kruti J Dangarwala. Moving object detection: Review of recent research trends. In *2015 International Conference on Pervasive Computing (ICPC)*, pages 1–5. IEEE, 2015.

[324] C. Lawrence Zitnick and Piotr Dollár. Edge boxes: Locating object proposals from edges. In David J. Fleet, Tomás Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland,*

September 6-12, 2014, Proceedings, Part V, volume 8693 of *Lecture Notes in Computer Science*, pages 391–405, Cham, 2014. Springer. ISBN 978-3-319-10601-4. doi: 10.1007/978-3-319-10602-1_26.

[325] Jasper R. R. Uijlings, Koen E. A. van de Sande, Theo Gevers, and Arnold W. M. Smeulders. Selective search for object recognition. *International Journal of Computer Vision*, 104(2):154–171, 2013. doi: 10.1007/s11263-013-0620-5.

[326] Ross B. Girshick. Fast R-CNN. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pages 1440–1448, 2015. doi: 10.1109/ICCV.2015.169.

[327] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(6):1137–1149, 2017. doi: 10.1109/TPAMI.2016.2577031.

[328] Jolanta Karpinski, Ginette Lajoie, Daniel Cattran, Stanley Fenton, Jeffrey Zaltzman, Carl Cardella, and Edward Cole. Outcome of kidney transplantation from high-risk donors is determined by both structure and function. *Transplantation*, 67(8): 1162–1167, 1999.

[329] Vincenzo Losappio, Giovanni Stallone, Barbara Infante, Antonio Schena, Michele Rossini, Annamaria Maiorano, Marco Fiorentino, Pasquale Ditonno, Giuseppe Lucarelli, Michele Battaglia, et al. A single-center cohort study to define the role of pretransplant biopsy score in the long-term outcome of kidney transplantation. *Transplantation*, 97(9):934–939, 2014.

[330] Patrick D Walker, Tito Cavallo, and Stephen M Bonsib. Practice guidelines for the renal biopsy. *Modern Pathology*, 17(12):1555–1563, 2004.

[331] Douglas B Murphy. *Fundamentals of light microscopy and electronic imaging*. John Wiley & Sons, 2014.

[332] Rafael C Gonzalez and Richard E Woods. *Digital Image Processing*. Pearson III edition, 2007.

[333] Sotiris B Kotsiantis, Ioannis D Zaharakis, and Panayiotis E Pintelas. Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26 (3):159–190, 2006.

[334] Saad Ullah Akram, Juho Kannala, Lauri Eklund, and Janne Heikkilä. Cell segmentation proposal network for microscopy image analysis. In *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pages 21–29. Springer, 2016.

[335] Alessandro Ferrari, Stefano Lombardi, and Alberto Signoroni. Bacterial colony counting by convolutional neural networks. In *Engineering in Medicine and Biology*

Society (EMBC), 2015 37th Annual International Conference of the IEEE, pages 7458–7461. IEEE, 2015.

[336] Zhimin Gao, Lei Wang, Luping Zhou, and Jianjia Zhang. Hep-2 cell image classification with deep convolutional neural networks. *IEEE journal of biomedical and health informatics*, 21(2):416–428, 2017.

[337] Muhammad Nasim Kashif, Shan E Ahmed Raza, Korsuk Sirinukunwattana, Muhammmad Arif, and Nasir Rajpoot. Handcrafted features with convolutional neural networks for detection of tumor cells in histology images. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 1029–1032. IEEE, 2016.

[338] Ha Tran Hong Phan, Ashnil Kumar, Jinman Kim, and Dagan Feng. Transfer learning of a convolutional neural network for hep-2 cell image classification. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 1208–1211. IEEE, 2016.

[339] Anat Shkolyar, Amit Gefen, Dafna Benayahu, and Hayit Greenspan. Automatic detection of cell divisions (mitosis) in live-imaging microscopy images using convolutional neural networks. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 743–746. IEEE, 2015.

[340] Jianwei Zhao, Minshu Zhang, Zhenghua Zhou, Jianjun Chu, and Feilong Cao. Automatic detection and classification of leukocytes using convolutional neural networks. *Medical & biological engineering & computing*, 55(8):1287–1301, 2017.

[341] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, pages 184–199. Springer, 2014.

[342] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[343] Mehdi Habibzadeh, Adam Krzyżak, and Thomas Fevens. Analysis of white blood cell differential counts using dual-tree complex wavelet transform and support vector machine classifier. *Computer Vision and Graphics*, pages 414–422, 2012.

[344] Xin Zheng, Yong Wang, Guoyou Wang, and Jianguo Liu. Fast and robust segmentation of white blood cell images by self-supervised learning. *Micron*, 107:55–71, 2018. doi: 10.1016/j.micron.2018.01.010.

[345] Vitoantonio Bevilacqua, Domenico Buongiorno, Pierluigi Carlucci, Ferdinando Giglio, Giacomo Tattoli, Attilio Guarini, Nicola Sgherza, Giacoma De Tullio, Carla Minoia, Anna Scattone, et al. A supervised cad to support telemedicine in hematology. In *Neural Networks (IJCNN), 2015 International Joint Conference on*, pages 1–7. IEEE, 2015.

[346] PS Hiremath, Parashuram Bannigidad, and Sai Geeta. Automated identification and classification of white blood cells (leukocytes) in digital microscopic images. *IJCA special issue on recent trends in image processing and pattern recognition RTIPPR*, pages 59–63, 2010.

[347] Vincenzo Piuri and Fabio Scotti. Morphological classification of blood leucocytes by microscope images. In *Computational Intelligence for Measurement Systems and Applications, 2004. CIMSA. 2004 IEEE International Conference on*, pages 103–108. IEEE, 2004.

[348] M Alagappan, B BanuRekha, R Arun, M Kalaikamal, S Muthukr-ishnan, CS Sai Ganesh, and S Sathishkumar. Extreme learning machine (elm) based automated identification and classification of white blood cells. In *International Conference on Mathematical Modeling and Applied Soft Computing*, pages 846–852, 2012.

[349] Atin Mathur, Ardhendu S Tripathi, and Manohar Kuse. Scalable system for classification of white blood cells from leishman stained blood stain images. *Journal of pathology informatics*, 4(Suppl), 2013.

[350] Tony Wyss-Coray. Ageing, neurodegeneration and brain rejuvenation. *Nature*, 539 (7628):180–186, 2016.

[351] Karel Smetana, Lukáš Lacina, Pavol Szabo, Barbora Dvořánková, Prokop Brož, and Alekst šedo. Ageing as an important risk factor for cancer. *Anticancer research*, 36 (10):5009–5017, 2016.

[352] Jan R Aunan, William C Cho, and Kjetil Søreide. The biology of aging and cancer: a brief overview of shared and divergent molecular hallmarks. *Aging and disease*, 8(5): 628, 2017.

[353] Heiko Braak and EVA Braak. Staging of alzheimer's disease-related neurofibrillary changes. *Neurobiology of aging*, 16(3):271–278, 1995.

[354] Hans Förstl and Alexander Kurz. Clinical features of alzheimer's disease. *European archives of psychiatry and clinical neuroscience*, 249(6):288–290, 1999.

[355] Clifford R Jack Jr, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, and John Q Trojanowski. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128, 2010.

[356] Marilyn S Albert, Steven T DeKosky, Dennis Dickson, Bruno Dubois, Howard H Feldman, Nick C Fox, Anthony Gamst, David M Holtzman, William J Jagust, Ronald C Petersen, et al. The diagnosis of mild cognitive impairment due to alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):270–279, 2011.

[357] Guy M McKhann, David S Knopman, Howard Chertkow, Bradley T Hyman, Clifford R Jack Jr, Claudia H Kawas, William E Klunk, Walter J Koroshetz, Jennifer J Manly, Richard Mayeux, et al. The diagnosis of dementia due to alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7 (3):263–269, 2011.

[358] Reisa A Sperling, Clifford R Jack Jr, Sandra E Black, Matthew P Frosch, Steven M Greenberg, Bradley T Hyman, Philip Scheltens, Maria C Carrillo, William Thies, Martin M Bednar, et al. Amyloid-related imaging abnormalities in amyloid-modifying therapeutic trials: recommendations from the alzheimer's association research roundtable workgroup. *Alzheimer's & Dementia*, 7(4):367–385, 2011.

[359] Lauren C Costantini, Linda J Barr, Janet L Vogel, and Samuel T Henderson. Hypometabolism as a therapeutic target in alzheimer's disease. *BMC neuroscience*, 9 (S2):S16, 2008.

[360] Claudio Babiloni, Antonio I Triggiani, Roberta Lizio, Susanna Cordone, Giacomo Tattoli, Vitoantonio Bevilacqua, Andrea Soricelli, Raffaele Ferri, Flavio Nobili, Loreto Gesualdo, et al. Classification of single normal and alzheimer's disease individuals from cortical sources of resting state eeg rhythms. *Frontiers in neuroscience*, 10:47, 2016.

[361] Thomas Dierks, Ralf Ihl, Lutz Frölich, and Konrad Maurer. Dementia of the alzheimer type: effects on the spontaneous eeg described by dipole sources. *Psychiatry Research: Neuroimaging*, 50(3):151–162, 1993.

[362] Thomas Dierks, Vesna Jelic, Roberto D Pascual-Marqui, Lars-Olof Wahlund, Per Julin, David EJ Linden, Konrad Maurer, Bengt Winblad, and Agneta Nordberg. Spatial pattern of cerebral glucose metabolism (pet) correlates with localization of intracerebral eeg-generators in alzheimer's disease. *Clinical Neurophysiology*, 111 (10):1817–1824, 2000.

[363] C Huang, L-O Wahlund, T Dierks, P Julin, B Winblad, and V Jelic. Discrimination of alzheimer's disease and mild cognitive impairment by equivalent eeg sources: a cross-sectional and longitudinal study. *Clinical Neurophysiology*, 111(11):1961–1967, 2000.

[364] NV Ponomareva, ND Selesneva, and GA Jarikov. Eeg alterations in subjects at high familial risk for alzheimer's disease. *Neuropsychobiology*, 48(3):152–159, 2003.

[365] Jaeseung Jeong. Eeg dynamics in patients with alzheimer's disease. *Clinical neurophysiology*, 115(7):1490–1505, 2004.

[366] Roberto Domingo Pascual-Marqui et al. Standardized low-resolution brain electromagnetic tomography (sloreta): technical details. *Methods Find Exp Clin Pharmacol*, 24(Suppl D):5–12, 2002.

[367] Bernd Saletu, Peter Anderer, Gerda M Saletu-Zyhlarz, and Roberto D Pascual-Marqui. Eeg mapping and low-resolution brain electromagnetic tomography (loreta) in diagnosis and therapy of psychiatric disorders: evidence for a key-lock principle. *Clinical EEG and neuroscience*, 36(2):108–115, 2005.

[368] Claudio Babiloni, Roberta Lizio, Filippo Carducci, Fabrizio Vecchio, Alberto Redolfi, Silvia Marino, Gioacchino Tedeschi, Patrizia Montella, Antonio Guizzaro, Fabrizio Esposito, et al. Resting state cortical electroencephalographic rhythms and white matter vascular lesions in subjects with alzheimer's disease: an italian multicenter study. *Journal of Alzheimer's Disease*, 26(2):331–346, 2011.

[369] Claudio Babiloni, Francesco Infarinato, Fabienne Aujard, Jesper Frank Bastlund, Marina Bentivoglio, Giuseppe Bertini, Claudio Del Percio, Paolo Francesco Fabene, GianLuigi Forloni, Maria Trinidad Herrero Ezquerro, et al. Effects of pharmacological agents, sleep deprivation, hypoxia and transcranial magnetic stimulation on electroencephalographic rhythms in rodents: towards translational challenge models for drug discovery in alzheimer's disease. *Clinical Neurophysiology*, 124(3):437–451, 2013.

[370] Claudio Babiloni, Francesco Infarinato, Antonio I Triggiani, Roberta Lizio, Claudio Del Percio, Nicola Marzano, and Jill C Richardson. Resting state eeg rhythms as network disease markers for drug discovery in alzheimer's disease. *Drug Discovery Today: Therapeutic Strategies*, 10(2):e85–e90, 2013.

[371] Claudio Babiloni, Roberta Lizio, Nicola Marzano, Paolo Capotosto, Andrea Soricelli, Antonio Ivano Triggiani, Susanna Cordone, Loreto Gesualdo, and Claudio Del Percio. Brain neural synchronization and functional coupling in alzheimer's disease as revealed by resting state eeg rhythms. *International Journal of Psychophysiology*, 103: 88–102, 2016.

[372] Tobias Friedrich, Frank Neumann, and Andrew M. Sutton, editors. *Genetic and Evolutionary Computation Conference, GECCO 2016, Denver, CO, USA, July 20-24, 2016, Companion Material Proceedings*, 2016. ACM. ISBN 978-1-4503-4323-7. doi: 10.1145/2908961.

[373] Frederick Wolfe, Daniel J Clauw, Mary-Ann Fitzcharles, Don L Goldenberg, Robert S Katz, Philip Mease, Anthony S Russell, I Jon Russell, John B Winfield, and Muhammad B Yunus. The american college of rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. *Arthritis care & research*, 62 (5):600–610, 2010.

[374] Fabiola Atzeni, Rossella Talotta, Ignazio Francesco Masala, Camillo Giacomelli, Ciro Conversano, Valeria Nucera, Bruno Lucchino, Cristina Iannuccelli, Manuela Di Franco, and Laura Bazzichi. One year in review 2019: fibromyalgia. *Clin. Exp Rheumatol*, 37:S3–S10, 2019.

[375] Phillip J Mease, Lesley M Arnold, Leslie J Crofford, David A Williams, I Jon Russell, Louise Humphrey, Linda Abetz, and Susan A Martin. Identifying the clinical domains of fibromyalgia: contributions from clinician and patient delphi exercises. *Arthritis Care & Research: Official Journal of the American College of Rheumatology*, 59(7): 952–960, 2008.

[376] Leonardo M Knijnik, Jairo A Dussán-Sarria, Joanna R Rozisky, Iraci LS Torres, Andre R Brunoni, Felipe Fregni, and Wolnei Caumo. Repetitive transcranial magnetic stimulation for fibromyalgia: systematic review and meta-analysis. *Pain Practice*, 16 (3):294–304, 2016.

[377] Neil E O'Connell, Louise Marston, Sally Spencer, Lorraine H DeSouza, and Benedict M Wand. Non-invasive brain stimulation techniques for chronic pain. *Cochrane database of systematic reviews*, (3), 2018.

[378] Asbjørn J Fagerlund, Odd A Hansen, and Per M Aslaksen. Transcranial direct current stimulation as a treatment for patients with fibromyalgia: a randomized controlled trial. *Pain*, 156(1):62–71, 2015.

[379] Felipe Fregni, Rafaela Gimenes, Angela C Valle, Merari JL Ferreira, Renata R Rocha, Luane Natalle, Riviane Bravo, Sergio P Rigonatti, Steven D Freedman, Michael A Nitsche, et al. A randomized, sham-controlled, proof of principle study of transcranial direct current stimulation for the treatment of pain in fibromyalgia. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, 54(12): 3988–3998, 2006.

[380] Angela Valle, Suely Roizenblatt, Sueli Botte, Soroush Zaghi, Marcelo Riberto, Sergio Tufik, Paulo S Boggio, and Felipe Fregni. Efficacy of anodal transcranial direct current stimulation (tdcs) for the treatment of fibromyalgia: results of a randomized, sham-controlled longitudinal clinical trial. *Journal of pain management*, 2(3):353, 2009.

[381] Rafael Polanía, Walter Paulus, and Michael A Nitsche. Modulating cortico-striatal and thalamo-cortical functional connectivity with transcranial direct current stimulation. *Human brain mapping*, 33(10):2499–2508, 2012.

[382] Wei-Ju Chang, Neil E O'Connell, Emma Burns, Lucy S Chipchase, Matthew B Liston, and Siobhan M Schabrun. Organisation and function of the primary motor cortex in chronic pain: protocol for a systematic review and meta-analysis. *BMJ open*, 5(11): e008540, 2015.

[383] Angela J Busch, Sandra C Webber, Mary Brachaniec, Julia Bidonde, Vanina Dal Bello-Haas, Adrienne D Danyliw, Tom J Overend, Rachel S Richards, Anuradha Sawant, and Candice L Schachter. Exercise therapy for fibromyalgia. *Current pain and headache reports*, 15(5):358, 2011.

[384] Giuseppe La Bianca, Massimiliano Curatolo, Marcello Romano, Brigida Fierro, and Filippo Brighina. Motor cortex trns ameliorates pain, anxiety, depression, quality of life and cognitive impairment in patients with fibromyalgia: preliminary results of a randomized sham-controlled trial.(p1. 219), 2017.

[385] Kurtuluş KÖKLÜ, Merve Sarigül, Zuhal ÖZİŞLER, Hülya ŞİRZAİ, and Sumru Özel. Handgrip strength in fibromyalgia. *Archives of rheumatology*, 31(2):158, 2016.

[386] Candice L Schachter, Angela J Busch, Paul M Peloso, and M Suzanne Sheppard. Effects of short versus long bouts of aerobic exercise in sedentary women with fibromyalgia: a randomized controlled trial. *Physical therapy*, 83(4):340–358, 2003.

[387] Jaeyoung Shin, Alexander Von Lühmann, Do-Won Kim, Jan Mehnert, Han-Jeong Hwang, and Klaus-Robert Müller. Simultaneous acquisition of eeg and nirs during cognitive tasks for an open access dataset. *Scientific data*, 5:180003, 2018.

[388] Eleonora Gentile, Katia Ricci, Marianna Delussi, Filippo Brighina, and Marina de Tommaso. Motor cortex function in fibromyalgia: A study by functional near-infrared spectroscopy. *Pain research and treatment*, 2019, 2019.

[389] Takuro Zama and Sotaro Shimada. Simultaneous measurement of electroencephalography and near-infrared spectroscopy during voluntary motor preparation. *Scientific reports*, 5:16438, 2015.

[390] Sangtae Ahn and Sung C Jun. Multi-modal integration of eeg-fnirs for brain-computer interfaces–current limitations and future directions. *Frontiers in human neuroscience*, 11:503, 2017.

[391] Richard C Oldfield. The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, 9(1):97–113, 1971.

[392] Naomi Kuboyama, Teru Nabetani, Ken-ichi Shibuya, Keishi Machida, and Tetsuro Ogaki. The effect of maximal finger tapping on cerebral activation. *Journal of physiological anthropology and applied human science*, 23(4):105–110, 2004.

[393] Rolf-Detlef Treede, Jürgen Lorenz, and Ulf Baumgärtner. Clinical usefulness of laser-evoked potentials. *Neurophysiologie Clinique/Clinical Neurophysiology*, 33(6): 303–314, 2003.

[394] Marina de Tommaso, Katia Ricci, Eleonora Vecchio, Marianna Delussi, Anna Montemurno, Giuseppe Lopalco, Florenzo Iannone, et al. Pain processing and vegetative dysfunction in fibromyalgia: a study by sympathetic skin response and laser evoked potentials. *Pain research and treatment*, 2017, 2017.

[395] Mark Cope and David T Delpy. System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Medical and Biological Engineering and Computing*, 26(3):289–294, 1988.

[396] Wesley B Baker, Ashwin B Parthasarathy, David R Busch, Rickson C Mesquita, Joel H Greenberg, and AG Yodh. Modified beer-lambert law for blood flow. *Biomedical optics express*, 5(11):4053–4075, 2014.

[397] Massimiliano Valeriani, Loic Rambaud, and François Mauguière. Scalp topography and dipolar source modelling of potentials evoked by co2 laser stimulation of the hand. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 100(4):343–353, 1996.

[398] Jong Chul Ye, Sungho Tak, Kwang Eun Jang, Jinwook Jung, and Jaeduck Jang. Nirs-spm: statistical parametric mapping for near-infrared spectroscopy. *Neuroimage*, 44 (2):428–447, 2009.

[399] Frank Leavitt and Robert S Katz. Cognitive dysfunction in fibromyalgia: slow access to the mental lexicon. *Psychological reports*, 115(3):828–839, 2014.

[400] Omid Rasouli, Egil A Fors, Petter Chr Borchgrevink, Fredrik Öhberg, and Ann-Katrin Stensdotter. Gross and fine motor function in fibromyalgia and chronic fatigue syndrome. *Journal of pain research*, 10:303, 2017.

[401] Catherine Mercier and Guillaume Leonard. Interactions between pain and the motor cortex: insights from research on phantom limb pain and complex regional pain syndrome. *Physiotherapy Canada*, 63(3):305–314, 2011.

[402] Marina De Tommaso, Maria Nolano, Florenzo Iannone, Eleonora Vecchio, Katia Ricci, Marta Lorenzo, Marianna Delussi, Francesco Girolamo, Vito Lavolpe, Vincenzo Provitera, et al. Update on laser-evoked potential findings in fibromyalgia patients in light of clinical and skin biopsy features. *Journal of neurology*, 261(3):461–472, 2014.

[403] L Garcia-Larrea, Maud Frot, and M Valeriani. Brain generators of laser-evoked potentials: from dipoles to functional significance. *Neurophysiologie clinique/Clinical neurophysiology*, 33(6):279–292, 2003.

[404] Catarina Mateus, Raquel Lemos, Maria Fátima Silva, Aldina Reis, Pedro Fonseca, Bárbara Oliveiros, and Miguel Castelo-Branco. Aging of low and high level vision: from chromatic and achromatic contrast sensitivity to local and 3d object motion perception. *PloS one*, 8(1):e55348, 2013.

[405] F Rémy, L Saint-Aubert, N Bacon-Macé, N Vayssière, E Barbeau, and M Fabre-Thorpe. Object recognition in congruent and incongruent natural scenes: A life-span study. *Vision research*, 91:36–44, 2013.

[406] Katherine K Mott, Brittany R Alperin, Phillip J Holcomb, and Kirk R Daffner. Age-related decline in differentiated neural responses to rare target versus frequent standard stimuli. *Brain research*, 1587:97–111, 2014.

[407] Gesine Marquardt. Wayfinding for people with dementia: a review of the role of architectural design. *HERD: Health Environments Research & Design Journal*, 4(2): 75–90, 2011.

[408] Enrico Denti. Novel pervasive scenarios for home management: the butlers architecture. *SpringerPlus*, 3(1):52, 2014.

[409] Robert M Chapman and Henry R Bragdon. Evoked responses to numerical and non-numerical visual stimuli while problem solving. *Nature*, 203(4950):1155, 1964.

[410] Samuel Sutton, Margery Braren, Joseph Zubin, and ER John. Evoked-potential correlates of stimulus uncertainty. *Science*, 150(3700):1187–1188, 1965.

[411] Jessica D Bayliss. Use of the evoked potential p3 component for control in a virtual apartment. *IEEE transactions on neural systems and rehabilitation engineering*, 11 (2):113–116, 2003.

[412] John Polich. Updating p300: an integrative theory of p3a and p3b. *Clinical neurophysiology*, 118(10):2128–2148, 2007.

[413] Marc Hofmann, Alexander Rösler, Wolfram Schwarz, Franz Müller-Spahn, Kurt Kräuchi, Christoph Hock, and Erich Seifritz. Interactive computer-training as a therapeutic tool in alzheimer's disease. *Comprehensive psychiatry*, 44(3):213–219, 2003.

[414] Paul Van Schaik, Anthony Martyr, Tim Blackman, and John Robinson. Involving persons with dementia in the evaluation of outdoor environments. *CyberPsychology & Behavior*, 11(4):415–424, 2008.

[415] Déborah A Foloppe, Paul Richard, Takehiko Yamaguchi, Frédérique Etcharry-Bouyx, and Philippe Allain. The potential of virtual reality-based training to enhance the functional autonomy of alzheimer's disease patients in cooking activities: A single case study. *Neuropsychological rehabilitation*, 28(5):709–733, 2018.

[416] Vitoantonio Bevilacqua, Gianpaolo Francesco Trotta, Claudio Loconsole, Antonio Brunetti, Nicholas Caporusso, Giuseppe Maria Bellantuono, Irio De Feudis, Donato Patruno, Domenico De Marco, Andrea Venneri, et al. A rgb-d sensor based tool for assessment and rating of movement disorders. In *International Conference on Applied Human Factors and Ergonomics*, pages 110–118. Springer, 2017.

[417] Dominique Twelves, Kate S M Perkins, Mcrp Uk, and Carl Counsell. Systematic Review of Incidence Studies of Parkinson' s Disease. *Mov. Disord.*, 18(1):19–31, 2003. ISSN 08853185. doi: 10.1002/mds.10305.

[418] Christopher G. Goetz, Barbara C. Tilley, Stephanie R. Shaftman, Glenn T. Stebbins, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Matthew B.

Stern, Richard Dodel, Bruno Dubois, Robert Holloway, Joseph Jankovic, Jaime Kulisevsky, Anthony E. Lang, Andrew Lees, Sue Leurgans, Peter A. LeWitt, David Nyenhuis, C. Warren Olanow, Olivier Rascol, Anette Schrag, Jeanne A. Teresi, Jacobus J. van Hilten, Nancy LaPelle, Pinky Agarwal, Saima Athar, Yvette Bordelan, Helen M. Bronte-Stewart, Richard Camicioli, Kelvin Chou, Wendy Cole, Arif Dalvi, Holly Delgado, Alan Diamond, Jeremy P. Dick, John Duda, Rodger J. Elble, Carol Evans, Virgilio G. Evidente, Hubert H. Fernandez, Susan Fox, Joseph H. Friedman, Robin D. Fross, David Gallagher, Christopher G. Goetz, Deborah Hall, Neal Hermanowicz, Vanessa Hinson, Stacy Horn, Howard Hurtig, Un Jung Kang, Galit Kleiner-Fisman, Olga Klepitskaya, Katie Kompoliti, Eugene C. Lai, Maureen L. Leehey, Iracema Leroi, Kelly E. Lyons, Terry McClain, Steven W. Metzer, Janis Miyasaki, John C. Morgan, Martha Nance, Joanne Nemeth, Rajesh Pahwa, Sotirios A. Parashos, J. S.Jay S. Schneider, Anette Schrag, Kapil Sethi, Lisa M. Shulman, Andrew Siderowf, Monty Silverdale, Tanya Simuni, Mark Stacy, Matthew B. Stern, Robert Malcolm Stewart, Kelly Sullivan, David M. Swope, Pettaruse M. Wadia, Richard W. Walker, Ruth Walker, William J. Weiner, Jill Wiener, Jayne Wilkinson, Joanna M. Wojcieszek, Summer Wolfrath, Frederick Wooten, Allen Wu, Theresa A. Zesiewicz, and Richard M. Zweig. Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders*, 23(15):2129–2170, 2008. ISSN 08853185. doi: 10.1002/mds.22340.

[419] Christopher G. Goetz, Werner Poewe, Olivier Rascol, Cristina Sampaio, Glenn T. Stebbins, Carl Counsell, Nir Giladi, Robert G. Holloway, Charity G. Moore, Gregor K. Wenning, Melvin D. Yahr, and Lisa Seidl. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: Status and recommendations. *Movement Disorders*, 19(9):1020–1028, 2004. ISSN 08853185. doi: 10.1002/mds.20213.

[420] Milica D. Djuric-Jovicic, Nenad S. Jovicic, Sasa M. Radovanovic, Iva D. Stankovic, Mirjana B. Popovic, and Vladimir S. Kostic. Automatic identification and classification of freezing of gait episodes in Parkinson's disease patients. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(3):685–694, 2014. ISSN 15344320. doi: 10.1109/TNSRE.2013.2287241.

[421] Evanthia E. Tripoliti, Alexandros T. Tzallas, Markos G. Tsipouras, George Rigas, Panagiota Bougia, Michael Leontiou, Spiros Konitsiotis, Maria Chondrogiorgi, Sofia Tsouli, and Dimitrios I. Fotiadis. Automatic detection of freezing of gait events in patients with Parkinson's disease. *Computer Methods and Programs in Biomedicine*, 110(1):12–26, 2013. ISSN 01692607. doi: 10.1016/j.cmpb.2012.10.016.

[422] Ilaria Bortone, Gianpaolo Francesco Trotta, Antonio Brunetti, Giacomo Donato Cascarano, Claudio Loconsole, Nadia Agnello, Alberto Argentiero, Giuseppe Nicolardi, Antonio Frisoli, and Vitoantonio Bevilacqua. A novel approach in combination of 3D gait analysis data for aiding clinical decision-making in patients with Parkinson's disease. In *Lecture Notes in Computer Science (including subseries Lecture Notes*

in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 10362 LNCS, pages 504–514, 2017. ISBN 9783319633114. doi: 10.1007/978-3-319-63312-1_44.

[423] Athanasios Tsanas, Max A. Little, Patrick E. McSharry, and Lorraine O. Ramig. Accurate telemonitoring of parkinsons disease progression by noninvasive speech tests. *IEEE Transactions on Biomedical Engineering*, 57(4):884–893, 2010. ISSN 00189294. doi: 10.1109/TBME.2009.2036000.

[424] Sabato Mellone, Luca Palmerini, Angelo Cappello, and Lorenzo Chiari. Hilbert-huang-based tremor removal to assess postural properties from accelerometers. *IEEE Transactions on Biomedical Engineering*, 58(6):1752–1761, 2011. ISSN 00189294. doi: 10.1109/TBME.2011.2116017.

[425] Dustin A Heldman, Alberto J Espay, Peter A LeWitt, and Joseph P Giuffrida. Clinician versus machine: reliability and responsiveness of motor endpoints in parkinson's disease. *Parkinsonism & related disorders*, 20(6):590–595, 2014.

[426] George Rigas, Alexandros T. Tzallas, Markos G. Tsipouras, Panagiota Bougia, Evanthia E. Tripoliti, Dina Baga, Dimitrios I. Fotiadis, Sofia G. Tsouli, and Spyridon Konitsiotis. Assessment of tremor activity in the parkinsons disease using a set of wearable sensors. *IEEE Transactions on Information Technology in Biomedicine*, 16 (3):478–487, 2012. ISSN 10897771. doi: 10.1109/TITB.2011.2182616.

[427] V. Bevilacqua, G.F. Trotta, C. Loconsole, A. Brunetti, N. Caporusso, G.M. Bellantuono, I. De Feudis, D. Patruno, D. De Marco, A. Venneri, M.G. Di Vietro, G. Losavio, and S.I. Tatò. *A RGB-D sensor based tool for assessment and rating of movement disorders*, volume 590. 2018. ISBN 9783319604824. doi: 10.1007/978-3-319-60483-1_12.

[428] Arash Salarian, Heike Russmann, Christian Wider, Pierre R. Burkhard, Françios J.G. Vingerhoets, and Kamiar Aminian. Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system. *IEEE Transactions on Biomedical Engineering*, 54(2):313–322, 2007. ISSN 00189294. doi: 10.1109/TBME.2006.886670.

[429] Houde Dai, Haijun Lin, and Tim C. Lueth. Quantitative assessment of parkinsonian bradykinesia based on an inertial measurement unit. *BioMedical Engineering Online*, 14(1), 2015. ISSN 1475925X. doi: 10.1186/s12938-015-0067-8.

[430] Robert I. Griffiths, Katya Kotschet, Sian Arfon, Zheng Ming Xu, William Johnson, John Drago, Andrew Evans, Peter Kempster, Sanjay Raghav, and Malcolm K. Horne. Automated assessment of bradykinesia and dyskinesia in Parkinson's disease. *Journal of Parkinson's Disease*, 2(1):47–55, 2012. ISSN 18777171. doi: 10.3233/JPD-2012-11071.

[431] Noël LW Keijsers, Martin WIM Horstink, and Stan CAM Gielen. Automatic assessment of levodopa-induced dyskinesias in daily life by neural networks. *Movement disorders: official journal of the Movement Disorder Society*, 18(1):70–80, 2003.

[432] Giovanna Lopane, Sabato Mellone, Lorenzo Chiari, Pietro Cortelli, Giovanna Calandra-Buonaura, and Manuela Contin. Dyskinesia detection and monitoring by a single sensor in patients with Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*, 30(9):1267–1271, 2015. ISSN 15318257. doi: 10.1002/mds.26313.

[433] Rachel Saunders-Pullman, Carol Derby, Kaili Stanley, Alicia Floyd, Susan Bressman, Richard B Lipton, Amanda Deligtisch, Lawrence Severt, Qiping Yu, Mónica Kurtis, et al. Validity of spiral analysis in early parkinson's disease. *Movement disorders: official journal of the Movement Disorder Society*, 23(4):531–537, 2008.

[434] Jerker Westin, Samira Ghiamati, Mevludin Memedi, Dag Nyholm, Anders Johansson, Mark Dougherty, and Torgny Groth. A new computer method for assessing drawing impairment in Parkinson's disease. *Journal of Neuroscience Methods*, 190(1):143–148, 2010. ISSN 01650270. doi: 10.1016/j.jneumeth.2010.04.027.

[435] C. Loconsole, G.F. Trotta, A. Brunetti, J. Trotta, A. Schiavone, S.I. Tatò, G. Losavio, and V. Bevilacqua. *Computer vision and EMG-based handwriting analysis for classification in parkinson's disease*, volume 10362 LNCS. 2017. ISBN 9783319633114. doi: 10.1007/978-3-319-63312-1_43.

[436] Eli Carmeli, Hagar Patish, and Raymond Coleman. The aging hand. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 58(2):M146—-M152, 2003.

[437] A W A Van Gemmert, H-L Teulings, Jose L Contreras-Vidal, and G E Stelmach. Parkinsons disease and the control of size and speed in handwriting. *Neuropsychologia*, 37(6):685–694, 1999.

[438] Peter Drotar, Jiri Mekyska, Zdenek Smekal, Irena Rektorova, Lucia Masarova, and Marcos Faundez-Zanuy. Prediction potential of different handwriting tasks for diagnosis of Parkinson's. In *E-Health and Bioengineering Conference (EHB), 2013*, pages 1–4. IEEE, 2013.