UNIVERSITA' POLITECNICA DELLE MARCHE

School of Engineering

Department of Information Engineering

Master of Science in Biomedical Engineering

Development of a deep-learning algorithm for autonomy evaluation in children with autism from RGB-D videos

Supervisor: Prof. Emanuele FRONTONI

Co-supervisor: Sara MOCCIA, PhD

Author:

Simone SALVONI

Academic Year 2019 - 2020

Abstract

Autism spectrum disorder (ASD) is a chronic childhood-onset neurodevelopmental condition with effects on adaptive functions throughout life [1]. The worldwide population prevalence for autism is 1%, increasing in the last decades. The underlying reasons for this increase are not fully understood [2].

The growing interest in ASD is due to the severe invalidation of affected subjects but also of the great conditioning of the relatives' lives which are the real caregivers for the autistic. The cost of autism, over the lifespan, is about double for an affected subject w.r.t. a person without this kind of disability [3]. The economic burden of ASD in US is of hundreds billion dollars projected to nearly double by 2025. People more able to communicate, care for themselves and participate in the workforce at greater levels, will need less financial support in their lives [4]. The average age for ASD diagnosis in the United States (US) stands at 5.7 years and the 27% of the subjects with autism remains undiagnosed at 8 years. ASD should be diagnosed in the early years of life because the affected subject, with a personalized therapy, shows encouraging improvements. This outcome has led the study of ASD in several fields, from genomics to biochemistry and physiology, passing through food science and psychology. Nowadays, the ASD diagnosis is based on assessing answers given by the subject's relatives during interviews focused on 3 developmental domains: communication and social interactions, restricted interests and behaviors, and stereotypical behaviors. Other approaches try to extract stereotypical motor movements patterns ASD-related from accelerometric signals, or

by video observation to analyze eye-gaze movements, the level of engagement and the emotional state of the children which can be associated to ASD syndrome.

In this context, the application of novel learning algorithms has gained a role in analyzing and infer over one or more types of information acquired by traditional methods to reduce the time requested for the autism assessment. These algorithms can extract more informative content from available datasets. Examples are: alternating decision tree (ADTree) [5, 6] and support vector machine (SVM) [7] in evaluating diagnostic interview (ADI); Naïve bayes and random forests [8] to determine ASD traits like developmental delay, less physical activity; neural networks, SVM and random forest to identify ASD patients using brain imaging; SVM [9] to analyze eye movements; convolutional neural networks(CNN) used to recognize stereotypical motor movements from accelerometric signals [10]. The applied behavioural analysis (ABA) is a science dealing to shape an individual's behavior [11], and is a standard in ASD therapy. In this work, a deep learning(DL) approach is proposed to evaluate the autonomy of autistic children in performing daily life activities, namely the hands-washing action. To the scope, a dataset of videos was acquired during the sessions of the ABA therapy at the facilities of "Il faro". The selected frames, have been annotated as belonging to class no-aid or to class aid, then used as input to the algorithms. The goal is to provide to the professionalities involved in the therapy, a tool that easily, rapidly and correctly gives them a picture of the subject's condition, reducing the assessment time and customizing the therapy. The proposed archictectures are 2 CNN, VGG16 and ResNet50. Both the nets were used in a from scratch version and in a pretrained version to achieve better results thanks to the knowledge acquired during the training on the ImageNet dataset [12]. Encouraging results were obtained even if the dataset was limited to 9700 frames. The fine-tuned ResNet50 was the best model with an accuracy of 0.83. To conclude, this work has shown that the use of DL methods with a simple acquisition setup, makes the evaluation faster and objective, allowing the therapy personalization.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

In this chapter, it will be introduced autism showing the clinical background on this neurodevelopmental disorder, its etiology and epidemiology. Then are reported the direct and indirect economic burden of the disease on the national healthcare systems. After the impact of this disorder on the lives of the affected subject and its family, the importance of an early diagnosis of ASD will be explained. The chapter ends with a description of the principal pharmacological and non pharmacological treatments actually used to alleviate symptoms and improve the overall autistic condition.

1.1 Clinical background on autism

In the World Health Organization (WHO) International Classification of Diseases eleventh edition (ICD-11), autism is reported as a neurodevelopmental disorder with this definition: *"Autism spectrum disorder is characterized by persistent deficits in the ability to initiate and to sustain reciprocal social interaction and social communication, and by a range of restricted, repetitive, and inflexible patterns of behaviour and interests"*. Deficits are sufficiently severe to cause impairment in personal, familiar, social, educational, occupational or other important areas of functioning. Individuals along the spectrum exhibit a full range of intellectual functioning and language abilities. Since the first accepted definition of autism, in the XXth century, many updates have followed. In the 5th edition of diagnostic and statistical manual of mental disor-

ders (DSM–5, American Psychiatric Association), four of the previous five pervasive developmental disorder were grouped under the definition of spectrum excluding the Rett's syndrome, now considered a discrete neurological disorder, and including the Asperger's syndrome (AS). Additionally, severity level descriptors were added to help categorize the level of support needed by an individual with ASD. Many studies report increased specificity and decreased sensitivity in the diagnosis using the DSM-5 respect to the DSM-IV criteria [1]. Respectively these terms refers to the percentage of healthy and sick people correctly identified as such. Then is higher the number of children whose ASD diagnosis is missed, particularly older children, adolescents, adults, or those with a former diagnosis of Asperger's disorder or PDD-NOS (pervasive developmental disorder-not otherwise specified). People diagnosed under the DSM-4, but not confirmed under the new DSM-5 appears to be declining over time, likely due to increased awareness and better documentation of behaviors [13]. The onset of the disorder occurs during the developmental period, typically in the early 3 years of age, but symptoms may not become fully manifest, until social demands exceed limited capacities. The economic impact associated with ASD is substantial and includes direct medical, direct non-medical and indirect productivity costs [14]. The lifetime cost of caring an individual with ASD and intellectual disability (ID) is $2.2 million in the US, and £1.5 million in the United Kingdom (UK), dropping to $1.4 million in the US and £0.92 million in the UK without co-morbid ID [15]. If unrecognized or untreated, ASD can contribute to poor educational attainment and difficulty with employment, leading to negative economic implications. The total economic impact of ASD in the US in 2015 (direct medical, non-medical and productivity costs combined) is $268 billion, from 0.9 to 2% of gross domestic product (GDP), expected to rise to $461 billion (from 0.99 to 3.6% of GDP) by 2025 [16]. Numbers comparable to diabetes, stroke and hypertension, separately considered.

The causes of ASD are not totally clear, but evidence suggests that a number of environmental and genetic factors are at play. A recent genetic analysis, combining several large population-based sources (more than 38,000 subjects), found genetic links between ASD and typical variations in social behavior and adaptive functioning [17]. This suggests multiple genetic risk factors as influencing the continuum of be-

havioral and developmental characteristics, with the extreme end of the continuum resulting in ASD. About 10% of children with ASD also have Down's syndrome or fragile X syndrome [18]. Parental history of psychiatric disorders (schizophrenia and affective disorders) has been linked to an increased risk for ASD. Other investigated risk factors are: parental age, premature born (before 33 week of gestation), low birth weight (<2500 g), fetal exposure to insecticides (chlorpyrifos) causing reduction in infant bodyweight and length, delayed psychomotor development, exposure of pregnant mothers to viral or bacterial infections promoting maternal immune activation (MIA, by 13%). MIA has been linked to increases in neuroinflammatory cytokines, abnormalities in synaptic protein expression and aberrant developments in synaptic connectivity, all of which underlie the pathophysiology of ASD. Exposure of pregnant mothers to psychotropic medication, to antiepilectics and anti-depressants (serotonin reuptake inhibitors) causes an 8-fold increased risk of developing ASD in the infant. Even after adjusting for maternal depression [19].

Neurobiological findings of behavioral functioning, report altered brain connectivity as a key feature of ASD pathophysiology. ASD is hypothetized as a disorder in long distance cortical and subcortical under-connectivity, with compensatory poorly formed shorter circuit over-connectivity. This altered connectiviy leads to the enhanced attention to simple stimuli with impairment in the sensory integration, of these stimuli, into a more complex perceptual representation. Children with ASD perform better than aged match peers in identifying the number of triangles in an Embedded Figure Test (EFT), but struggle when asked to identify the larger figure represented by the triangles. This dichotomy is reported also for auditory stimuli such as absolute pitch, a phenomenon that enables to replicate a single tone without any external reference. When trying to link singular pitches to form a melody, ASD individuals demonstrate poorer rhythmic entrainment when compared to non ASD controls. This reflects in language impairment of ASD because rhythm is a complex auditory perceptual skill, necessary for the discrimination of normal speech patterns. Sensory information forms the building blocks for higher-order social and cognitive functions and deficits in multisensory integration are critical for characterizing and understanding ASD [20]. This suggests that aberrations in successfully integrating multisensory processing are not

merely a feature of ASD, but the foundation upon which the other core symptoms develop. This inability to synthesize various sensory inputs, alters the ability to form metaphors and complex cognitive representations, necessary to effectively abstract, grasp language and respond to social cues, all representing the core symptoms of ASD. Over the 70% of autistic subjects presents another comorbid disorder:

- **Psychiatric**, attention-deficit hyperactivity disorder (ADHD), anxiety, bipolar disorder, obsessive-compulsive disorder (OCD), schizophrenia, irritability, aggressive behaviors, mood symptoms (depression), epilepsy, gender dysphoria, non-verbal learning disorder, sleep disorders.

- **Neurological**, inflammatory bowel disease, fragile X syndrome, intellectual disability, neuroinflammation and immune disorders, sensory problems, tuberous sclerosis, and Tourette syndrome, tic disorders [21, 22].

The prevalence of ASD has been steadily increasing in the past two decades. Starting from 1 in 2000 eight-year-old children in 1988, in 2000, the Center for Disease Control's Autism and Develop mental Disabilities Monitoring (ADDM) Network estimated the incidence of ASD to be 1 in 150 children. Further increasing has been recorded in 2006 (to 1 in 110 children), in 2008 (to 1 in 88), in 2012 (to 1 in 68 children), in 2016 (to 1 in 36 children) [23, 24]. This ratio is thought to be the same across all racial, ethnic or socioeconomic backgrounds, however, gender variations exist since males are affected 4-5 times more frequently than females. Increased ASD screening frequency in children and adults, better diagnostic criteria and more accurate behavioral and neuropsychological scales may, all, have contributed to the steady rise in the prevalence of ASD. In 2020 the estimated worldwide affected population is approximately 1%, approximately 70 million people living with ASD worldwide, the 85% living in developing countries. These data are underestimated due to insufficient population awareness, selection of studies and diagnostic capabilities, as well as cross-cultural appropriateness and comparability of the ASD screening, measurement and epidemiological data. The limited detection and screening of developmental delays in home, primary healthcare and education settings derive from:

- Parents not aware of the presence of developmental delays. Knowledge of ASD was also associated with a higher education level and school type [25, 26].

- Cultural stigma around neurodevelopmental disorders which discourage parents from seeking attention and delays the screening, even when suggested by a medical professional [27].

- Lacking knowledge among educators and primary care clinicians about ASD and how to assess them and the benefits of the therapies. Therefore, developmental delays may go unnoticed and not picked up at early signs [28].

- Limited number of clinical providers in pediatrics, psychiatry, neurology and psychology having the expertise to diagnose children with ASD.

- Challenges in widespread use of standardized diagnostic instruments which can be very expensive, require extensive and costly training (i.e. thousands of dollars per trainee), and be lengthy to administer [29, 30].

These aspects are particularly important in some low-resource and rural settings [31].

The ASD condition affects the quality of life for those with ASD as well as their families. Although some people with neurodevelopmental disorders can lead independent lives, for many, the impact of these conditions is severe. The disorder interferes with the productivity of their parents [32], who may experience increased anxiety and depression, and may need to decrease the worked hours outside due to their child's condition [33]. As such, evaluating the effects of ASD is a complex exercise that should include investigation of the effects on family members and caregivers, as well as on those people with the condition. It is possible to state, nowadays, that using a prudential approach to screen and eliminate other medical conditions, sufficient diagnostic tools are available to confirm an ASD diagnosis. Currently there is no cure for autism. The aim of current interventions is to affect developmental trajectories to lead children toward a more neurotypical outcome. This approach requires to coordinate services across health, education and social sectors [34].

1.2 Diagnosis and importance of early detection and intervention

In many cases, ASD is not identified until after four years of age, despite the 87% of these children had noted developmental concerns before age of three [23]. A major issue in screening for ASD is that it is a broad-spectrum condition [35]. The symptoms included marked impairment in non-verbal behaviors such as eye-to-eye gaze, facial expression, body postures, as well as stereotyped repetitive behaviors and loss of interest in social functions, communications and activities. Based on these criteria, a patient diagnosed with autism would have exhibited at least 3 deficits in the domain of social communication and, at least, 2 symptoms of restricted interests and repetitive behaviors. Social communication and interaction criteria, include problems in reciprocal social or emotional interaction, severe problems in maintaining relationships and nonverbal communication problems. Restricted and repetitive behaviors criteria include stereotyped or repetitive speech (repeated use of short, out of context, phrase), motor movements (repetitive hand motions) or use of objects, excessive adherence to routines, ritualized behavior, or excessive resistance to change, highly restricted interests, hyper or hypo-reactivity to sensory inputs or unusual interest in sensory aspects of environment. These symptoms must cause functional impairment for a diagnosis to be made [36]. High variability exists within the spectrum regarding to, for example, language or cognitive abilities, and severity of core symptoms. This heterogeneity is relevant for intervention planning and long term outcomes [37].

Studies show that with age (in general) the diagnosis of ASD remains stable, but adaptive functioning improves and co-morbid behavioral symptoms become less severe, whereas social functioning, cognitive ability and language skills have more variable outcomes. A valid, early and time-saving diagnostic process is clinically important for several reasons:

- To allow to families and children the necessary access to (early) intervention.
- To produce a minimum number of false positives, which avoids unnecessary interventions as well as unnecessary apprehension and fear among parents.

- To use resources effectively and allow for additional cognitive and language testing as well as diagnosis of co-morbid psychiatric and medical disorders.

1.3 Treatments

The ASD is a disease nowadays diagnosed evaluating dysfunctional behavioural aspects, as described in Chapter 2. It is already known that symptoms are the effects of underlying causes that are yet not completely understood and still investigated. These symptoms cover a broad range of areas, from neuropsychiatry to dietology and endocrinology. Looking to each different symptom to be faced and reduced, two broad types of current treatments have been identified, pharmacological and non-pharmacological, which are the core of the Sec. 1.3.1 and Sec. 1.3.2.

1.3.1 Pharmacological treatments

Pharmacological treatments are beneficial in improving co-morbidities. Showing variable levels of efficacy, this set of treatments provide relief from the disruptive repetitive behaviors of the ASD condition rather than modify the underlying disease process (i.e. neurodevelopmental abnormalities) [38, 39]. Moreover, they must be used carefully, especially with children, because of their side effects (i.e. metabolic deficits and sedation). More common treatments are:

- Psychostimulants drugs (methylphenidate and amphetamines) against hyperactivity and impulsivity.

- Atypical antipsychotic drugs (risperidone, aripiprazole, quetiapine, ziprasidone, and olanzapine) to reduce irritability and agitation.

- Antidepressant drugs (fluoxetine, sertraline, citalopram, escitalopram, and fluvoxamine) to reduce repetitive behaviors and improve anxiety and aggression.

- Alpha-2 adrenergic receptor agonists clonidine and guanfacine for treating aggressive behaviors, anxiety disorders, improving sleep disturbances.

- Cholinesterase inhibitors and NMDA receptor antagonists, for irritability and hyperactivity.

- Antiepileptic and mood stabilizers.

1.3.2 Non-pharmacological treatments

Non-pharmacological treatments refer to a broad domain of healing resources, known as "complementary and alternative medicine (CAM)". CAM refers to therapies not usually taught in medical schools or generally available in hospitals and include a broad range of practices and beliefs such as acupuncture, chiropractic care, relaxation techniques, massage therapy, and herbal remedies. It is also known as "complementary and integrative health (CIH)" and it is used in parallel to traditional medical practices, to augment traditional therapies outcomes. The National Institute of Health (NIH), groups CIH interventions into two main categories: natural products (biological ones, their use relies on a biological mechanism) and mind/body (or non-biological) practices. Biological CAM treatments usually include dietary interventions, vitamin supplements, and herbal remedies while non-biological CAM therapies are divided in three groups: mind-body medicine (i.e., auditory and sensory integration practices such as prayer, yoga, meditation, music, dance, and art in general), manipulative and body-based practices (i.e., massage, chiropractic care, and acupuncture), and energy medicine (i.e. homeopathy) [40, 41].

1.3.2.1 Biological CAM treatments

- **Omega-3 fatty acids** are essential polyunsaturated acids, derived mainly from fish (the eicosapentaenoic acid (EPA) and the docosahexaenoic acid (DHA)) or grains (the alpha-linolenic acid (ALA)). These fatty acids play a central role in neurodevelopment, including both structural and functional effects on neurotransmission, oxidative stress, inflammation and immunity. Few open trials have provided evidence that oral supplementation of Omega-3 fatty acids improved social, behavioral and attention deficits in young ASD patients [42]. In contrast, randomized control trials have found no evidence for the efficacy of omega-3 fatty

acid supplementationon improvement of ASD symptoms [43]. While generally safe, these supplements are associated with gastrointestinal side effects associated with behavioral symptoms including irritability, hyperactivity, stereotypy and social withdrawal [44].

- **Herbal remedies** such as Gingko biloba, Zingiber officinal (ginger), Astralagus Membranecaeus, Centella asiatica (gotu cola), and Acorus Calamus (Calamus) may have therapeutic benefits in ASD patients based on their somatic effects including increasing cerebral circulation, enhancing cognitive functions, exertion of a calming or sedative effect, and enhancing immune response. One recent systematic review concluded that while being safe, herbal medicines, when used in combination with conventional therapy, showed promising results in improving abnormal behaviors and inattention in ASD patients. Among herbal remedies, a recent study from Chan et al. [45] investigated the efficacy of intranasally administered Borneol and Borax (two herbs which in Chinese traditional medicine were thought to enhance cognitive abilities) showing a reported higher flexibility in problem solving, greater attention, and planning capacities. Yokukansan, a Japanese herbal remedy used for restlessness and behavioral symptoms of dementia, was tested in a 12-week, open label trial [46] on 40 subjects with Asperger syndrome or PDD-NOS. 90% of the sample showed a clinically significant response, and no serious adverse event was reported (only mild nausea in five patients).

- **Nutritional supplements** are considered since ASD patients are at an increased risk of malnutrition due to the very specific food preferences such as too wet/dry, the color of food, food shape, packaging type and brand, which causes lowered energy intake. The lower intake of energy is also caused by intestinal dysfunction, indigestion, and bad absorption of nutrients [47]. The nutritional status should always be assessed in ASD patients to rule out any nutrient deficiency and to plan interventions using nutritional supplements to correct evenctual deficiencies. Biochemical processes involved in ASD symptoms include:

  - Vitamin-A (VA) plays a role in regulation of gut microbiota. Except re-

stricted interests, all the symptoms regarding neurodevelopment deficits significantly improve following VA supplementation [48].

- Vitamin-C decreases the level of oxidative stress and the severity of stereotyped behaviors and other autism symptoms [49].
- Vitamin-D, neuroactive hormone for normal brain homeostasis and neurodevelopment, significantly improved outcome, which was mainly in the sections of the Child Autism Rating Scale (CARS) and aberrant behavior checklist subscales that measure behavior, stereotypies, eye contact, and attention span [50].
- Vitamin-E, an important antioxidant, plays a major role in the protection and development of the embryo nervous system, supplementation reduces brain oxidative stress [51].
- Vitamin-B6, B12, whose deficiencies cause symptoms of attention deficit, hyperactivity, diverse behavioral processes, including sleep, learning, memory, sensation of pain, epilepsy. Its supplementation improves symptoms of ASD as the rated Clinical Global Impression Scale of Improvement (CGI-I) score was statistically significantly better (lower) [52].
- Vitamin-K, has a role in neural development.
- Multivitamins supplement, containing several vitamins, minerals, iron, folic acid (helps make DNA and other genetic material, it is especially important in prenatal health), and antioxidants such as coenzyme Q10 and n-acetylcysteine) was chosen as active treatment with recorded improvement in parent-rated scores of irritability.

Even if a recent review of clinical trials assessing their effects failed to provide convincing evidence for the therapeutic benefits of these agents on ASD core symptoms, supplementation of nutritional deficiencies may be necessary to overcome any malnutrition side effects in ASD patients.

- **Dietary interventions**, consisting in the use of a specific dietary regimen, rely on the absence of specific food allergen (such as casein or gluten) which could en-

hance immune response in predisposed subjects or trigger autoimmunity. Several gastrointestinal abnormalities have been observed in subjects with ASD, such as increased permeability of the gut barrier and bacterial overgrowth which could benefit from elimination diets aimed at omitting foods with negative effects on ASD symptoms. The ketogenic diet is low-carbohydrate (10%), high-fat (90%) diet which has been successfully administered in children with refractory epilepsy showing some effect on improving social and communication skills in ASD children: this dietary regimen determines a better seizure control and has an effect comparable to antiepileptic drugs [53, 54]. An elemental formula (containing free amino acids-Neocate) diet, with exclusion of all milk products, after 4 months, reported a significant reduction of hyperactivity. Among diets, the Chanyi approach suggests to decrease the intake of some foods (like meat and fish, eggs, ginger, garlic, and onion) which are thought to produce higher internal heat and exert a negative impact on the child's mood and cognitive functions. A double blind randomized study in which 24 ASD children were assigned either to a specific diet modification based on the Chanyi approach or to their usual diet for one month, have reported a significant improvement in parent-rated social problems and repetitive behaviors [55]. Dietary approaches may be important contributors to the increased overall well-being of ASD patients.

- **Nutraceutical** is defined as "any substance that is food or a part of food and provides medical or health benefits, including the prevention and treatment of disease" [56]. Usually consisting of dietary supplements (such as vitamins, minerals, amino acids, and herbal substances) or functional food, nutraceuticals could represent a potential treatment for autism with limited or no side effects. To date, the only functional food tested in autism is **camel milk** which contains less cholesterol and lactose and more vitamins and enzymes, such as the peptidoglycan recognition protein (PGRP), than cow milk. This milk plays a role in preventing food allergy, modulating the immune system, and has showed significant improvement in CARS scores and in antioxidant activity in children treated either with raw or boiled camel milk for 2 weeks compared to placebo. **L-Carnosine**, is another CAM therapy tested in autism, is based on the con-

nection between carnosine and GABA functioning, which seems to be altered in ASD [57]. Particularly, it could alter neurotransmission by interacting with zinc and copper at GABA receptor level [58], improving receptive speech and social behavior, with no side effect (apart from rare hyperactivity which disappeared after lowering the dose). **Natural flavonoids**, precisely quercetin and luteolin, exert a powerful antioxidant activity and have a low redox potential which could in turn be useful in autism wherein altered redox status and concomitant subclinical inflammation has been reported [59]. Significant changes in adaptive functioning and aberrant behaviors were observed. The most relevant adverse event was irritability, which was experienced by half of the sample usually at the beginning of therapy (1–8 weeks). **Probiotics** are living microorganisms which could exert health benefits on the host. Generally, they are bacteria belonging to two groups, Lactobacillus or Bifidobacterius. In recent years, the gut-brain reciprocal influence has obtained much relevance in autism since gut inflammation, or altered microflora, could determine a detriment on brain development and function [60]. Outcomes reported a significant improvement in core symptoms of autism, such as eye contact and correct recognition of human emotion. **Digestive enzymes** are another treatment of gut abnormality consisting of three plant-derived enzymes (peptidase, protease 4.5, papain). In a double blind, placebo controlled, randomized, crossover trial, enzymes were administered for 3 months [61]. The ASD group receiving digestive enzyme therapy had statistically significant improvement in emotional response, general impression autistic score, general behavior and gastrointestinal symptoms, demonstrating the usefulness of digestive enzymes which are inexpensive, readily available and have an excellent safety profile [62].

- **Homeopathic products** come from plants (such as red onion, arnica mountain herb, poison ivy, belladonna deadly nightshade, and stinging nettle), minerals (such as white arsenic), or animals (such as crushed whole bees). Homeopathic products are often made as sugar pellets to be placed under the tongue; they may also be in other forms, such as ointments, gels, drops, creams, and tablets. Treatments are “individualized” or tailored to each person so it’s common for

different people with the same condition to receive different treatments.

- **Hyperbaric Oxygen Therapy (HBOT)** is generally used to treat carbon monoxide poisoning or air embolism. The exact mechanism of action is not yet fully understood but HBOT seems to exert positive effects on different neurological symptoms [63]. It is of note that both groups seemed to improve from baseline. A small open label trial reported improvement in several symptoms of ASD [64, 65, 66, 67].

- **Chelation treatment** involves administration to an individual of various chemical substances for the purpose of binding and then withdrawing specific metals from the person's body [68]. Conducted double blind randomized trials did not demonstrate any significant evidence supporting the utility of chelation treatment in ASD [69]. An open label trial in which children underwent chelation and antiandrogen therapy reported significant improvement, but the study design and the multicomponent intervention refrained to draw solid conclusions [70].

1.3.2.2 Non-biological CAM treatments

- **Music therapy** is efficacious in ASD due to its ability to potentially change both the structure and functional connectivity of the cortex. This changes allow for multisensory integration across cortical and subcortical domains in early developmental stages, the absence of which is the core of aberration in ASD [71]. The role of music as a treatment for psychiatric conditions (i.e., depression, schizophrenia, substance dependence and abuse disorder, and dementia) has been studied for many years, for its effectiveness in physical recovery, cognitive improvement, communication skills, and social and emotional rehabilitation [72]. Musical improvisation in autism could represent a sort of nonverbal shared language that could enable both verbal and nonverbal patients to reach communication [73]. In fact, it has been reported that the learning of language in infants is highly based on the musicality of sounds [74]. ASD children appeared to respond better to music than to spoken words [75]. The use of songs could help people with ASD to understand emotion which they have difficulties in detecting

in words. Several trials outcomes evidence that music therapy may help children with ASD to improve their skills in areas like social interaction, verbal communication, initiating behavior, and social-emotional reciprocity. A study showed significant improvement in several standardized scales (CGI-S, CGI-I, and BPRS) [76]. The cohort was divided into two groups of severity and the study demonstrated that while simple music was more effective in severe ASD patients' joint attention, complex music was more effective in children with mild or moderate autism. Patients diagnosed with ASD generally show an activated bilateral temporal brain networks during sung-word perception and functional front-temporal connectivity, disrupted during spoken-word perception, is preserved during sung-word listening. There is also synchronization of frontal activity with activity in posterior and other areas. Both passive listening and active playing, activate areas of the brain involved in cognitive, sensorimotor and perception-action mediation through increasing the oscillation synchrony between these cortical areas. This sinchrony promotes heightened sensory-integration. Short-term and long-term music listening and music playing, involve multisensory and motor networks and create connections between functionally related brain regions with continued music exposure. Children engaged in long-term instrumental practice have larger corpus callosum, frontal, temporal and motor areas relative to controls. Previous evidence demonstrate a change in the volume and fiber density of the arcuate fasciculus in not only professional musicians but also in adult patients with Boca's aphasia. Both, musicians and aphasic subjects demonstrated both clinical improvement and structural changes in this front-temporal tract following an intensive course of music-based speech therapy. Music improve deficits in the mirror neuron system (MNS) in children with autism priming them to speak. Music's ability to modulate emotion and mood in individuals both with and without ASD is well established: both experimentally and anecdotally. Both, the ASD and the healthy subjects, show a preserved, and even heightened, sense of musicality that extends into adulthood, with an ability to interpret and respond to the emotions conveyed in songs or music even when unable to do so in speech. All the mentioned evidences design a role for music-based therapies to

recover these deficits.

- **Cognitive behavioral therapy (CBT)** is a psychotherapeutic intervention previously evaluated to be effective for ASD patients for targeting core symptoms improving affective communication, social skills, cognition and facial emotion perception, but also for comorbid anxiety, depression and obsessive-compulsive disorder. Anxiety often leads to several other problems including increased irritability, disruptive behaviors, inattention, and decreased functionality. These programs always include elements of psychoeducation and social coaching to develop social skills, self-care skills, highly structured worksheets and visual aids. CBT techniques and protocols have been introduced for ASD patients to address difficulties arising from their problems in recognizing and understanding both their own and others' thoughts and feelings. CBT can be administered individually, can involve the presence of family members and/or can be done in a larger group setting. While individual treatment is more effective due to its flexibility and personalization to needs of the individual patient, group CBT allows for increased social interaction, sharing of experiences, promotion of self-acceptance and improved insights of both strengths and impairments related to ASD.

- **Social behavioral therapy (SBT)** in ASD focuses on functional independence and quality of life by targeting improvement of emotional regulation, social skills and communication. Different types of SBT interventions include both specific targeted approaches focusing on each symptom domain, as well as more complex and comprehensive approaches. Comprehensive intensive behavioral interventions are based on Applied Behavioral Analysis (ABA), an approach that evaluates the impact of environmental events on behavior and employs structured specific teaching methods focusing on language, cognitive, sensorimotor skills, social interactions, everyday living skills and specific problem behaviors. This is nowadays the main used therapy in autism. Such comprehensive ABA programs include early intensive behavioral intervention (EIBI) in children under 5 years old. EIBI works by decomposing complex skills into more elementary subskills, teaching them individually, increasing intellectual functioning. Other

ABA programs include the Learning Experiences: the Alternative Program for Preschoolers and Parents (LEAP), focuses on integrated teaching with non ASD peers. SBT also includes developmental intervention models, which conduct an evaluation based on developmental history, assessment of functioning, and clinical observations of interactions to yield a developmental profile for every patient and design targeted interventions. Examples of such models include the Denver Model and the Early Start Denver Model (for toddlers) (ESDM) [77], Responsive Teaching [78], and the Developmental Individual-Difference Relationship-based model (DIR/Floortime) [79]. Besides comprehensive programs, there are also targeted interventions focusing on specific cognitive skills and domains tailored individually to the existing level of functioning, skills and needs. Such interventions include enhancement of functional communication, emotion recognition, social skills, and promoting independence [80]. Successful programs include the Picture Exchange Communication System (PECS) [81], the use of speech generating devices, self-management [82] or Reciprocal Imitation Training (RIT).

- **Oxytocin (OT) and Vasopressin(V)** are closely related neurohypophyseal nine-amino acid peptide hormones, differing in only 2 aminoacids, synthesized in the hypothalamus and stored in the neurohypophysis. OT and V receptors are highly expressed in the amygdala, hippocampus, and nucleus accumbens, and play a role in social behaviors, bonding, and parental care. Several studies detected lower OT plasma levels and parallel higher levels of its precursors in ASD patients [83, 84] as well as associations between ASD and genetic variation in OT receptors. Several randomized clinical trials or open label studies have evaluated the effects of OT in ASD using various dosing regimens and behavioral and cognitive outcome measures. One trial reported that acute intravenous administration of OT decreased repetitive behaviors and enhanced social cognition in ASD patients [85], while another trial found that acute intranasal OT administration improved social cognition [86]. Functional neuroimaging studies reported increased activation in brain regions involved in processing of social information and reward processing in ASD patients after OT administration. An important limitation of OT treatment is its intranasal vs. intravenous administration routes

by which only a small fraction reaches the brain, and therefore the majority of the administered agent can trigger peripheral side effects (nasal discomfort, skin irritation, diarrhea, irritability and fatigue). Vasopressin enhanced responses to social communications and interactions suggest that vasopressin V1a receptor antagonists may exert pro-social benefit for disorders where social and emotional functions are core deficits. A limited number of clinical trials have evaluated the efficacy and safety of V1a antagonist in ASD. The Food and Drug Administration (FDA) in USA has recently granted this compound "Breakthrough Therapy Designation", raising the hope for approval of the first pharmacotherapy to improve core social and communication deficits in ASD.

- **Sensory Integration Therapy**. ASD affected subjects often display impairments in sensory information processing resulting in overwhelming situations when they are solicited with lights, sounds, smells, tastes, or textures [87]. Sensory integration therapy commonly uses activities specifically studied to modulate how the brain responds to sight, touch, sound, and movement [88]. All studies yielded significant improvement in several autistic core symptoms (communication, social reciprocity and motor activity).

- **Dance Therapy**. Dance and movement therapies are based on the mirroring of the movements performed by the therapist, focusing more on "attunement" than on simple imitation, to achieve a more mature form of social reciprocity [89].

- **Acupuncture** is a form of traditional chinese medicine widely used also in western countries and consists in placing needles in the skin and near tissues in specific points, known as acupuncture points. The needle could convey also electricity (electro-AP) or laser or heat [90].

- **Massage** is used because the touch alleviates sensory impairment (hypo/hypersensitivity) and reduce anxiety [91]. Significant increase in socialization and communication and a reduction of sensory impairment were observed in different trials.

- **Yoga** is a movement therapy which could ameliorate behavioral problems and anxiety. It is of note that yoga appears to increase GABA brain levels, even after

one session [92]. As GABA is considered to play a key role in autism pathogenesis, yoga may represent a potential treatment candidate involving simple body movements which has to be performed in a relaxed and natural manner. Study findings reported increased self-control, social interaction, control of disruptive behaviors and reduced parent-rated autistic symptoms [93].

- **Pet Therapy**. The use of animals in ASD relies on the hypothesis that animal movements and behaviors are more predictable and repetitive and could help children with ASD to interpret social cues even in more subtle contexts. Traditional chosen animals are dogs and horses but in 2014, also Guinea pigs were used with 64 children with ASD [94] and the authors reported significant improvement in social functioning compared to the control situation.

- **Chiropractic manipulation** is a popular and widely used CAM in ASD, focused on the relationship between the body's structure, primarily of the spine, and function. It makes use of a type of hands-on therapy called manipulation (or adjustment) as their core clinical procedure [95]. Examples of different types of chiropractic care are: the Atlas Orthogonal Upper Cervical Spinal Manipulative Therapy (a form of manipulation involving the instrumental percussion of the atlas to correct possible misalignments) and the full-spine Spinal Manipulative Therapy (characterized by high velocity and low amplitude thrusts) in children with autism. The Atlas Orthogonal Group has showed the major improvement.

Chapter 2

LITERATURE REVIEW

In this chapter a review of the literature on the approaches for ASD evaluation is presented. Sec. 2.1 describes the methods considered as standards in the clinical practice of autism diagnosis, while Sec. 2.2 reports the more recent approaches. After a brief resume of the limitations in the state of the art in Sec. 2.3, the survey prosecutes with Sec. 2.4, introducing the new paradigm of learning algorithms applied to the evaluation of autism. The chapter ends explaining the objective of this work with a brief thesis overview in Sec. 2.5.

2.1 Approaches to evaluate the degree of ASD

ASD are, as a rule, diagnosed by trained clinicians via direct behavioural observation or interviews with parents/primary caregivers or adult patients, or both. The 3 main areas of investigation are:

- Communication and language skills, that is child's history of speech development and current abilities to sustain the conversation.
- Social interaction issues, how the child interacts with other people and how show or interpret emotional responses.
- Repetitive and obsessive behaviors, also called stereotypical behaviors such as an obsession on unusual items, repetitive hand motions, or repeated use of short,

out-of context phrases.

The behavior is a complex ensemble of more basic gestures, gazes, vocalizations and postures. Depending on the area under investigation, the autistic patterns are individuated consequently but the ways (i.e. recorded signals) by which the subjects are monitored could be the same for more than one area. Nowadays the observation is accomplished by mean of the direct acquisition of signals such as accelerometric, audio, gaze tracking, skin impedance, but also using video capturing, offline coding, and analysis of the children's condition. These "signals" are then used to infer about the behavioral state of the children. The aim of the observation is the understanding of the level of engagement in social relations and other skills. Novel approaches use computational methods such as learning algorithms to analyze and infer over one or more types of information acquired by the above methods to speed up the evaluation process, making it more objective and allowing for a personalization of the therapy.

2.1.1 Paper-and-pencil rating

Interview based scales are collections of questions, concerning the different ASD symptoms and the possible comorbidities, posed to the parents/caregivers. To the received answers are assigned scores ranging from 0 to 3 (increasing severity) plus additional scoring possibilities [96]. Scales are different for several developmental stages (from early childhood to adulthood) and are helpful in differentiating between ASD and other developmental disorders. They are divided into 4 main categories:

- Interview with parent/primary caregiver:
  - 3di - Developmental Dimensional and Diagnostic Interview [97]
  - ADI-R - Autism Diagnostic Interview - Revised [98]
  - ASDI - Asperger Syndrome Diagnostic Interview [99]
  - DISCO - Diagnostic Interview for Social and Communication Disorders [100, 101]
  - ASDDA - Autism Spectrum Disorder-Diagnosis Scale for Intellectually Disabled Adults [102]

  - ABI - Autistic Behavior Interview [103]

- Interview with the affected adolescent or adult:

  - AAA - Adult Asperger Assessment [104]

- Direct behavioural observation by a trained clinician:

  - ADOSG - Autism Diagnostic Observation Schedule - Generic [105]
  - ASDOC - Autism Spectrum Disorder - Observation for Children [106]
  - BOS - Behaviour Observation Scale for Autism [107]

- Combination of interview and direct observation:

  - CARS-2-ST - Childhood Autism Rating Scale - Second Edition -Standard Version [108]

These rating scales require long time to be acquired and are limited by the subjectivity of the interviewed subject and, in some cases, of the interviewer that must interpret the answers.

2.1.2 Direct acquisition of signals

In ASD, anxiety and poor stress management can intensify social interaction difficulties, increase levels of ritualized or repetitive behaviors, and magnify irritability and aggression. Monitoring the physiological changes, associated with negative emotions, can give to care-givers insights into the internal emotional changes in a real-time fashion, allowing them to take necessary actions to alleviate the symptoms and to manage the stress. Due to the multiple ways by which stress is manifested, the evaluating process makes use of different measurable physiological parameters of the body [109]:

- Heart Rate Variability (HRV) is the oscillation in the interval between consecutive heart beats.

- Respiration rate (RR) is the total number of respiratory cycles occurring each minute and is a useful indicator for stress.

- Electrodermal activity (EDA) refers to measures of the conductance of the skin surface by using one electrode to injects a small AC current into the skin and another to calculate the impedance of the skin using Ohm's Law given a certain voltage. It can be a useful indicator of stress [110].

- Skin temperature (ST) refers to the temperature measured on the surface of the skin, including body and peripheral temperature is a parameter of stress, although this depends on the location of temperature measurement.

- Cortisol is a hormone released by the adrenal gland when the person is exposed to particular external stimuli that causes stress and can be measured conventionally from saliva. There are also wearable patches attached to the skin for measuring cortisol level non-invasively [111].

- Blood pressure can monitor changes in the emotional state due to stress since, when exposed to a stressful situation, the body produces hormones which increase the blood pressure.

- Blood Volume Pulse (BVP), measured by photoplethysmography (PPG), indicates dynamic changes in blood volume underneath the sensor. These oscillations reveal changes in the vascular bed due to vasodilation or vasoconstriction (increase or decrease in blood flow) and to changes in the elasticity of the vascular walls, both related to stress [112].

- Blood Oxygen Saturation refers to the extent to which hemoglobin is saturated with oxygen which is altered by stress.

- Electromyography (EMG) records electrical activity in response to a nerve's stimulation of the muscle. Muscle activity of certain body parts such as face, shoulder, and lower back exhibit increased EMG activity during stress [113].

Recently developed stress monitor solutions have been implemented which exploit both physiological data and other additional informations such as physical activity, and sleep data for a comprehensive assessment of stress and the associated activities [114]. The emergence of wearable assistive technologies provides a mean to detect emotional

arousal and corresponding changes in the autonomous nervous system non-invasively. Imani et al. [115] have developed a wearable sensor that includes, in a patch, sensors monitoring physical exertion (via a lactate sensor), and electrocardiogram. This hybrid device misses the algorithms to recognize the emotional states from such signals. Yoon et al. [116] have developed a flexible patch detecting skin conductance, temperature, and arterial pulse wave. The sensor has a small contact area (25 mm×15 mm) and flexible material. It is useful for people with ASD as they have hyper-sensory issues. The device should be embedded into their clothing or by wearable technology (e.g., wristbands and smart watches).

A smart scarf has been developed by Guo et al. [117] which uses a heart rate sensor and a Electrodermal Activity (EDA) sensor to recognize emotional information. It responds to negative emotions when detected, by changing its color and emitting an odor to promote positive emotions. The authors did not develop the algorithms to recognize relevant affective emotional states.

A smart glove that is designed for ASD population was developed by Koo et al. [118]. The glove included an EDA sensor and pulse oximeter sensor (heart rate/heart rate variability). The EDA sensor was made of conductive thread sewn into the glove to make electrical contact with the skin. The glove incorporated a wireless module enabling a remote monitoring and notification for individuals with ASD and their parents.

More recently, solutions are under development that utilize advanced artificial intelligence technologies such as machine learning and deep learning for emotion recognition to make meaningful information out of the collected physiological data.

Another device that is still under development and targeted for people with ASD is reported in [109]. This device collects heart rate, electrodermal activity and skin temperature data. The solution integrates a patented technology, called Anxiety Meter, to assess anxiety level in a natural setting and can notify the caregiver when anxiety levels start to elevate. By applying data analytics techniques, the device allows to make “smart” clinical decisions.

Another class of solutions aim monitor stress level but also manage and reduce stress utilizing stimulating electrical pulses. This solution uses a patented technology

named Bi-lateral alternating stimulation–tactile (BLAST) [119]. Muaremi et al. [120] have included contextual information such as voice, physical activity, and sleep data for comprehensive assessment of stress and the associated activities. A new generation of stress management and stress relief devices uses two neuro-stimulation technologies, namely Transcutaneous Electrical Nerve Stimulation (TENS) and Transcranial Direct Current Stimulation (tDCS) [121]. This technology is currently being validated for humans. As can be seen, some technologies are in their early stage of testing and not thoroughly validated and will require extensive validation to show their usefulness in clinical studies. In some of the cited studies there is the lack of a proper algorithm that uses the recorded signal to predict the emotional state and, by this information, is able to evaluate the presence of ASD. Other of these studies lack of a methodologically reliable experimentation. Anyway, adopting such wearable devices can potentially carry enormous benefits for people with ASD and their caregivers.

Screening for ASD has evolved recently from subjective clinical assessments to objective metrics acquired from sensing devices such as gaze-tracking, motion sensors, and speech analytics. One domain in which autistic people behave unusually relates to oculomotor behavior, including low levels of eye contact during communication, and low levels of directional signaling via eye gaze [122]. A study conducted by Frazier et al. [123, 124] confirmed that an ASD child avoids looking at the faces, and specifically, the eyes. Aggregating gaze dwells time to social and non-social Region of Interests, strongly discriminated children with ASD from those without ASD. More researches have been conducted using variety of commercially available eye tracking systems with similar experimental setups and procedures. Using visual stimuli (e.g., still images or dynamic videos) with predefined Region of Interests such as face, eyes and mouth. Tracking the target's gaze patterns through a device, atypical patterns in the gaze behavior are expressed using the subject x and y coordinates of gaze fixations with respect to time. Approaches used in these studies ranged from facial viewing patterns to video viewing patterns. There are three main types of eye tracker: screen-based (also called remote or desktop), glasses, (also called mobile) and eye tracking within virtual reality headsets.

The display of stereotypical body movements is a symptom of ASD and the iden-

tification of movement patterns has been focusing on data from accelerometer sensors. The accelerometric signals are used to detect hand-flapping movements, body rocking, fingers flapping, hand on the face and hands behind back. Even coarser movement indices (including, for instance, gesture indices or general movement patterns) also provide meaningful information for the identification and differentiation of autistic behavioral markers.

Prosody collects those elements of speech that are not individual phonetic segments (vowels and consonants) but properties of syllables and phrases, including intonation, tone, stress, and rhythm. Prosody may reflect various features of a subject: its emotional state, the form of the utterance (statement, question, or command), the presence of irony or sarcasm, emphasis, contrast, and focus. It is usual to distinguish between auditory measures (subjective impressions of the listener) and acoustic measures (physical properties of the sound wave). In auditory terms, the major variables are:

- The pitch of the voice (varying between low and high)
- Length of sounds (short and long)
- Loudness, or prominence (soft and loud)
- Timbre (quality of sound)

In acoustic terms, these correspond closely to:

- Fundamental frequency (hertz, or cycles per second)
- Duration (time units such as milliseconds or seconds)
- Intensity, or sound pressure level (measured in decibels)
- Spectral characteristics (distribution of energy at different parts of the frequency range)

In the context of autism, researchers focused on deficits in vocal emotional communication and pointed out that these deficits tend to affect voices exactly as faces and body movement. Deficits in verbal communication may also affect vocal identity perception, and vocal expression of autistic people in communication could be affected

beyond emotional expressions. Detecting auditory markers for autism in voices is linked to the symptom of vocal stereotypies. Usually differentiation is done between vocal stimming (a nonverbal vocalization often observed in autism) and other noises using dedicated dictionnaires. ASD children lacking verbal communication present vocal stimming and frustration. For verbal children, other potential vocal markers could include prosody [125]. Marchi et al. [126] created an evaluation database in three languages with ASD and typical developing (TD) children's emotionally toned voices recordings during an imitation task for sentences with different intonations (e.g., rising, falling). The major drawbacks of the methods described in this section is that they are invasive and conditionating the spontaneous behaviour of the subject. In some cases the described methods require the accomplishment of unusual actions such as, for example, looking to a monitor for a prolonged time.

2.1.3 Video-based methods

The use of videos in autism belongs to two broad categories:

- Video based ASD identification. In many cases it has been demonstrated low agreement between parent report and more objective measures of ASD symptoms, togheter with a lower reliability for screening instruments when used in rural, low income, less educated, and racially diverse samples. This is due to bad comprehension and interpretation problems of the queried constructs or inadequate knowledge of developmental milestones. The Video-referenced Infant Rating System for Autism (VIRSA) is a complementary tool employing a large library of video clips depicting a wide range of social-communication ability and relying solely on video in the ratings, with no written descriptions of behavior. Video segments were rated by 9 clinical research staff on a scale from 1 (least socially competent) to 10 (most competent). The semantic clarity of the videos and their grouping per age, improve early discrimination of infants at highest risk for ASD, allowing for a few minutes for the test completion (7 to 10) and concurrent symptoms indexing.

- Video based ASD interventions [127], facilitate the participant with a definitive

appropriate model, and appropriate set of desired behaviours without additional stimuli causing confusion. Its use is based on the fact that human learning often comes from observing and imitating a skilled person or 'model', proficient in performing the skill adequately. Children with ASD have specific problems with this type of learning due to poor motor imitation, reduced interactions with peers, inappropriate 'frame of reference', lack of attention and eye contact that cause misunderstanding of how and when it is appropriate to apply the behaviour. VBI include:

- Video feedback (VF), the individual is recorded and can review appropriate or inappropriate behaviours while the experimenter provides direction and assistance in the modification. VF aims to develop self-perceptions and improve peer interactions and behaviours.
- Video modelling (VM), desired behaviours are acquired by watching a video demonstration of the desired behaviour by a correct model and then imitating the behaviour.
- Video self-modelling, to the ASD subject is asked to self-critique watching itself filmed during the target action execution after the experimenter has edited out all the undesirable behaviours.
- Point-of-view modelling, the camera is pointed to embrace the scene as the participant would see it, directed at a specific set of hands performing the desired task reducing irrelevant stimuli and optimising the focus on the specific task.
- Video prompting (VP), presenting the learner with a subjective viewpoint, it is not a fluid clip, like the VM procedure, but many mini clips based on a task analysis of the entire task, allowing the subject to perform each step in time with the video. VP is more effective in teaching daily living skills to adults with learning disabilities than VM, even if it is slower and more difficult to administer than VM.
- Computer-based video instructions present a variety of media, text, music, pictures and video footage. Computers use have positive effects on students,

due to higher attention rates, recreational associations of media files and increased successful performance.

In [128], VIRSA demonstrates that videos can be used to clarify developmental phenomena improving parent reporting of early development and allowing for web-based screening. The sensitivity of the VIRSA at 18 months is comparable to existing measures, suggesting that it is useful in identifying toddlers with ASD. Anyway, its specificity and positive predictive value, were lower than recommended standard and do not support the use of the VIRSA as a stand-alone ASD screener but only as an initial step in a screening process. Despite these limitations, the VIRSA demonstrates that it is possible to develop a parent-report instrument for identifying ASD risk in the first year of life. It also demonstrates that video can be used to clarify the developmental state. An innovation of the VIRSA is its web-based, mobile-optimized application deploying the smartphones diffusion even in lower income, rural, and minority communities. The VIRSA, with its low-burden, quick, online ratings, has the potential to reduce disparities in communities with limited access to screening and provide the possibility of initiating intervention before the symptoms set of ASD have emerged.

Many evaluations of the efficacy of VBI do exist. An individual's progress under VBI can be influenced by external factors such as preexisting social skills, the patient's age, mental ability and the duration and intensity of the program. In Bellini and Akullian (2007) [129] is reported that the obtained results suggest video modeling as an effective intervention strategy for addressing social-communication skills, functional skills, and behavioral functioning in children and adolescents with ASD. Results also indicate that these procedures promote rapid skill acquisition and these skills are maintained over time and transferred across persons and settings. Based on these results, video modeling intervention strategies meet criteria for designation as an evidence-based practice. A limitation of the video-based approach could be the perceived presence of the acquisition setup, causing inhibition of the natural behavior. Another one is that the methods requiring an offline evaluation are affected by the subjectivity, knowledge and experience of the operator.

In [130] the children interact with mobile, non-humanoid robots in whatever position they prefer (e.g., lying on the floor, crawling, standing). They are also free to

choose how to interact with the robot (touching, approaching, watching from a distance, picking it up, etc.). Interference by adults is only necessary when the child is about to damage the robot, or when the child switches off the robot. The robot is used to guide the children towards more 'complex' forms of interaction, as found in social human-human interactions. A purely reactive robot engages children with autism in simple, imitative, interaction games, based on elements of turn-taking. The robot's behaviour is guided by a small set of rules making it more predictable and less complex than human behaviour. The same 'approach child' behaviour is never repeated precisely but performed in variations to avoid perpetuating stereotypical and repetitive behaviour. The authors show that most children responded very well and with great interest to the autonomous robot. For a group of 18 children with autism, the statistical results showed a significant increase in the interaction levels of the children with the robot when considering the amount of eye gaze and attention directed at the robot. These results support the robots use in education and therapy of children with autism. However, given the nature of autism, one issue in this project, is the role of the affective aspects in child-robot interactions. There are ethical issues of encouraging the development of affective attachment of a child with autism with a robot that is not more than a machine without emotions.

2.2 Novel approaches

All the methods reported in Sec.2.1 use parameters derived from the clinical assessment of the presence and severity of ASD. Even if the use of digital technologies has moved the measure of such parameters toward a more objective and reliable level, the amount of data these methods provide has increased the needs of computational methods to speed up the evaluation process. Novel approaches to accelerate the diagnosis process, use learning algorithms for analyze and infer over one or more types of information acquired by the above methods [10]. The applications usually belong to two main categories which are the behavioural observation and the physiological signal analysis. Examples of applied learning algorithms are:

- In [131] a decision tree is fed with fractional anisotropy (FA), radial diffusivity

(RD), and cortical thickness from magnetic resonance data as features to perform classification between ASD and typical development children.

- Bone et al. (2016) trained and cross-validated an SVM classifier to differentiate ASD from other developmental disorders (DD) based on data from two standardized assessments, the Autism Diagnostic Interview, Revised (ADI-R) and the Social Responsiveness Scale (SRS) [132].

- Voice prosody was examined by Nakai et al. (2017), comparing the performance of an SVM classifier with cross-validation on 24 features vs the clinical judgment of speech therapists in classifying children with ASD and children with typical development (TD) based on single-word utterances. The SVM proved more accurate than the 10 speech therapists [132].

- Functional near-infrared spectroscopy (fNIRS) optical brain imaging modality, providing time-series of oxygenated hemoglobin (HbO_2) , deoxygenated hemoglobin (Hb), and total hemoglobin ($Hb_T = HbO_2CHb$), used in a multilayer convolutional neural network (CNN) for predicting ASD [133].

- Body-worn sensors (three axis accelerometer and a two-axis gyroscope) collected time series used for human activity recognition in a CNN to provide a higher level abstract representation [134, 135].

- Using curated lists of genes known to be associated with ASD and intellectual disability (ID), Kou et al. (2012) used two network-based classifiers and one attribute-based classifier, to classify known and predict new genes linked to these diagnoses. Finally, 10 SVM classifiers were employed using positive gene sets (i.e. genes associated with ID) and negative sets (i.e. genes associated with ASD), which were randomly generated with 200 genes in each set. The SVM performed better than both network-based classifiers [132].

- To accelerate the diagnostic process, Wall et al. (2012a, b) tried to identify a subset of ADI-R and of ADOS items that could be used to accurately classify ASD. An ADTree classifier was found to perform best with an accuracy of 99.9% in both cases.

- Support Vector Machine classification algorithm exploring patterns of eye movements during face recognition tasks used to differentiate children with ASD from children with typical development [9].

- Two CNN have been used to analyze speech dialog of autistic children in 3 stages [136]:

  - Thresholding for silence detection and Vocal Activity Detection for vocal isolation.

  - The 1^{st} neural network with frequency domain representations classifies utterance for the isolated vocals.

  - The 2^{nd} neural network recognizes autistic traits in speech patterns of the classified utterances.

- Neural networks have also been implemented by Linstead et al. (2015, 2017) in a task to find the relationship between treatment intensity and learning outcomes in the context of applied behavior analysis (ABA) treatment for ASD. Compared to simple linear regression, neural networks were more accurate. Even if the used neural networks had a single hidden layer and considered patients only within the early intervention age range, this work highlighted the capacity of the networks to learn non-linear relationships without any prior knowledge of the functional form of those relationships [132].

- Patnam et al. (2017) tried to identify behaviors that precede meltdowns and self-injurious behavior, in children with ASD. The authors used a recurrent CNN trained on video and images collected from various databases and Internet sources. On average, the model was trained in 30 to 60 min and tested on video recording of five individuals with 92% accuracy. The gestures were identified in less than 5 s and used to implement an alarming mechanism to alert caregivers in real time.

2.3 Limitations in the state of art

The main limitations in the current state of the art are:

- Subjectivity affecting the procedure. Be it a parent answering the questionnaire or the operator conducting the evaluation, the different knowledge of the ASD aspects and development milestones, even between two professionals, can lead to different diagnosis or to the missing of relevant aspects in the assessment. This subjectivity can be found also in the feature engineering process in case of handcrafted features and can cause poor discrimination and generalization.

- Invasivity of the devices used to record physiological signals. Even if it is a wearable device, the tools adopted to the scope introduce a disturbing element that isn't present in real life thus conditioning/biasing the observed behaviour or signal. The operator could be conditioning the natural response of the child. On the other hand, signal acquisition (MRI, functional MRI, electroencephalography) require the subject to be still for 5–10 min or longer. It is not an easy task for conscious (not sedated) children, especially for children with ASD.

- Complexity of the behaviour and heterogeneity in symptoms presentation of the different children. The autism is a spectrum of many disorders each of which can be present with a variable severity degree and manifested in a very great variety of manners. This aspect is worsen by the fact that each child has its own sets of repetitive behaviours, vocalizations, disruptive behaviours, stress and engagement response, what makes difficult the standardization and the hard coding of the evaluation process.

- The lack of specific datasets for autism makes the evaluation process affected by poor generalization power and less objective. This cause a delay in the stratification of subtypes within the ASD population, in the development of more targeted and effective therapies and drugs, and in evaluating their success remediating the core symptoms.

- Perceived engagement, as many other behaviours, displays cultural differences among children with ASD and there is still little research examining ASD char-

acteristics in cross-cultural settings. Cross-cultural analyses are important to provide better insights into the perception of symptoms and expression of different behaviors. [137, 138].

- Assessment is time consuming failing to provide quicker access to health care services.

- Diagnostic reliability and validity, without tools able to manage big amount of data to identify the most diagnostic features in ASD, are, in many cases, reduced.

2.4 How to go beyond the state of the art

A central problem in ASD is the assessment of the condition of the subject, be it for a first potential diagnosis or for evaluating the progresses of an already diagnosed subject. Still lacking ASD biomarkers, actual evaluation approaches are based on observation of the subject and in finding patterns in various types acquired signals related to the disease symptoms. The increase in data acquisition, has led to a broad application of computational algorithms, known as machine learning, in many fields, included ASD. Looking to the previous works in which the learning algorithm is focused on the ABA therapy, in one of the cases the therapy has been evaluated using supervised learning to assess the intensity of the therapy (i.e. the amount of therapy hours) with respect to the objective reached by the child. In brief a statistical approach which does not provide improvements to the therapy. Instead, the second reported case, points to understand and to signal an incoming crisis of the child from video recorded during the ABA therapy. Even if it is useful for preserving the child from stressful situations, also in this case there are no improvements for the therapy itself. Based on the limitations listed in Sec. 2.3 and on the examples in Sec. 2.2, it is possible to search for an approach allowing for a less invasive observation and a more reliable and fast evaluation of the ASD children. The core objective of the new approach should be to provide significant insights in the therapy, lightening the economic burden both for families and healthcare systems. Such an approach should be based on the video recording of the children during the sessions of ABA therapy aiming to classify their

progresses and to customize the therapy for each child based on their condition. This can be achieved by feeding a DL algorithm with the frames extracted from the acquired videos to obtain a classification of the autonomy level in performing actions. Such an approach is beneficial to the scope for several reasons:

- CNN architectures are widely used for image analysis tasks and easy to implement.
- The setup to acquire data is cheap and easy to build and use.
- Previous application of deep learning algorithm to behavioural observation reveal good accuracy in detecting patterns hidden in the images without relying on predetermined knowledge of the operator. The discovered hidden patterns will help so in find other relevant features to the diagnosis and assessment.
- Observation accomplished in this way allows for a more natural response of the child that is free from the constrains of the wearable sensors, and can be extended to the acquisition and analysis of other important markers (i.e. speech analysis, stereotypical motor movements detection, emotional state evaluation via face/eye-gaze recognition).
- The diffusion of smartphone applications joined with the high memory capabilities of the devices could move the detection of the ASD symptoms from health facilities towards the home environment also in lower income countries.

2.5 Aim of the work and thesis overview

The presented project, “Come a casa”, has been developed by the Università Politecnica delle Marche and ”Il Faro” Società Cooperativa Sociale in partnership with Azienda Ospedaliera Ospedali Riuniti Marche Nord, Clementoni S.p.A., SixS, Elicos S.R.L., Human Foundation. It consists in the development of a CNN algorithm from video monitoring of the behaviour of the guests of the “Il Faro”, at the Centro Orizzonte, during their daily life activities, gaming and simple movements, and evaluate their level of autonomy. The acquired material is used as input to the algorithm that provide real

time results useful to define the therapy and measure the improvements in relation to the performed therapy. In this work, two different algorithms have been implemented to compare their performances and choose the best in accomplishing the task. After a description of the relevant aspects of the disorders, a review of the literature on the current state of the art and its limitations is reported, followed by a description of the most used learning algorithms. Then are described the protocol used in data acquisition, the chosen implemented algorithms and the used performance metrics, to conclude with a report of the obtained results and a discussion of the future perspectives.

Chapter 3

METHODS

In this chapter, starting with Sec. 3.1 that introduce the world of artificial intelligence, will be done an overview of the more diffused learning algorithms, the machine learning family in Sec. 3.1.1 and the deep learning in Sec. 3.1.2. Then, after the description of the architectures used in this work in Sec. 3.2, the adopted data acquisition protocol will be explained in Sec. 3.3. The chosen training strategy, experimental settings and performance metrics are reported in Sec. 3.4. The chapter ends exposing the reliability test implemented using the Grad-Cam technique in Sec. 3.5 and the used programming language for the algorithms implementation in Sec. 3.6.

3.1 Overview on machine and deep learning

Artificial intelligence is a science which studies ways to build intelligent programs and machines that can solve problems as humans do. Machine learning is a subset of artificial intelligence providing different algorithms able to learn and improve by experience. A computer program is said to learn from experience E with respect to some class of tasks T and performance P, if its performance at tasks in T, as measured by P, improves with experience E [139]. Experience means from the data processed, the task describes how the data's examples have to be processed and the performance is the measured quantity (metric) used to evaluate the goodness of the achieved results (output) in executing the task. The learning is the mean by which

the ability to perform the task is gained. These algorithms take the crude data and, after a basic preprocessing step, return the natural patterns hidden in the data. This approach helps in classification decisions but also in make predictions on unseen data. Deep learning (DL) is a subset of machine learning that has gain increasing interest in the last years. Traditional machine learning algorithms work on a wide variety of important problems but cannot solve problems, such as recognizing speech or objects, because the mechanisms used to do this are insufficient to learn complicated functions in high-dimensional spaces. Such spaces, often, impose high computational costs. ML techniques, listed in Sec. 3.1.2, only transform the input data into one or two successive representations spaces, via simple transformations such as high-dimensional non-linear projections (SVMs) or decision trees. But complex problems require more refined representations, meaning the necessity to manually engineer good representations for the data, what is called features engineering. DL completely automates this step, learning all features in one pass rather than let the programmer to do it. This has greatly simplified the learning workflow, replacing sophisticated multistage pipelines with a single, simple, end-to-end DL model. Everything is controlled by a single feedback signal (the error). When "learning", a ML algorithm explores only a subset of all the infinite functions $f : R(d) \rightarrow T$, and chooses the best representing function among the subset it can select, its hypothesis space H. "Learning" means selecting "good" values for the parameters of the function so that these values are able to produce low values of the chosen error function, the loss function L. Parameters are then the values that control the behaviour of the system [140]. The two essential characteristics of how DL learns from data are:

- An incremental layer-by-layer way to develop more complex representations.
- The joint learning of the intermediate incremental representations, during which, each layer is updated to follow the representational needs of the layer above and the needs of the layer below.

3.1.1 Machine learning algorithms

There are two main categories of algorithms:

- UNSUPERVISED LEARNING (UL), used to find internal patterns based only on input data (clustering tasks without labeled samples).
- SUPERVISED LEARNING (SL), based on inputs and their associated labels, is used for classification into categories and regression for continuous variables such as temperature forecasting. Can be considered function-learning algorithms. In SL, the dataset presented to the algorithm is composed of input features and the corresponding label of the membership class. The term supervised means the view of the target (label) being provided by an instructor showing the system what to do. In unsupervised case the algorithm must learn to make sense of the data without this guide.

Examples of (UL) are:

- K-means
- Apriori algorithm
- Principal component analysis (PCA)
- Singular value decomposition (SVD)
- Indipendent component analysis (ICA)

While belong to the SL:

- Decision trees (DT)
- Naïve bayes classification (NBC)
- Support vector machines (SVM)
- Random forest (RF)
- Linear regression (LR)
- Ordinary least square regression (O-LQR)
- Logistic regression
- Ensemble methods (EM)

3.1.2 Deep Learning models

DL fundamental algorithms are known as neural networks and their structure is derived from the human neural system even if this link is only figurative. Their basic mathematical units, the neurons, are organized in successive layers. The number of these layers determines the depth of the model which can range from few to hundreds of layers (from which the term "deep") which increase the complexity of the learned features. The layers between the input (1^{st} layer) and the output (last layer) are called hidden layers. A neuron represents a linear transformation of the input followed by a chosen nonlinear activation function. X, the input, could be time series, images, sequence of data which are all represented by numbers and this extend the field of application of this type of algorithms to almost all the domains, from health informatics to energy, passing through economy, bioinformatics, psichology and mechanics.

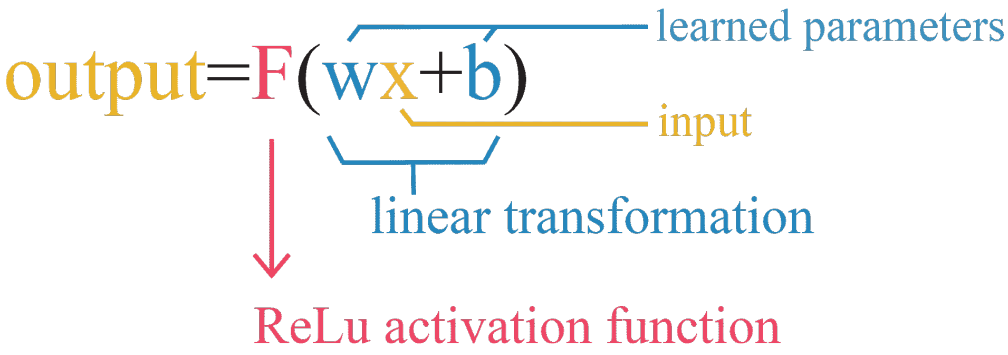

Figure 3.1: *Mathematical representation of a neuron's output*

Passing through the net, a sequence of data transformations, parameterized by the weights (w and b) of each neuron, are accomplished. The number of parameters can reach several millions and the modification of one parameter affects the behavior of the others. The summation output passes through an activation function F, that is a Heaviside step function, providing the network's output. Such learning systems are trained, meaning that they are fed with examples from which the most frequent patterns are extracted and used for the same classification task on new data. The loss function computes then the distance score between the output of the network and the true target (i.e. the training error) and uses it as the feedback signal to adjust the weights. The parameters update is starts using the backpropagation algorithm

(computation of the gradients of the loss w.r.t. the parameters w and b using the chain rule and storing till the call to optimizer). Once the update is done, in the optimization step, involving learning rate and momentum, the gradients are discharged and a new step of train (epoch) is performed in order to minimize the loss [141].

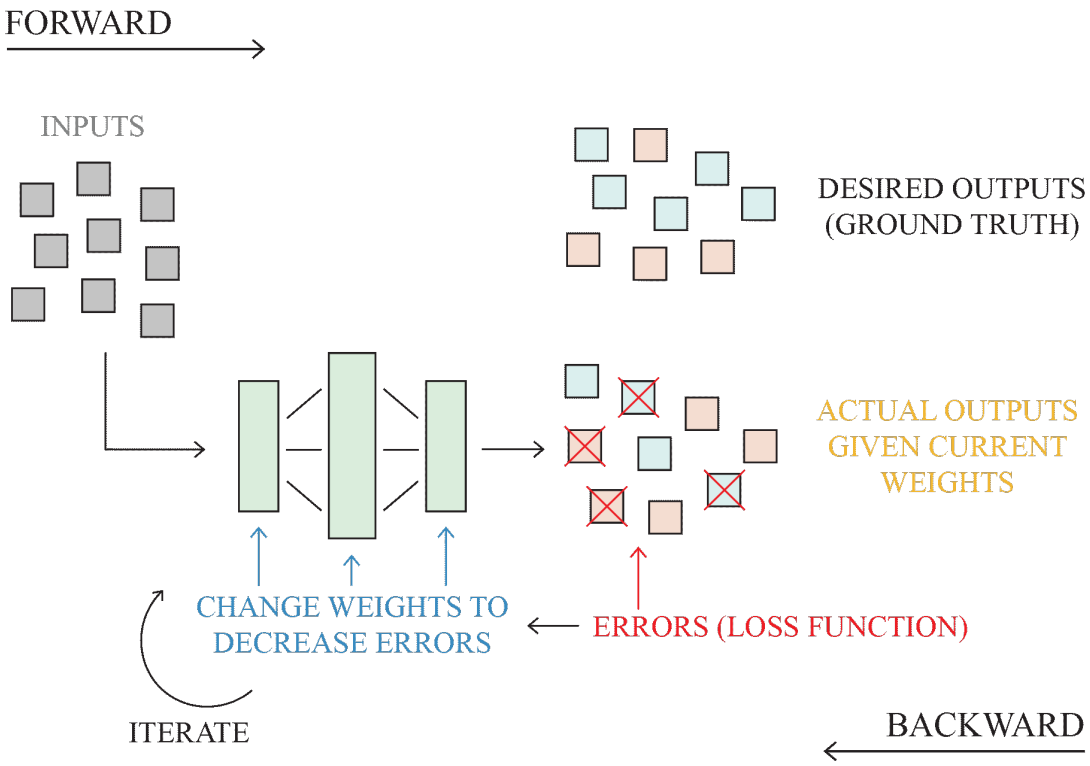

Figure 3.2: *Schematization of the learning process*

A good trained algorithm makes the training error small and the gap between the training and test error small. When the training error is too big, the algorithm is underfitting the training data. When the gap between the training error and the test error is too large, the algorithm is overfitting the training data and it is unable to generalize over new data samples. Loss trends and overfitting are shown in Fig [139, 142].

Another problem, related to the training of the model, is that of the gradient which can explode or vanish after some iterations. To make the model performing well in both training and test phases, many regularization techniques are used. By Dropout, for example, it is possible to prevent complex co-adaptations on the training data and so overfitting. In the presentation of each training case, each hidden unit is randomly

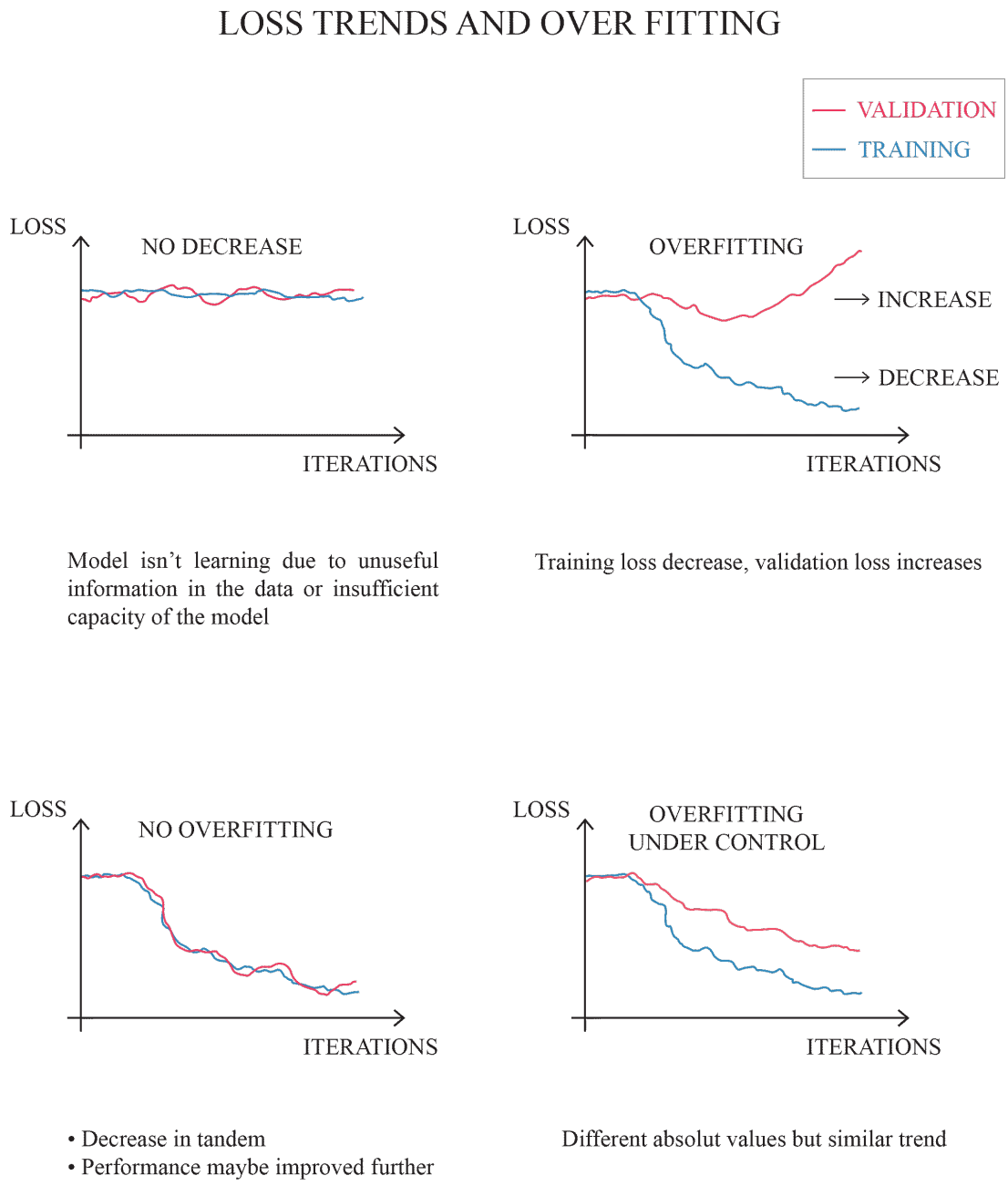

Figure 3.3: *Loss trends in training and validation phases and overfitting*

omitted from the network with a probability of p (usually 0.5). In this way a hidden unit cannot rely on other hidden units being present. During the test time, a scale of the output of each node by a value p is performed since each node is activated only p times [143]. Another regularization technique is the Batch Normalization (BN) and is based on the same concept on which relies the input normalization [144]. By normalizing the inputs, all the inputs features values are brought to the same scale.

This magnitude reduction avoids the updates, associated with the backpropagation, to be large and the learning algorithm to oscillate in the plateau region before it finds the global minima. The network can so train faster thanks to the managing of lower values. Similarly, the activation values for 'n' number of hidden layers present in the network, need to be computed. The activation values will act as an input to the next hidden layers present in the network and would vary a lot going deeper into the network, based on the weight associated with the corresponding neuron. To bring all the activation values to the same scale, the mean and standard deviation from a single batch are computed. BN is done individually at each hidden neuron in the network. In order to maintain the representative power of the hidden neural network, BN introduces two extra parameters, Gamma and Beta. Once normalized the activation, one more step to get the final activation value that acts as the input to another layer, is needed.

$$\mathrm{h}_{ij}^{norm} = \frac{\mathrm{h}_{ij} - \mu_j}{\sigma_j} \tag{3.1}$$

$$\mathrm{h}_{ij}^{final} = \gamma_j \cdot h_{ij}^{norm} + \beta_j$$

The parameters Gamma and Beta are learned along with other parameters of the network. If (γ) is equal to the mean (μ) and (β) is equal to the standard deviation (σ), then the activation h^{final} is equal to the h^{norm}, thus preserving the representative power of the network. With BN the loss of the network reduces much faster than the normal network because of the covariate shifting of the hidden values for each batch of input. This, along with the reduced magnitude of the scalars, helps in faster converge of the network, and reduces the training time.

The most popular DL methods are:

- Feedforward neural network (FNN)
- Convolutional neural network (CNN)
- Recurrent neural network (RNN)

- De-noising autoencoder (DAE)
- Deep belief networks (DBNs)

All the listed models are based on the concepts of neuron and layers but the have different topology and peculiarities. CNN's characteristics will be discussed in Sec. 3.1.2.1. Feedforward networks (a.k.a. multilayer perceptrons, MLP) is a class of networks consisting of multiple layers of neurons connected in a feed-forward way. The width of scope of applications of FNN comes from their ability to approximate complex functions and to modeling non-linear relationships. Each neuron in one layer has connections to all the neurons of the subsequent layer and use an activation function to determine the value feeding the following neuron in the connection. The learning techniques used by MLP are various but the most diffused is the back-propagation one, in which the outputs are used to compute the value of an error function, and the gradient of this error function respect to the weights is used for adjusting these weights and to decrease the error function. Back-propagation can only be applied on differentiable activation functions.

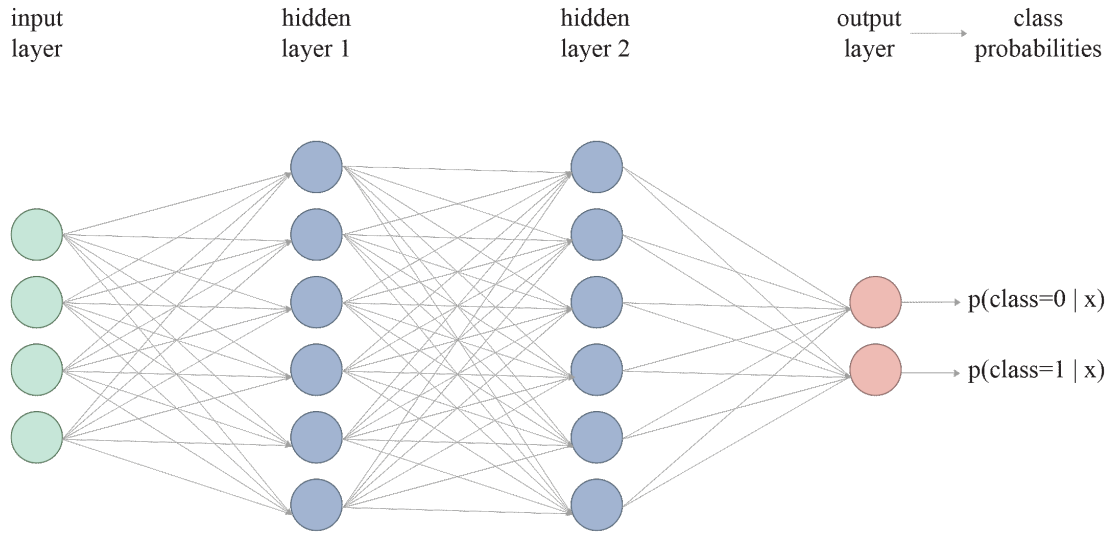

Figure 3.4: *Feedforward network example*

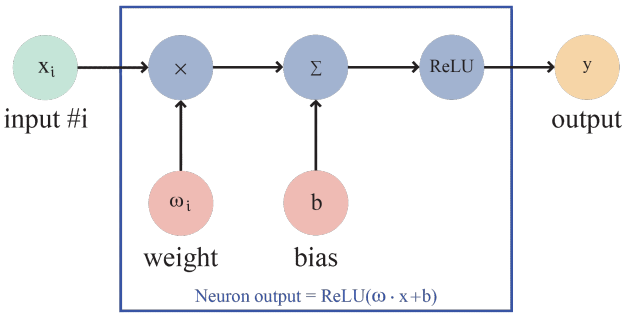

Figure 3.5: *Neuronal process in a perceptron*

Recurrent neural networks are a type of artificial neural network used in speech recognition and natural language processing (such as Apple's Siri and Google's voice search). These RNNs in fact are designed to recognize sequential characteristics and use the extracted patterns to predict the next likely scenario. RNNs differ from other types of artificial neural networks for the use of the feedback loops to process a sequence of data. The loop informs the final output, which can also be a sequence of data. These feedback loops allow information to persist and this characteristic is described as memory.

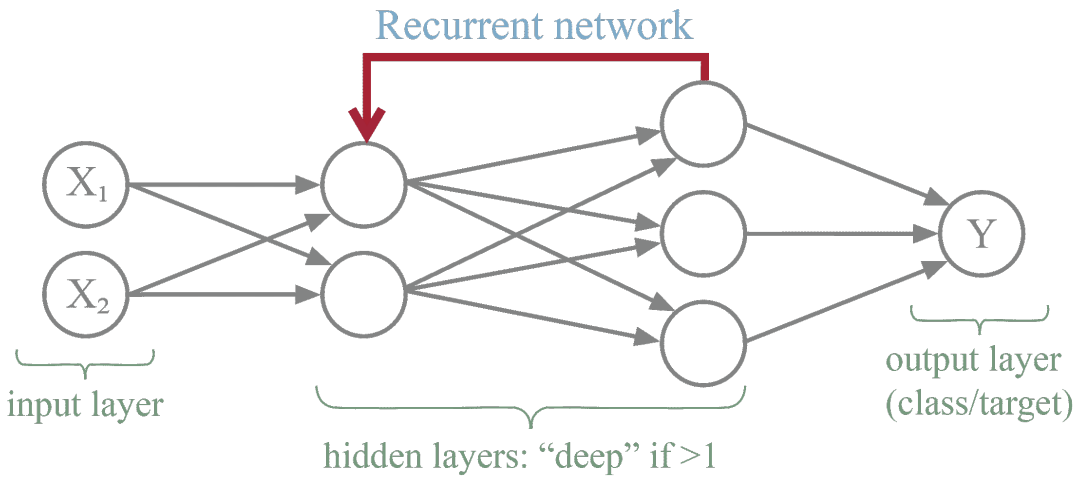

Figure 3.6: *Recurrent neural network*

Autoencoders are neural networks (for language translation tasks) used to learn a representation (encoding), to reduce the dimensionality (i.e. cut the noise), and to

generate an output, from this reduced code, closer to the original input. Between the encoding and decoding parts there is a hidden layer (code) which must have the number of nodes lower or equals to the number of input nodes to avoid an output perfectly equal to the (eventually corrupted) input. In DAE the input corruption by noise is made on purpose, setting a fixed percentage of input nodes to zero in a random way.

Deep belief networks are probabilistic models producing all possible values which can be generated for the case at hand. They are used as nonlinear feature learners, formed by a set of binary hidden units h. The variables in h are meaningful, not observable but inferred from other directly measured. Then there is a set of (binary or real-valued) visible units v, and a weight matrix W associated with the connections between the two layers.

3.1.2.1 Convolutional neural networks

Convolutional neural networks (CNNs) are a kind of neural network for processing data arranged in a grid-like topology such as time-series (form 1-D grid taking samples at regular time intervals), image data (are 2-D grid of pixels). The term "convolutional" indicates that the network employs the convolution mathematical operation (a kind of linear operation) in place of the general matrix multiplication. In the raditional neural networks layers every output unit interacts with every input unit, while CNNs have sparse interactions (also referred to as sparse weights) since they use a kernel (or filter) that is smaller than the input. For example, when processing a digital image, the input image might have thousands of pixels (i.e. pitcure elements), but we can detect small, meaningful features, such as edges, with kernels that occupy only tens or hundreds of pixels. This means storing fewer parameters, reduced memory requirements and computational costs, improved efficiency. Convolution consists of adding each pixel of the input image to its local neighbors, weighted by the elements of a kernel (i.e. a matrix of smaller dimensions with respect to the image), in order to create a new output image. Each pixel, in fact, store a scalar representing the intensity of the color (or grayscale for black and white pitcures). Scalars are coded with different numerical precision types (8-bit, 16-bit integers or floating) depending on the type of image, its resolution and quality. These are organized in a grid structure with height and width.

Coloured images are composed of 3 grids, referred to the intensity level of red, green and blue respectively, by which almost all the spectrum of visible colors is reproducible. Images can so be represented with tensor of dimensions, for example, of

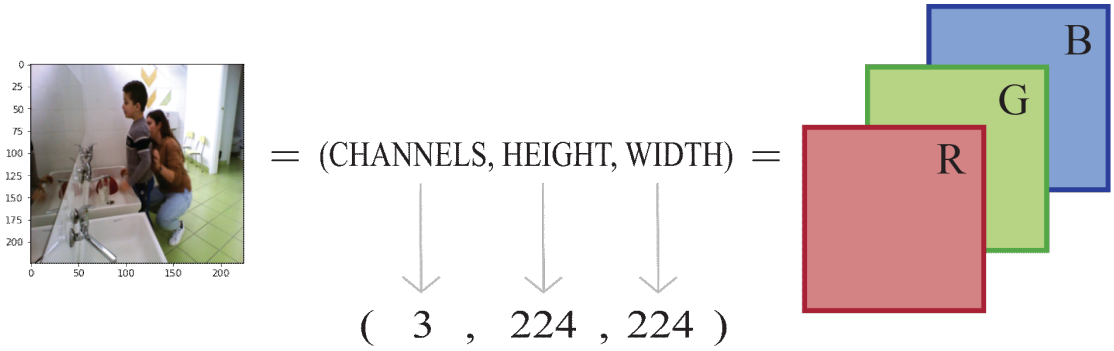

Figure 3.7: *Typical image channels composition*

In case of a depth image (RGB-D), the channels will be 4 (one more channel for the depth value). Software implementations usually work in batch mode to optimize the computational effort, so RGB images are transformed in 4-D tensors, with the first axis indexing a number of examples equal to the the batch dimension.

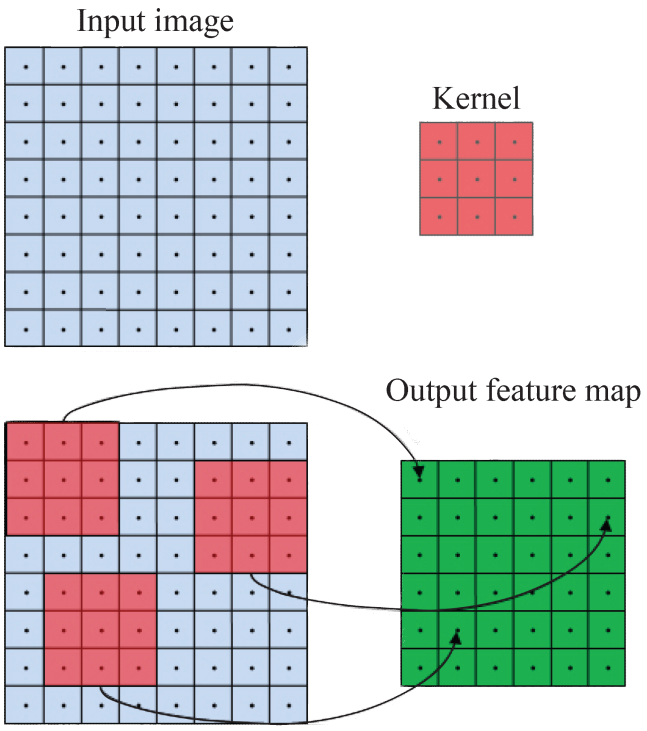

Figure 3.8: *Convolution and kernel*

$$\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} * \begin{bmatrix} a & b \\ d & e \end{bmatrix} = (1 \cdot a) + (2 \cdot b) + (4 \cdot d) + (5 \cdot e) \tag{3.3}$$

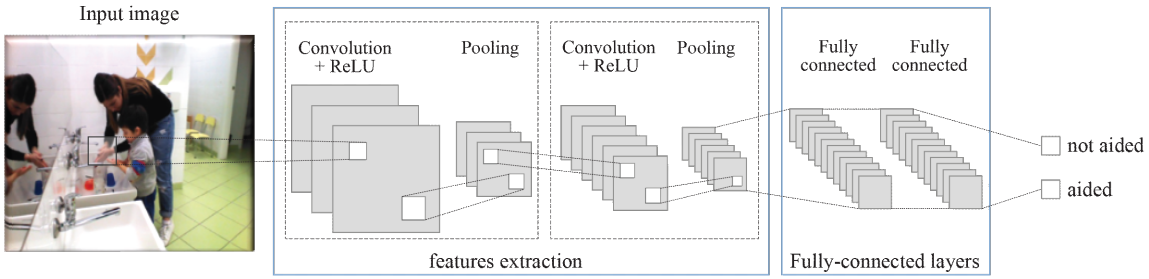

Figure 3.9: *Convolution, kernel, features extraction in a classification task*

If there are m inputs and n outputs, then matrix multiplication requires m x n parameters (per example). Limiting the number of connections each output may have to k, then the sparsely connections require only k $\times$ n parameters. In a deep convolutional network, units in the deeper layers indirectly interact with a larger portion of the input (field of view). This allows the network to efficiently describe complicated interactions, between many variables, by constructing such interactions from simple building blocks, each describing only sparse interactions. Parameter sharing refers to using the same parameter for more than one function in a model in such a way that the network performs convolution operation on the image. In a traditional neural net, each element of the weight matrix is used exactly once when computing the output of a layer. It is multiplied by one element of the input and then never revisited. It is the same to say that a network has tied weights, because the value of the weight applied to one input is tied to the value of a weight applied elsewhere. In CNNs, each member of the kernel is used at every position of the input (except perhaps some of the boundary pixels, depending on the design decisions regarding the boundary). In this way, the parameter sharing allows to learn only one set of parameters rather than learning a separate set of parameters for every location. This does not affect the runtime of the forward pass, but it does further reduce the storage requirements of the model to k parameters. Convolution is thus dramatically more efficient than dense matrix multiplication in terms of memory requirements (not every "layer" has parameters) and statistical efficiency. The parameter sharing causes the layer to have a property called

equivariance to translation, meaning that if the input changes, the output changes in the same way. Convolution creates a 2-D map of where certain features appear in the input. If we move the object in the input, its representation will move the same amount in the output. This is useful when we know that some function of a small number of neighboring pixels can be applied to multiple input locations (in edges detection the same edges appear practically everywhere in the image). Convolution is not naturally equivariant to some other transformations, such as changes in the scale or rotation of an image. Other mechanisms are necessary for handling these kinds of transformations [140]. In a CNN, the operations performed can be said to be a three stages procedure. In the first stage, the convolutional layer performs several convolutions with a set of K kernels in parallel each generating a new feature map X_K. In the second stage, each feature map is given as input to an element-wise nonlinear activation function f, such as the rectified linear activation function. In the third stage, a pooling function modifies the output of the layer replacing it with a summary statistic of the nearby outputs. For example, the max pooling reports the maximum output within a rectangular neighborhood.

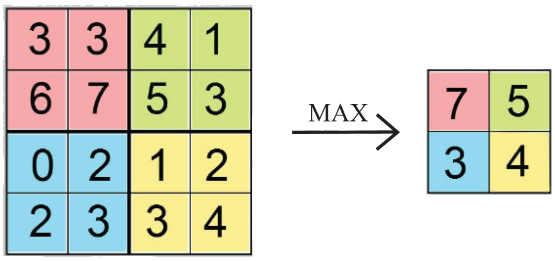

Figure 3.10: *Max pooling*

Other popular pooling functions include the average of a rectangular neighborhood, the $L2_{norm}$ of a rectangular neighborhood, or a weighted average based on the distance from the central pixel. In all cases, pooling helps to make the representation invariant to small translations of the input meaning that translating the input by a small amount, the values of most of the pooled outputs do not change. This can be a useful property if a concern is whether some feature is present than exactly where it is. In other contexts, it is more important to preserve the location of a feature. For example, if we want to find a corner defined by two edges meeting at a specific orientation, we need to preserve

the location of the edges well enough to test whether they meet. Pooling is also useful to reduce memory requirements for storing the parameters. The dot product between the kernel and the overlapping region of the input is computed at each new location of the kernel. At each layer, the output volume size is controlled by two parameters: the depth of the layer and the stride. The depth is the number of kernels used, each learning to look for something different in the input (edges, colors, etc.). The stride is a step size with which the kernel is slipped over the image, namely the number of pixels skipped at each convolution. It controls the output volume in its width and height. Finally, after several convolutional and pooling layers, the CNN ends with one or more fully connected layers, that produce non-spatial output. Fully connected layers are one-dimensional layers with full connections to all activations in the previous layer.

3.2 Proposed architectures

To perform the frames classification task, two different network types have been chosen: VGG16 and ResNet50. Both architectures were tested in the from scratch and pretrained version, for a total of 4 models, to find the one which performs better. Using the pretrained versions, the fine-tuning methodology has been adopted to migrate the knowledge, learned by the two models, during the training on Imagenet. In sections 3.2.1 and 3.2.2, will be explained the details of each model and how them have been adapted to the goal of this work.

3.2.1 VGG-16 Neural Network

The first chosen model is a convolutional network called VGG16. The number in the name means that the layers with weights to be updated in training are 16. The dimensional input requirement is 224x224 pixels. Its layout comprises specifically:

- 13 convolutional layers (the feature extraction part of the net)
- 3 fully connected layers
- 5 max pooling layers

The size of the filters receptive field of the convolutional blocks is 3x3 pixels and is activated by a rectified linear unit (ReLU) activation function. Every two or three convolutional blocks (depending on the network depth), max pooling layers are used to progressively reduce the spatial size of the feature map. Specifically, the Max-Pooling 2D is placed after the 2^{nd}, 4^{th}, 7^{th}, 10^{th}, 13^{th} convolutional layers. After the last max pool operation, 3 fully connected layers in series end the net. The 3 fully connected layers with 4096, 4096, and 1000 neurons, respectively, are separated by dropouts to reduce the effects of overtraining of the neural network (training set tracking). The last fully connected layer is followed by a softmax layer, which returns the probability of the image to belong to each class of the Imagenet dataset, the dataset of images used to train originally the pretrained version of the VGG16. To accomplish the binary classification task, the last fully connected layer has been replaced with a fully connected layer with 2 neurons 3.11.

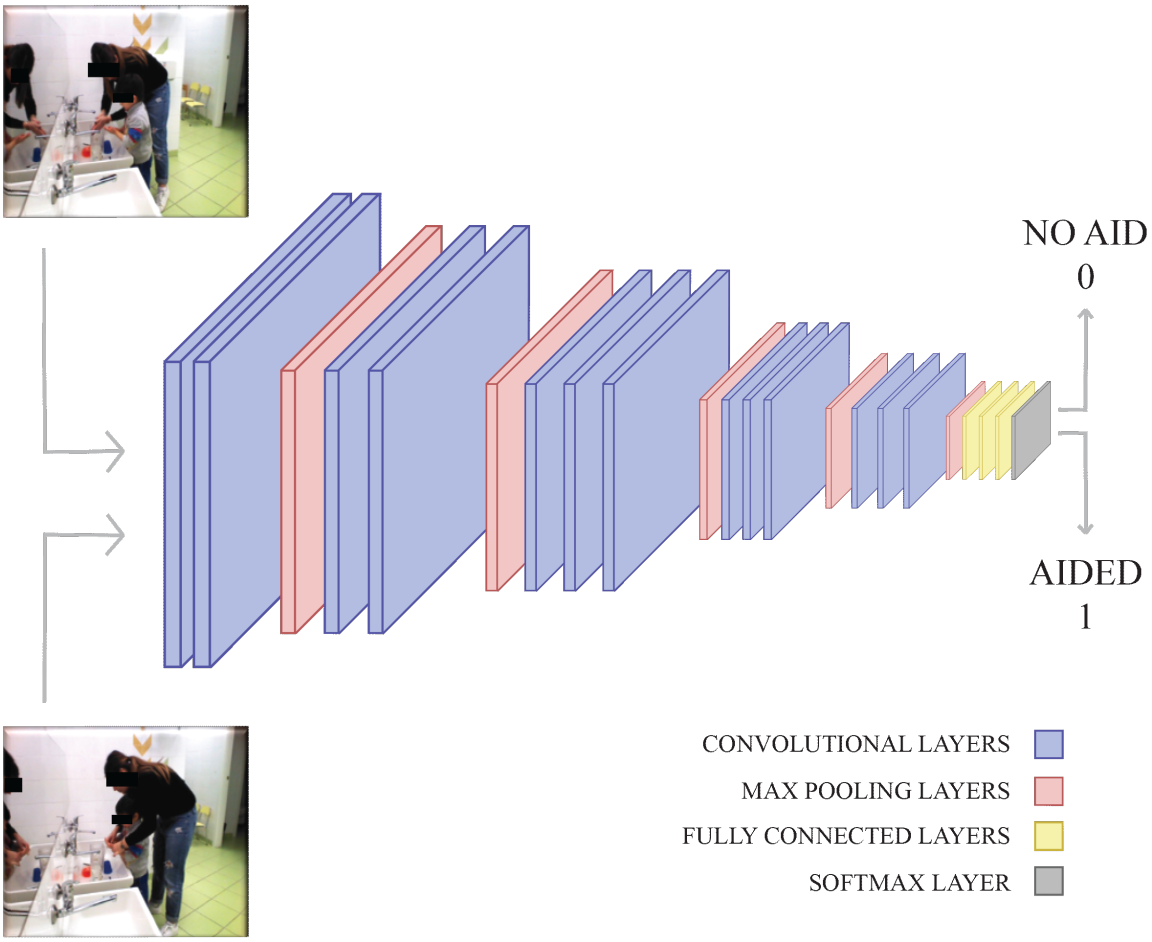

Figure 3.11: *Vgg16 layers composition*

3.2.2 ResNet-50 Neural Network

Using deeper networks, the level of features can be enriched by the number of stacked layers (depth). A not negligible hinderance to their applicatioin is that, during the application of Back-Propagation and chain rule methods (to find the gradients and send back to hidden layers for weights update), the gradient could assume low values or even become null. Consequently, this nullity of the gradients causes unchanging weights and no effective learning. ResNet50 is another type of convolutional network which solves the problem. It makes use of residual connections to add the value of the input x at the beginning of the block to the end of the block (F(x)+x). This connection doesn't pass through the convolutional layers, so the derivatives are not reduced but result in a higher overall derivative of the block. With the help of residual blocks, it can be increased the number of hidden layers as much as is wanted. The residual function creates a duplicate of the given input to preserve the previous output from the possible disastrous transformations without introducing neither extra parameter nor computation complexity [145]. After storing the original value of x, it undergoes a series of convolution operations as it tries to maximize the learning. Since the original weights of x are preserved (shortcut), they are finally added (element-wise channel by channel) with the transformed x to eliminate the possible negative effects of transformations. The weights act as an identity function. Gradients propagating through the identity connection path does not encounter any weight so doesn't change. Otherwise, the newly learnt weights are summed to what the network has learnt. The number of hidden layers is not a concern now for the neural network architecture, since using a ResNet, we will take care of the maximum learning of weights possible without over-fitting or negative deviation of accuracy through means of vanishing or exploding gradients problem. ResNet50 has the same types of building layers (convolution, pooling, activation and fully connected layers) with a different layout and in higher number:

- Initial Convolutional layer (kernel 7x7), Batchnorm, ReLU and Maxpool (kernel 3x3).
- Then starts a series of 4 blocks (or stages) each one hosting a different number of residual blocks stored in the layers vector (dimension 4, a value for each block)

A residual block contains:

- 3 convolutional layers
- Skip connection that can be:
  - Identity Shortcut (input/output dimensions are equal)
  - Projection Shortcut (a convolution operation to match the volumes sizes to sum - dotted line in the diagram)

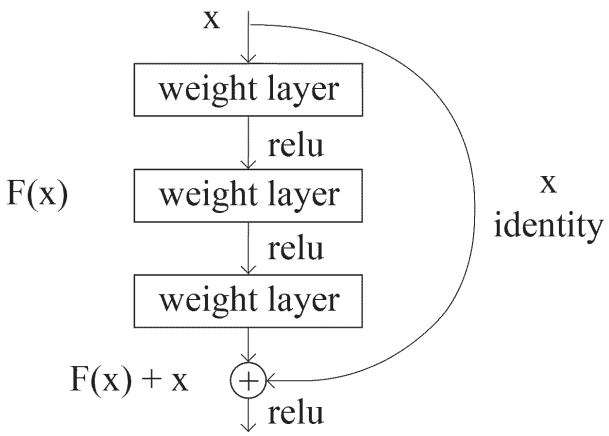

Figure 3.12: *Shortcut connection scheme*

To reduce the training time, the building block has been modified as a bottleneck design. For each residual function, it has been used a stack of 3 layers, 1x1, 3x3, and 1x1 convolutions, where the 1x1 layers are responsible for reducing and then increasing (restoring) dimensions, leaving the 3x3 layer a bottleneck with smaller input/output dimensions. The parameter-free identity shortcuts are particularly important for the bottleneck architectures. Using instead the projection type, the time complexity and model size are doubled, as the shortcut is connected to the two high-dimensional ends. The most frequent skip connections are identity, which make more efficient the model for the bottleneck design. From 1 stage to another, the channel width is doubled and the input size is halved. Finally, there is an Average Pooling layer and a fully connected layer having 2048 neurons to flatten the dimensionality to 1x1xN (neurons). The last layer has been changed to 1x1x2 dimension to deal with the task requirements. As in

the VGG16 model, the last fully connected layer is followed by a softmax layer, which returns the probability of the image to belong to each class.

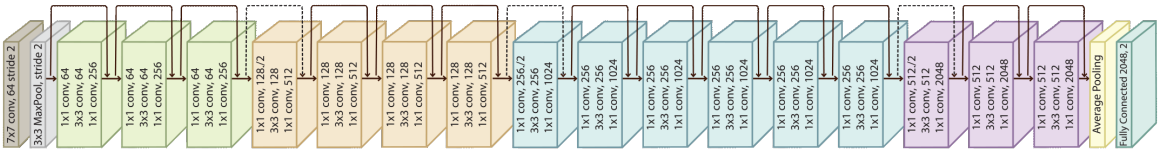

Figure 3.13: *ResNet50 layers composition*

3.3 Data Acquisition Protocol

The aim of this study is to develop a deep learning algorithm able to detect, directly from video frames, if the observed subject performs the target action by itself or aided by the operator. The results of the classification will allow the operators to evaluate the level of autonomy and the achieved progresses during the ABA therapy. Written informed consent has been obtained from the parents of all participants. The experimental protocol is focused on the hands-washing action observation and videos were acquired by an RGB-D camera (Astra Mini S-Orbbec®) and a minipc Intel® NUC core i5, installed to be imperceptible and to not distract the child during the therapy.

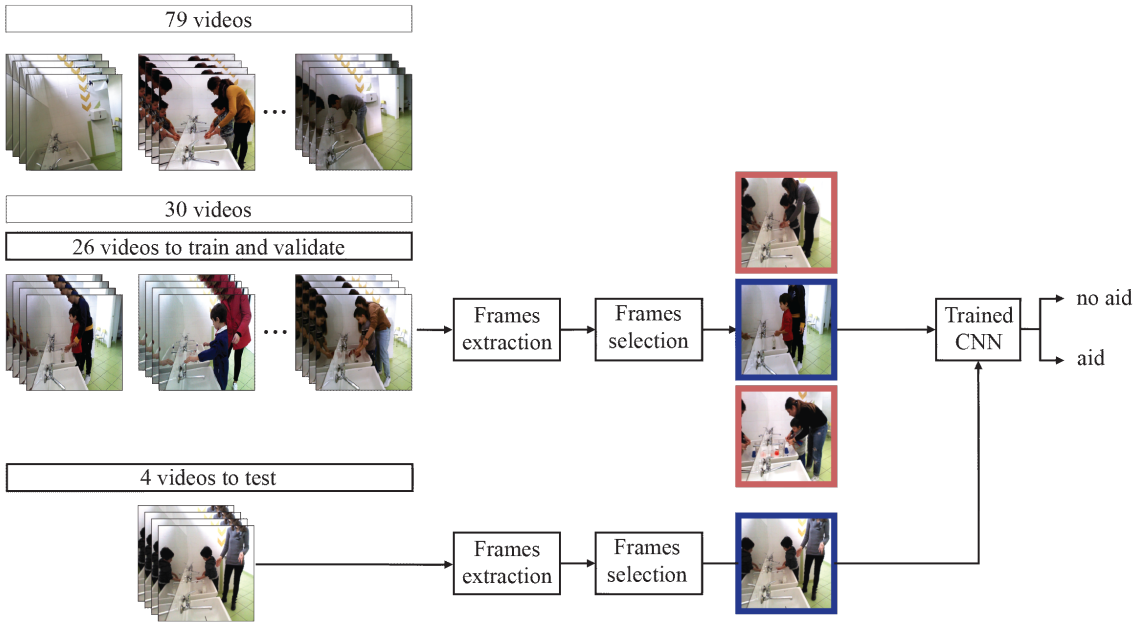

Figure 3.14: *Workflow of the images classification task*

The acquired videos have a duration of 5 minutes and a frame rate of 25 fps. Each RGB frame has a resolution of 640x480 pixels. From the collection of frames of each video, were extracted 1 frame every 6 frames and, among the extracted, only those frames in which the subject is in the field of view were selected for the successive annotation phase. The annotation has been done following these rules:

- Subject and operator in front of the sink
- Frames where the subject act by itself are labeled as 0 (NO AID)
- Frames in which the operator helps in performing the action are labeled as 1 (AIDED)
- Physical contact of the operator hands with the arm of the subject is labeled as 0
- Operator standing near or behind the subject without touch is labeled as 0
- Operator mimicking the action is labeled as 0

From a total of 1247 acquired videos, only 46 shows the subject in the field of view and, among these, only in 30 videos there is a clear execution of the observed action. A total of 142200 frames has been selected and, after the annotation phase, only 9712 frames have been validated to form the ground truth dataset, 4607 labeled as 0 (NOT AIDED) and 5105 as 1 (AIDED).

3.4 Training strategy and experimental settings

All the frames formìng the dataset have been resized to a 224x224 resolution to meet the input requirements of the networks and then transformed to tensors of shape (3, 224, 224) before to load them into 4-D batch tensors of shape (64, 3, 224, 224). To ensure a better generalization capability to the developed algorithms, the held-out cross validation has been accomplished dividing the dataset in training, validation and test subsets paying attention that the frames coming from one video are present in only one of the three subsets. This ensure that unseen frames are available for validation and

testing. The dimensions of the subsets reflect a 65/20/15 percent proportions respectively. These are approximated percentages since the author has tried to maintain a balance between each class in all the sets to avoid class imbalance affecting the metrics. Before to insert into each dataset, images have been normalized channel-wise, subtracting and dividing by appropriate values of mean and standard deviation respectively, as suggested from the best computer vision practices.

Table 3.1: *Dataset splitting proportions reported in number of: children, videos and frames*

Training set		**Validation set**		**Test set**	
No-aid	Aid	No-aid	Aid	No-aid	Aid
3255	3433	805	960	547	712
26 videos				4 videos	
6 children				4 children	

Both the architectures types have been trained from scratch and using a pretrained version on ImageNet dataset (1.4 million images) [12] to transfer the knowledge previously acquired on such dataset. When trained from scratch, the weights are initialized using kaiming normal method (He initialization) in "fan out" mode to preserve the magnitude of the weights variance in the backward pass and to avoid the risk of the exploding gradient problem and non effective learning that is real using a totally random initialization. In both cases of pretrained versions, has been made the load of the Imagenet weights for the conv blocks and for the connections between neurons until the last fully connected layer, while this last layer was initialized with the standard Glorot initialization. The fine tuning has been done retraining the entire net (with "freezed" layers till the fully connected layers) on the current dataset. In both Vgg16 and ResNet50 architectures, in both the pretrained and from scratch modalities:

- Many trainings have been performed to optimize the hyperparameters (learning rate, learning rate decay, number of epochs, batch size) to reach the better configuration of the models.
- Both "stochastic gradient descent" and "Adam" optimizers have been tested to choose the better for each architecture. Finally, the SGD was chosen for the

VGG16 and the ADAM for the ResNet50.

- For the SGD case, the initial learning rate is decayed by a factor of 2 every 7 epochs
- For all the architectures, the batch size was set to 64, as a trade-off between memory requirements and training convergence, while the number of epochs was set to 30 for the VGG16 models and to 20 for the ResNet50. The decision is due to the uneffective learning presented by the models after the chosen limit of epochs.
- The best weights configuration among epochs, for each model, was retrieved according to the highest accuracy on the validation set.

Table 3.2 reports the final choice of hyperparameters for each model.

Table 3.2: *Hyperparameters choice of each model*

PARAMETER	VGG SCRATCH	VGG PRETR	RES-50 SCRATCH	RES-50 PRETR
Batch normalization	YES	YES	YES	YES
Batch size	64	64	64	64
Number of epochs	30	30	20	20
Initial learning rate	0.0001	0.0001	0.0001	0.0001
Optimizer	SGD	SGD	ADAM	ADAM

3.4.1 Performance metrics

To evaluate the performances of the models the following metrics have been computed:

- Training and validation losses using "Crossentropy loss" method:
  - It measures how well the net behaves after each iteration of optimization.
  - It is the distance between the true values of the problem and the values predicted by the model and should reduce after each, or several, iterations.

  - It is, practically, the sum of the errors made for each example in training or validation sets.

- Training, validation and test accuracies. Each phase accuracy:
  - Is determined after the model parameters are learned and fixed. In training, this computation is done using the weights of the current epoch.
  - Measures the algorithm's performance in an interpretable way.
  - Is calculated in the form of a percentage (correct predictions over the total of samples) and there is no relationship between loss and accuracy metrics.

- Confusion matrix in test (reported in normalized version):
  - Rows represent the instances in a true class, columns represent the instances in a predicted class (or vice versa). It makes easy to see if the system is mislabeling the two classes (one as another).
  - Classification accuracy is the ratio of the correct predictions to the total of predictions made, presented as a percentage by multiplying the result by 100.

- Classification report in which are reported the following metrics, the scores of which correspond to every class and tell the accuracy in classifying in that particular class compared to all other classes. These metrics can be:
  - MACRO AVERAGED, that is metrics are calculated for each label finding their unweighted mean. This doesn't take label imbalance into account.

$$\frac{(\text{metric}_0 + \text{metric}_1)}{2} \tag{3.4}$$

  - WEIGHTED AVERAGED, meaning metrics are computed for each label, finding their average weighted by the support (the number of true instances for each label). This alters the 'macro averaged' one to account for label imbalance; it can result in an F-score that is not between precision and recall.

$$\frac{(\text{support}_0 * \text{metric}_0 + \text{support}_1 * \text{metric}_1)}{\text{support}_0 + \text{support}_1} \tag{3.5}$$

– Precision (correctly positive predicted observations over the total positive predicted observations, high precision relates to the low false positive rate).

$$\text{Prec}_{class-j} = \frac{\text{TP}_j}{\text{TP}_j + \text{FP}_j} \tag{3.6}$$

– Recall (or sensitivity, is the ratio of correctly positive predicted observations to the all observations in actual class, good is above 0.5).

$$\text{Rec}_{class-j} = \frac{\text{TP}_j}{\text{TP}_j + \text{FN}_j} \tag{3.7}$$

– F1-score (weighted average of Precision and Recall, reaches its best value at 1 and the worst score at 0. It takes both false positives and false negatives into account and is more useful than accuracy in case of unbalanced class distribution).

$$\text{F1}_{class-j} = 2 * \frac{\text{Prec}_{class-j} * \text{Rec}_{class-j}}{\text{Prec}_{class-j} + \text{Rec}_{class-j}} \tag{3.8}$$

– ROC curve and AUC are performance measurements for classification problems at various thresholds settings. ROC is a probability curve, AUC represents the degree or measure of separability. Both tell how much the model is capable to distinguish between classes. Closer to 1 the AUC, the better the model is at predicting 0s as 0s and 1s as 1s. The 0.5 value (represented by the diagonal) means that the model hasn't class separation capacity. The ROC curve is plotted with TPR (Recall) against the FPR (1-Specificity) where TPR is on y-axis and FPR is on the x-axis.

$$\text{Spec}_{class-j} = \frac{\text{TN}_j}{\text{TN}_j + \text{FP}_j} \tag{3.9}$$

$$\text{FPR} = 1 - \text{Spec}_{class-j} = \frac{\text{FP}_j}{\text{TN}_j + \text{FP}_j} \tag{3.10}$$

Sensitivity and specificity are inversely proportional to each other. Increased sensitivity means decreased specificity and vice-versa. Decreasing the threshold, leads to get more positive values thus increasing the sensitivity and decreasing the specificity. Similarly, increasing the threshold, leads to more negative values thus to higher specificity and lower sensitivity. Since FPR is 1 – Specificity, increase in TPR means also FPR increase and vice versa in case of decrease.

3.5 Reliability test and visual explanation: Grad-Cam

Neural networks can be considered as black boxes regarding to how they assign a class to a processed sample. Grad-Cam is a method allowing to highlight the regions of the image where the net is focusing when it makes the class choice provided as output. It is a way to test the reliability of the predictions. Applying the method, it is generated a heatmap, of the same size of the original image, presenting increasing color intensities in correspondence of the pixels capturing the net attention. Superimposing the heatmap to the input image, an observer can argue which are the features of the image having more importance for the net and which are the patterns the net has learnt from the dataset. The method is applyied after the training and in evaluation modality (fixed parameters). The basic concept behind the Grad-Cam is to exploit the spatial information preserved through the convolutional layers using the feature maps produced by the last convolutional layer. It is in the last convolutional layer, in fact, that the authors of [146], expect to have the best compromise between high-level semantics and detailed spatial information. The part of the network relevant to Grad-Cam is represented by the feature maps (A1, A2, A3) of a chosen layer as depicted in Fig. 3.15.

The only requirement is that the layers, inserted after the A1, A2, and A3 feature maps, have to be differentiable (i.e. gradient computation allowed) so to get a gradient to use for computing the so called “alpha values” by which the feature maps are weighted. This makes Grad-CAM applicable to any architecture because, gradients

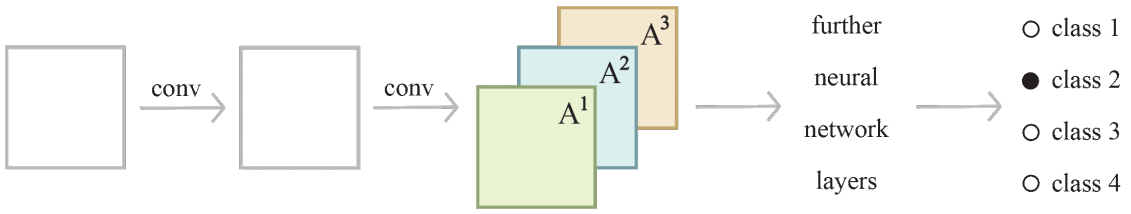

Figure 3.15: *Feature maps exploited by the Grad-Cam technique to compute the heatmaps*

can be computed through any kind of neural network layer. The "Grad" in Grad-CAM stands for "gradient". The output of the Grad-CAM is a "class-discriminative localization map", a heatmap, in which the red part corresponds to a particular class.

$$L^{c}_{GradCAM} \in \mathbb{R}^{u\ x\ v} \qquad \text{Localization map width } u\text{, height } v\text{, class } c \tag{3.11}$$

If the classification involves 2 classes, then, for an input image, 2 different Grad-CAM heatmaps are computed, one heatmap for each class. The steps of Grad-Cam are:

- 1. Compute the gradient of y^c (the output of the neural network for class c) with respect to the feature map activations A_k of a convolutional layerdy_c/dA_k (partial derivative). This gradient depends on the particular image because this image determines the y_k feature maps as well as the final class score y^k that is produced. For a 2D input image, this gradient is 3D, with the same shape as the feature maps. There are k feature maps, each of height v and width u, i.e. collectively the feature maps have shape [k, v, u]. This means that the gradients calculated in 1 are also going to be of shape [k, v, u]. In the pitcure below, k=3 so there are three u x v feature maps and three u x v gradients.

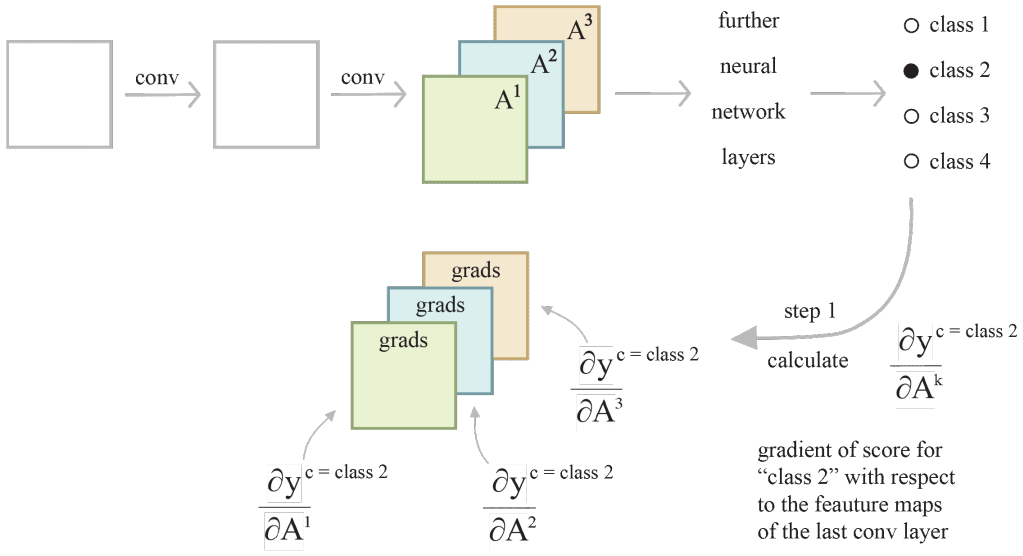

Figure 3.16: *Grad-Cam: computation of the gradients of the scores of a class with respect to the feature maps of the last convolutional layer. Considered class is the class 2.*

- 2. Globally averaging the gradients over the width and height dimensions (i, j) to obtain neuron importance weights, the alpha values α_k^c.

$$\alpha_k^c = \overbrace{\frac{1}{Z}\sum_i\sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A_{ij}^k}}_{\text{gradients via backprop}} \tag{3.12}$$

The alpha value for the class c is going to be used in the next step as a weight applied to the feature map k A_k. Pooling over the height v and the width u ends up with shape [k, 1, 1] or just [k]. These are the aimed k alpha values.

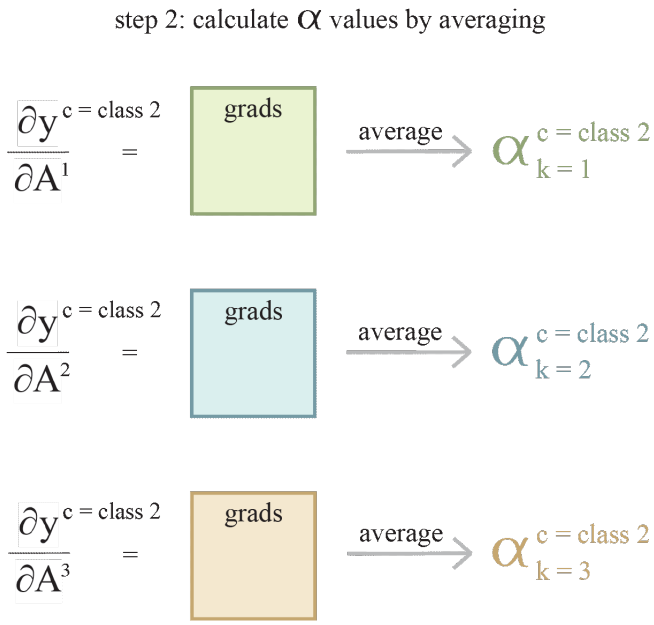

Figure 3.17: *Grad-Cam: alpha are values computed via global averaging pooling of the gradients over the width and height dimensions which will be used to compute a weighted sum of all the feature maps.*

- 3. Compute the final heatmap performing a weighted sum of each feature map activations A_k multiplied with the corresponding α_k^c values and apply a ReLU operation to turn into 0 all the negative values and emphatising the positive.

step 3: calculate the final heatmap as a weighted combination of the feature maps

$$\text{Grad-CAM}^{\text{class 2}} = \alpha_1 A^1 + \alpha_2 A^2 + \alpha_3 A^3, \text{ i.e. } \sum_k \alpha_k^{\text{class 2}} A^k$$

In Grad-CAM we also apply a ReLU:

$$\text{Grad-CAM}^{\text{class 2}} = \text{Re LU}\,(\alpha_1 A^1 + \alpha_2 A^2 + \alpha_3 A^3)$$

Figure 3.18: *Grad-Cam: weighted sum of the feature maps using the alpha values and successive application of the ReLU function to obtain the final heatmap for that class.*

- 4. The heatmap is a lot smaller than the original input image size, so up-sample (using interpolation) is needed before the final visualization.

3.6 Programming language and Colab environment

The two architectures have been implemented on the Google Colaboratory cloud platform notebooks (Colab notebooks) using Python object-oriented programming language. The main reasons for this choice are:

- The Python language offers the PyTorch deep-learning library which is built to work with tensors by which multidimensional entities (inputs and weights) are easily representable.
- Even if the operations performed within a neural network are simple, they occur in huge number and require a lot of time to be accomplished. This time can be shortened exploiting the higher number of operations per second offered by the GPUs and TPUs available on Colab.
- Using Pytorch allows to write the code only one time since it can run on CPUs and GPUs without changing it.
- Pytorch is built around the concept of neural network and offers very powerful functionalities and alternatives to make transparent and easy all the steps of training, evaluation and metrics computation.
- A Colab notebook shows itself as a page in the browser through which the code can be interactively run and executed by a server that send back the results. It maintains the variables defined during the pre-compiling and the execution of the code, in memory until the runtime it's terminated or restarted. Python is an interpreted language so it uses an interpreter (which is said to pre-compile the code) rather than a compiler before the execution.
- The code can be divided in multiple cells that one can run in different times, allowing the new cells to see the variables created in the already executed cells. This greatly reduces the code debug times.

Chapter 4

RESULTS

The best configuration of hyperparameters for each used network is reported in table 3.2 which summarizes the results achieved by VGG16 and ResNet50, both fine-tuned and trained from scratch. Then, for each model, it will be reported: confusion matrix, classification metrics and the ROC (Receiver Operating Characteristic). At the end of this section, for the best performing model, images of the heatmaps computed via Grad-Cam are inserted. The loss recorded in the training phase is reported together with the loss in the validation phase for the 4 models in Fig. 4.1

The same thing is done for the accuracies in training and validation, reported in Fig. 4.2

Table 4.1: *Metrics*

MODEL	PREC		REC		F1		ACC
	No-aid	Aid	No-aid	Aid	No-aid	Aid	
VGG16 scratch	0,68	0,79	0,74	0,74	0,71	0,76	0,74
fine-tuned VGG16	0,78	0,79	0,71	0,84	0,74	0,82	0,74
ResNet50 scratch	0,74	0,82	0,78	0,79	0,76	0,81	0,79
fine-tuned ResNet50	0,95	0,78	0,65	0,97	0,77	0,87	0,83

The best accuracy is showed by the pretrained ResNet50. The two networks trained from scratch achieved the lowest performance w.r.t. the homologous fine-tuned ones.

Figure 4.1: *Loss in training and validation*

The from-scratch versions have unbalanced values of the perclass metrics, exception made for the recall of the VGG16 from scratch that is equal for both classes. The same model achieved the worst accuracy. The results highlighted that both the architectures trained from-scratch are more confident in predicting the aid class with respect to the no-aid one. The recall is the ability of a model to maximize the true positives. The only model showing a recall below the 70% is the fine-tuned ResNet50 (65% for the no-aid class) which, however, shows a 97% recall for the aid class. The precision is the ability of a model to minimize the false positives. The only model having a precision below the 70% is the from scratch version of the VGG16 (68% for the no-aid class). All the other precision percentages are higher than the 74%, with a peak of the 95% for the no-aid class of the fine-tuned ResNet50 (the best performing model). The F1 score combines precision and recall into a single metric describing the model performance.

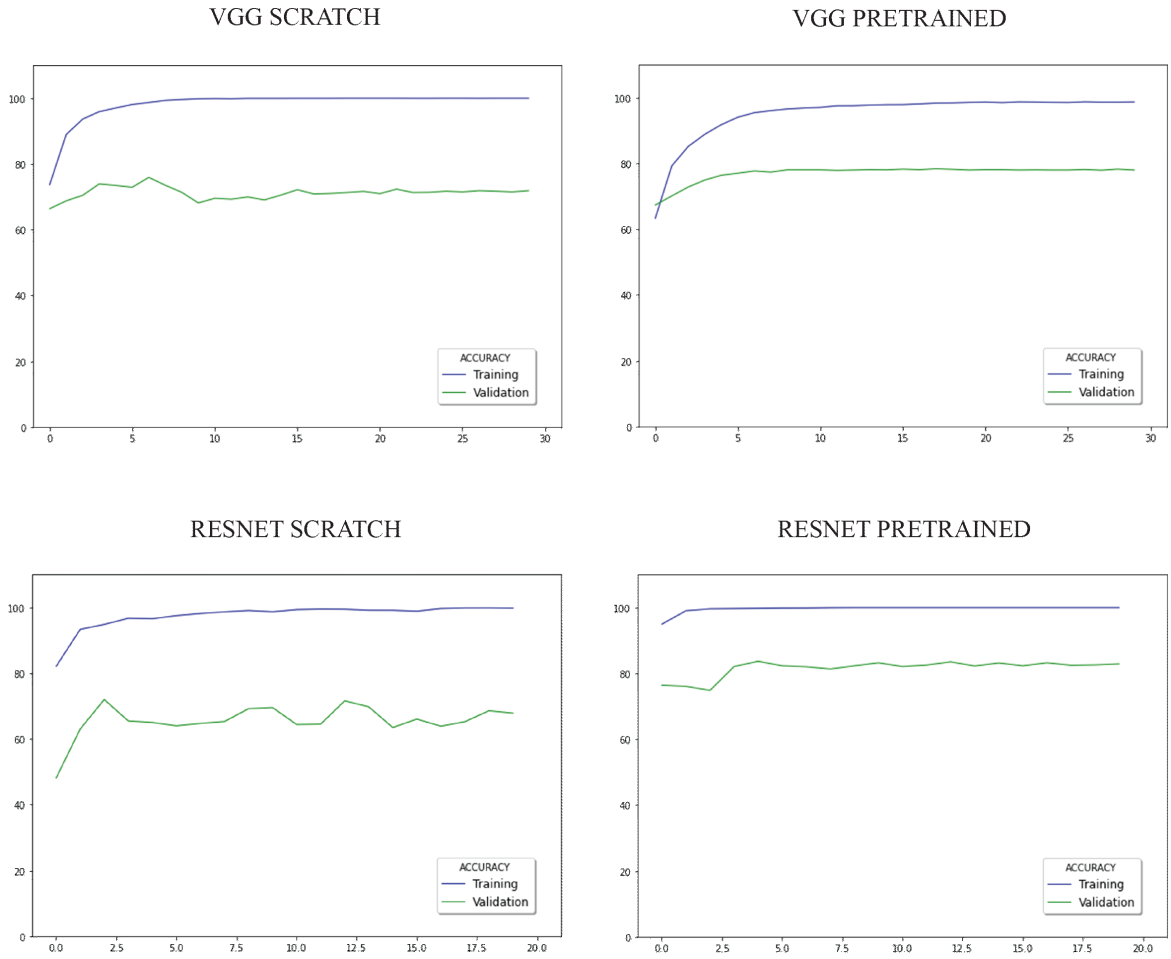

Figure 4.2: *Accuracy in training and validation*

It determines if a change, in training or of the model, results in an improvement or not. Moving from the model having the worst accuracy to the model showing the best accuracy, the perclass f1-score increases at each step (new model), with the only exception of the aid class passing from the pretrained VGG16 to the from scratch ResNet50 in which it passes from the 82% to the 81%. This is a negligible difference considering the good training results defined as having a positive and increasing F1 score. The confusion matrices of the models are shown in Fig. 4.3. In each confusion matrix are reported the percentages of true positives (TP), false negatives (FN), false positives (FP) and true negatives (TN), reading in clockwise fashion from the first quadrant of the matrix.

Besides the AUC for each label of each model, in Tab. 4.1 are reported the macro-average AUCs (mean of the perclass AUCs) and the μ-average AUCs (mean of the

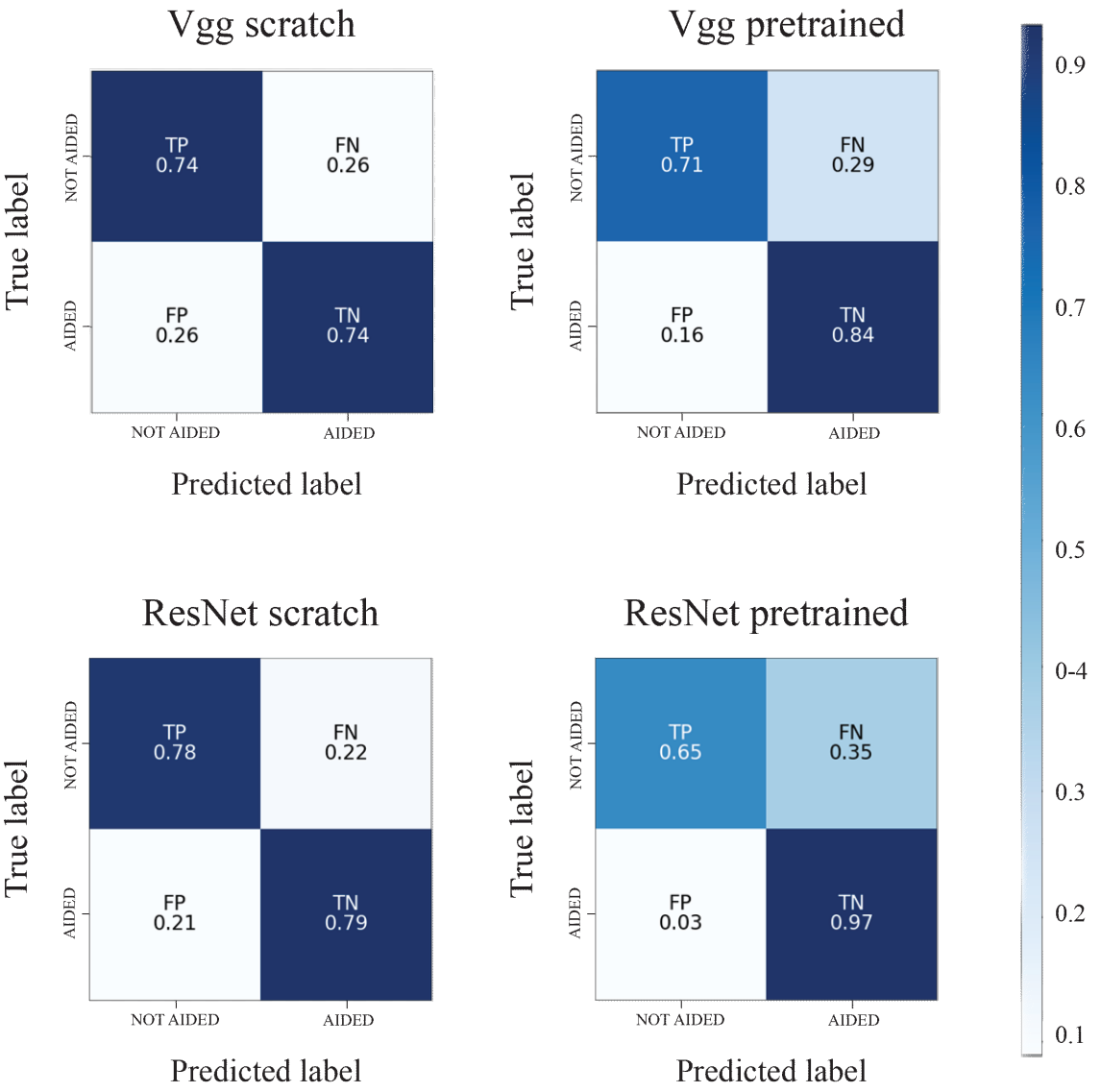

Figure 4.3: *Confusion matrices of the 4 models used. The upper matrices refer to the VGG16 from scratch (left) and pretrained (right), the lower ones to the ResNet from scratch (left) and pretrained (right)*

perclass AUCs, weighted by the number of the ground-truth samples of each class). This last, takes into account the label imbalance and confirms that the dataset is built ensuring the balance in the number of samples between the 2 classes.

In Fig. 4.4 the ROC curves with the corresponding AUC of the four models are reported.

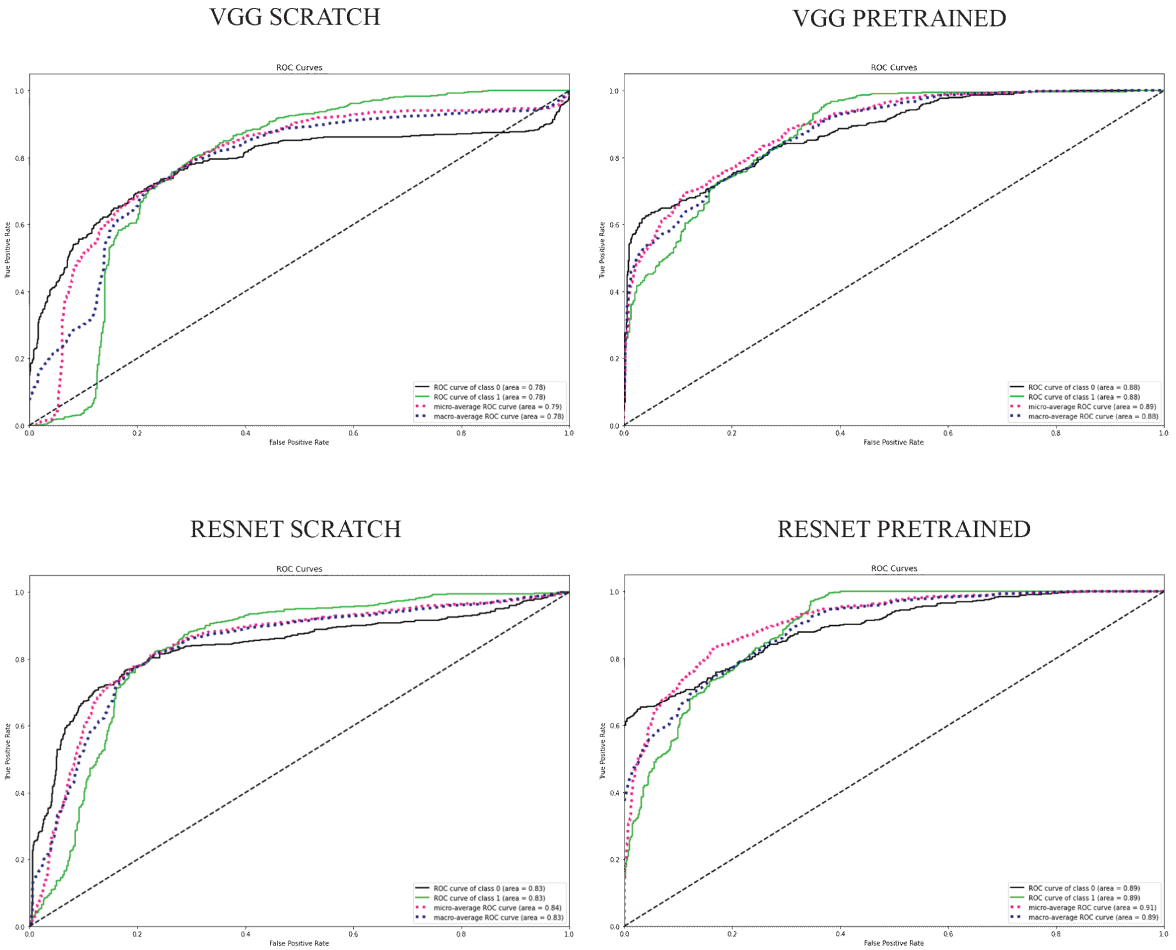

Figure 4.4: *ROC curves of the 4 used models and resulting AUCs. Upper graphs are for the VGG16 model, the lower ones for the ResNet50*

In the following is reported a series of Grad-Cam output images of the fine-tuned ResNet50, reporting the heatmap and the perclass classification probabilities.

Figure 4.5: *Grad-Cam heatmap, ground truth is "No aid". The output perclass probabilities are: 0-No aid (Pr=0.96713) 1-Aid (Pr=0.03287).*

Figure 4.6: *Grad-Cam heatmap, ground truth is "Aid". The output perclass probabilities are: 0-No aid (Pr=0.00080) 1-Aid (Pr=0.99920).*

Figure 4.7: *Grad-Cam heatmaps showing a too large camera field of view. Ground truth is "Aid". The output perclass probabilities are: 0-No aid (Pr=0.00131) 1-Aid (Pr=Pr=0.99869).*

Figure 4.8: *Grad-Cam heatmap confirms the too large camera field of view. The presence of misleading objects causes errors in the correct prediction. This can be solved focusing the camera on a more restricted region.*

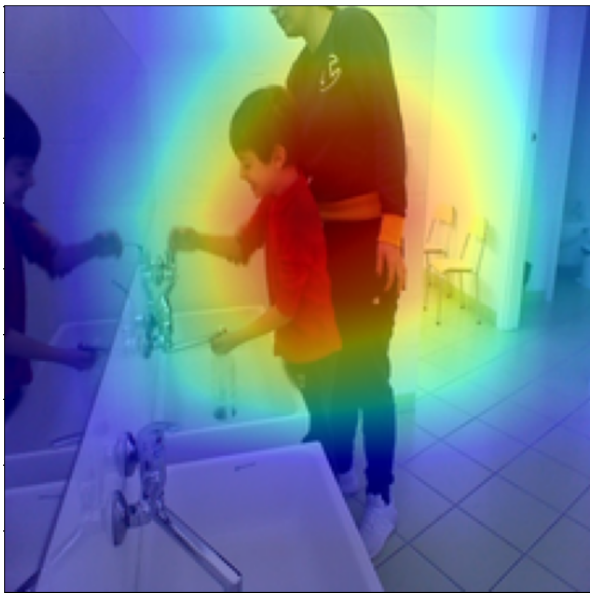

Figure 4.9: *Grad-Cam shows the region chosen by the net to classify a no-aid frame with a probability near the 100%. Ground truth is "No-aid": 0-No aid (Pr=0.99999) 1-Aid (Pr=0.00001).*

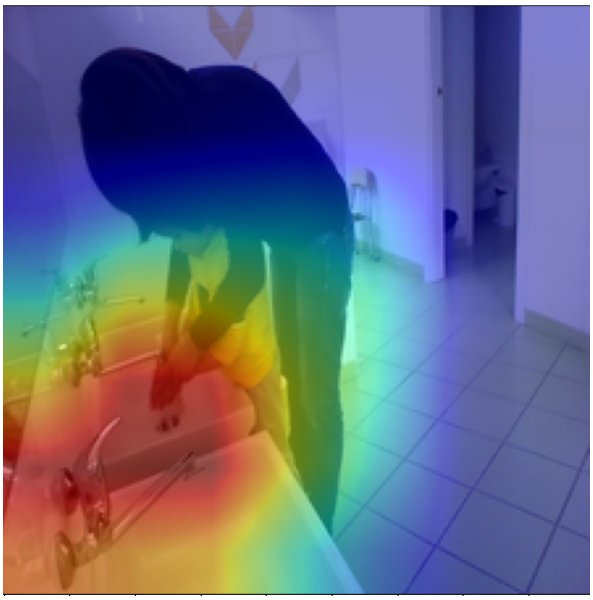

Figure 4.10: *Grad-Cam shows the region chosen to classify an aid frame with a probability near the 100%. Ground truth is "Aid": 1-Aid (Pr=0.99996) 0-No aid (Pr=0.00004).*

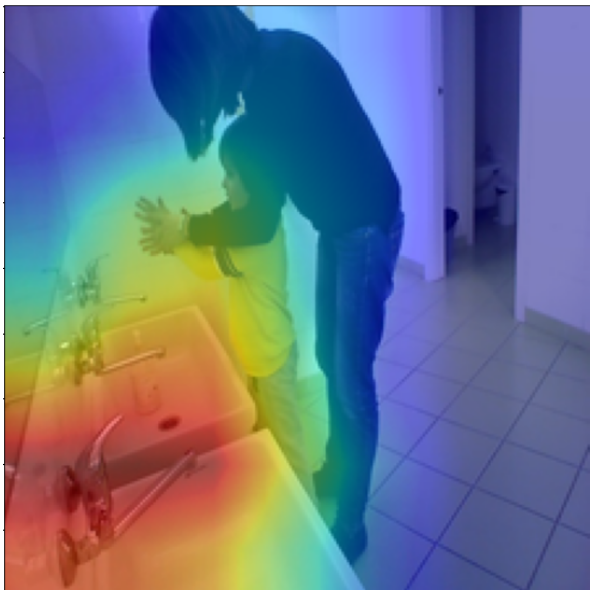

Figure 4.11: *Perfect classification with no error. Grad-Cam shows the region chosen to classify an aid frame with a probability of the 100%. Ground truth is "Aid": 1-Aid (Pr=1.00000) 0-No Aid (Pr=0.00000).*

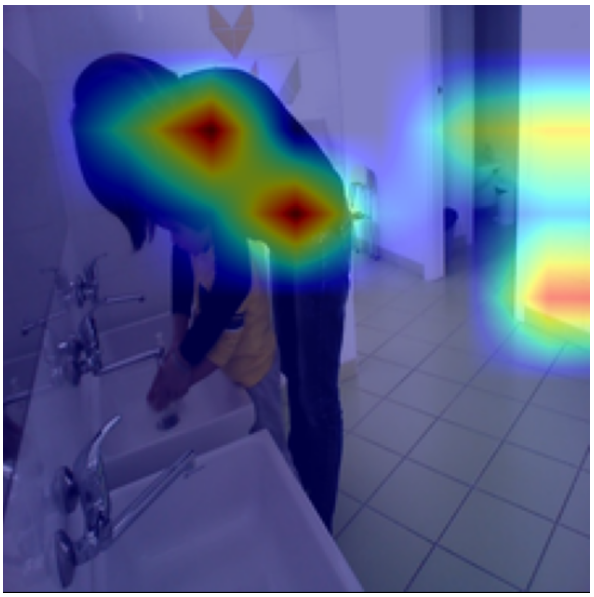

Figure 4.12: *Grad-Cam heatmap shows a frame belonging to the class no-aid. It's clear the importance of the body pose for the operator. To note the high probability value for the correct class (near the 100%) and the low one for the remaining class. Ground truth is "Aid": 1-Aid (Pr=0.99849) 0-No aid (Pr=0.00151).*

Figure 4.13: *The net recognizes the no-aid class even if the pose of operator could suggest the aid class. To note the high probability value for the correct class (near the 100%) and the low one for the remaining class. Ground truth is "No aid": 0-No aid (Pr=0.99941) 1-Aid (Pr=0.00059).*

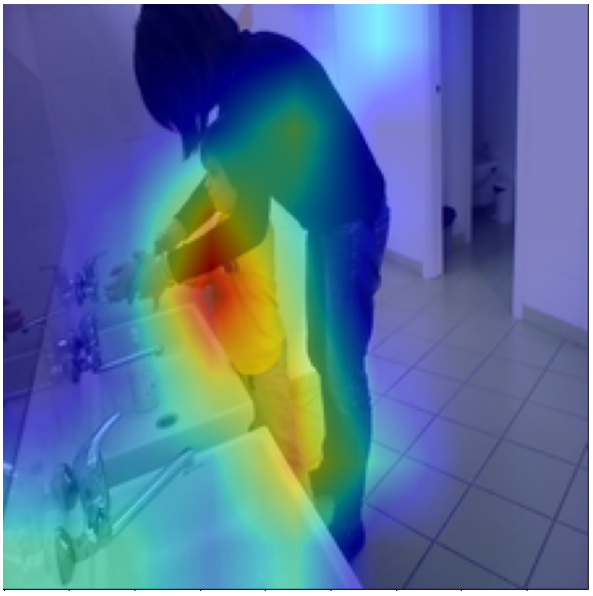

Figure 4.14: *Another perfect classification with no error. It is visible the wide region on which the classification is made with a probability of the 100%. Ground truth is "Aid": 1-Aid (Pr=1.00000) 0-No aid (Pr=0.00000).*

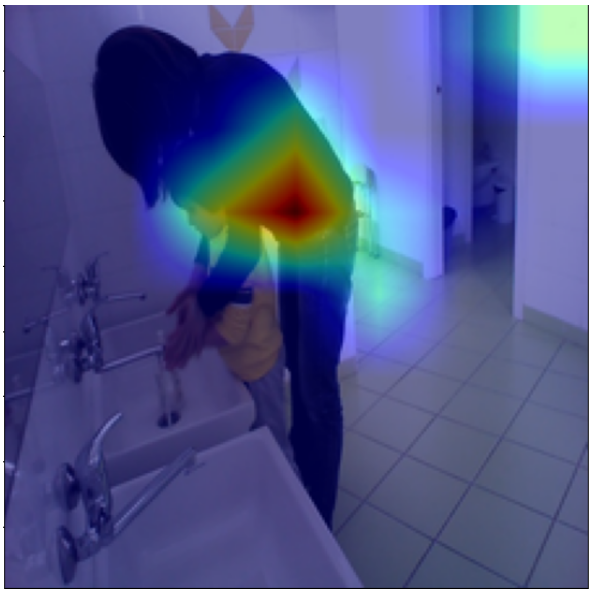

Figure 4.15: *Another correct classification. It is visible the small region on which the classification is made with a probability near the 100%. The heatmap region is limited but considers at the same time: pose of the operator, pose of the child, operator arm position, relative distance between the two subjects. Ground truth is "Aid": 1-Aid (Pr=0.99996) 0-No aid (Pr=0.00004).*

Chapter 5

DISCUSSION AND FUTURE WORK

The ABA therapy is based on experimental behavior analysis with the aim of improving the dysfunctional behaviours of autistic subjects. During the therapy sessions, the operators need to constantly observe the child and take paper-and-pencil rating-scales to evaluate his/her progress and the eventual difficulty in the everyday life tasks. To support the ABA therapists during their actual practice, in this work, the author developed a DL-based application to monitor the children hosted in the structures of "Il Faro" while performing the hand-washing task. By analysing RGB frames, the implemented models detected whether the child accomplishes the task autonomously (no-aid class) or supported by the ABA operator (aid class). Two models were chosen, VGG16 and ResNet50, both trained from scratch and fine-tuned. The results show that the best performance accuracy has been obtained by the pretrained ResNet50 model (83%), followed by the from-scratch version of the same model (79%), then the VGG16 versions in the same order (78% and 74%). The best performance ranking is confirmed by the ROC curves and the corresponding AUCs. This is due to the already acquired knowledge of the pretrained versions on the huge ImageNet dataset. The fine-tuning technique allowed to migrate the knowledge of the training on ImageNet dataset to the presented classification task, improving the networks generalization ability. In particular, the micro-average AUCs highlight the good performances of all the 4 models

in classifying both classes. In fact, being preserved the perclass samples balance, a value of the μ-AUC close to values of the other types of AUC means that the model is proficient in well predicting both classes. Except the VGG16 from scratch model, that shows an equally distributed percentage of false positives and false negatives, the other models report a higher quote of false negatives respect to the false positives. All the 4 models report a higher percentage of true negatives respect to the true positives. Correct predictions of both classes are beyond the 70% in 3 models and only in the pretrained version of the ResNet50, the true positive percentage is of the 65% with the true negative reaching the 97%. In particular, the VGG16 from scratch has an equal distribution of correct prediction (positives and negatives) while the remaining 3 models show a true negatives percentage over the 90%. The perclass metrics of the fine-tuned ResNet50 model require some considerations. The low recall value for the no-aid class reveals a high number of no-aid frames classified as aid. The high precision value for the no-aid class tells that a low number of aid frames are classified as no-aid . In the same way, the high value of the recall for the aid class signals a low number of aid frames classified as no-aid while the lower value of the precision for the aid class with respect to the higher precision value for the no-aid class, reveals a consistent number of no-aid frames classified as aid. All the previous considerations highlight 3 important things:

- The shift between the same metric of the two classes is present even if attention has been paid in balancing the two classes in all the 3 datasets.
- The aid class is well labeled and recognized.
- The presence of a set of challenging frames wrongly annotated as no-aid frames that the net correctly classifies as aid, decreasing the performances of the model. This kind of frames is that in which the arms of the operator and of the child are close enough to mislead the net choice, even if there is no contact.

The problem can be solved annotating as belonging to the autonomy class only the frames in which the operator do not accomplish any movement during the action of the child. Due to the already challenging task, the author decided to not apply any data augmentation strategy because he reputed that the costs of this choice would have

been higher than the benefits in terms of introduced misleading frames. Considering the absence of overfitting in 3 out of 4 models, and the fact that the overfitting is under control in the remaining model, the choice demonstrated to be correct. The best weights per epoch have been saved during the training phase using the lower validation loss as parameter to do it. The trained models, all show good metrics values on the testing data, meaning that the task is well accomplished. This is promising for further applications. Noteworthy is the fact that, despite its deepness, the ResNet50 model has a lower number of trainable parameters respect to the VGG16 model. This reflects in a short time of training. A not negligible aspect in case of comparable metrics values. The ResNet50 model shows an overall best performance. This is probably due to the higher quality of the features extracted exploiting its deepness which, thanks to the residual connections characterizing this model, does not causes the vanishing gradient problem. Increasing the dataset size may improve the model performances. Another possible way to improve the performances of the models could be the use of a larger batch size coupled with a higher computational capability for training. In fact, during the several trials of training, only in 1 (not reported) case it was possible to use a batch size of 128 images, due to the not available memory and to the computational limits imposed by the used Colab environment. In this case, the heatmaps showed a larger area of focusing in the more relevant regions of the images for the task in question. Another noteworthy aspect is represented by the details used by the networks to make the classification choice. In fact, looking at the Grad-CAM heatmaps, it is visible the nets' attention focused on details and patterns not detected by a human expertise. This greatly support the use of deep learning because the identified patterns are strictly related to the task, unbiased by the experience or knowledge of the professional, and provide significant insights to the evaluation process. This is significant considering that the whole process didn't require any previous mathematical modeling or handcrafted features engineering, but it started from real, near untreated data. The Grad-CAM attention maps also reveal that the camera field of view should be reduced to include only one sink and the region where the action is performed. In fact, many classification errors are induced by the presence of the sink near the objective of the camera or by the fact that the net has

focused on the mirror in front of the child or on a door far behind the operator. In some cases, the mirror helped in the right choice but, in other, it confused the net. The elimination of unuseful zones in the image, choosing a restricted region of focusing when tuning the setup, will help to perform a more precise cropping of the image to have, in its center, only the child, the operator and the sink. The time required to make a classification is heavily reduced. A more satisfying evaluation of the ASD condition and treatment outcomes, could be achieved observing different actions or aspects of children behaviour. For example, the pose of both, the child and the operator, revealed to be another index of autonomy. Namely, the distance between the two subjects, but also between them and the wall behind the sink, was another feature that the algorithms used in making the classification. Coupling the images classification task with the identification of other markers of autism could be another good choice. Useful input types to train the networks, in this view, could be, for example, speech analysis data. Interesting could be the development of applications able to automate also the frames selection phase from videos which would make the screening possible also by the child's parents. This would aid the early diagnosis, overcoming the acceptance problem of the parents due to cultural beliefs in many countries.

Chapter 6

CONCLUSIONS

This thesis presented a learning-based application to monitor already diagnosed ASD children, aiming to define if they are autonomous in accomplishing a teached daily live activity such as washing their hands. To the scope, two learning algorithms have been implemented to recognize patient's capacity from frames acquired from a camera. The chosen algorithms showed encouraging results in both the pretrained and from scratch versions. The best results have been shown by the ResNet50 pretrained version. This fact can be explained by the transfer of the knowledge acquired in the training on the million of images of ImageNet dataset. Another reason of the better results is the deeper structure of the net which allows to learn more refined features, solving the problem of vanishing gradient by its short-cut connections. However, to better support the operators in quantifying the progresses achieved by the children who underwent the ABA therapy, further researches are required. There is the need of building larger datasets, by further acquisitions, to improve the performances and achieve a higher generalization capacity. Anyway, the obtained results promote the use of this method in other ABA activities (i.e., the tooth brushing), looking to a more automated way of evaluating the ASD condition. The proposed approach can be used in parallel with a pose-estimation model and a voice pitch analyzer to assess the communication skills. All the computer aided solutions, relevant to the monitoring of the subjects affected by autism, could be included in a single framework. Such a framework will help the operators during their actual practice, especially in lower income countries

where the autistic syndrome is spreading and not faced at the right time to reach the best treatment outcomes.

Chapter 7

MY GRATITUDE

I would thank the Prof. Emanuele Frontoni for giving me the opportunity to work on a such interesting project that could help people with ASD to improve their lives. In the same way i thank Lucia Migliorelli for the proposal to spend my thesis time in the deep learning field. It was rewarding beyond all expectations. Thank also to Sara Moccia and Daniele Berardini for the critical suggestions that have made this a better work. I need to thank my sister Serena Tesei for the images done exactly as i needed. Most of the problems arised during the code implementations have been solved thanks to the big community of passionate in the Python and Latex languages i've found in the web. Their work in sharing the solutions was precious, thank you all. I thank the other students of the Master in Biomedical Engineering of the academic year 2016/2017 for their help and for making these years funnier. Thank to Andrea Tigrini, Luca Pettinari, Lorenzo Marchesini, Stefano Cardarelli, Alessandro Mengarelli. Their dedication to details and to the knowledge was an opportunity for discussion and fun. I am grateful also to ALL the personnel (but really all the personnel) of the Università Politecnica delle Marche that every day works to make this istitution greater and valuable, you made my dreams real. Thank to my friends Luca Giuliani and Eleonora Magnarello, Fabrizio Togni and Lucia Bacci for the quiet saturdays. Thanks to Jerry Capitanelli and Marco David for having tolerated my nervous times and my saturday absences near the exams. A special thank goes to my family which supported me in the busier days divided between my job and the study. In particular to my mom

Giannetta for her saving dinners and to my sisters Sara and Silvia for their help in the hard days. I need to thank my father Costantino for allowing me to combine job and study. I need to thank Francesca Cecato and Augusto Castelli, Carla Panicucci and Tommaso Notaristefano, Federica Pergolesi. Your support in difficult times was and is invaluable. Last but not least, I thank all the people near and far from me, those who supported me in the way they could, even the simplest, this goal is mine as yours. I can't fit all of you here due to space problems (you are too many!), but I embrace you. I wish you the best.

Bibliography

[1] H. Hodges, C. Fealko, and N. Soares, "Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation," *Translational Pediatrics*, vol. 9, no. Suppl 1, p. S55, 2020.

[2] M.-C. Lai and S. Baron-Cohen, "Identifying the lost generation of adults with autism spectrum conditions," *The Lancet Psychiatry*, vol. 2, no. 11, pp. 1013–1027, 2015.

[3] K. Munir, T. Lavelle, D. Helm, D. Thompson, J. Prestt, and M. Azeem, "Autism: a global framework for action," in *Report of the WISH Autism Forum 2016*, 2016.

[4] J. L. Taylor, N. A. Henninger, and M. R. Mailick, "Longitudinal patterns of employment and postsecondary education for adults with autism and average-range iq," *Autism*, vol. 19, no. 7, pp. 785–793, 2015.

[5] D. P. Wall, R. Dally, R. Luyster, J.-Y. Jung, and T. F. DeLuca, "Use of artificial intelligence to shorten the behavioral diagnosis of autism," *PloS one*, vol. 7, no. 8, p. e43855, 2012.

[6] K. S. Omar, P. Mondal, N. S. Khan, M. R. K. Rizvi, and M. N. Islam, "A machine learning approach to predict autism spectrum disorder," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1–6, IEEE, 2019.

[7] D. Bone, S. L. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S. Narayanan, "Use of machine learning to improve autism screening and diagnostic instruments:

effectiveness, efficiency, and multi-instrument fusion," *Journal of Child Psychology and Psychiatry*, vol. 57, no. 8, pp. 927–937, 2016.

[8] B. van den Bekerom, "Using machine learning for detection of autism spectrum disorder," in *Proc. 20th Student Conf. IT*, pp. 1–7, 2017.

[9] W. Liu, M. Li, and L. Yi, "Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework," *Autism Research*, vol. 9, no. 8, pp. 888–898, 2016.

[10] N. M. Rad and C. Furlanello, "Applying deep learning to stereotypical motor movement detection in autism spectrum disorders," in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pp. 1235–1242, IEEE, 2016.

[11] C. Ricci and E. Mattei, "Storia dell'aba in italia: tra miti e false credenze," *Autismo e disturbi dello sviluppo*, vol. 16, no. 3, pp. 327–336, 2018.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[13] M. J. Maenner, C. E. Rice, C. L. Arneson, C. Cunniff, L. A. Schieve, L. A. Carpenter, K. V. N. Braun, R. S. Kirby, A. V. Bakian, and M. S. Durkin, "Potential impact of dsm-5 criteria on autism spectrum disorder prevalence estimates," *JAMA psychiatry*, vol. 71, no. 3, pp. 292–300, 2014.

[14] D. Amendah, S. D. Grosse, G. Peacock, and D. S. Mandell, "The economic costs of autism: A review," *Autism spectrum disorders*, pp. 1347–1360, 2011.

[15] A. V. Buescher, Z. Cidav, M. Knapp, and D. S. Mandell, "Costs of autism spectrum disorders in the united kingdom and the united states," *JAMA pediatrics*, vol. 168, no. 8, pp. 721–728, 2014.

[16] J. P. Leigh and J. Du, "Brief report: Forecasting the economic burden of autism in 2015 and 2025 in the united states," *Journal of autism and developmental disorders*, vol. 45, no. 12, pp. 4135–4139, 2015.

[17] E. B. Robinson, B. St Pourcain, V. Anttila, J. A. Kosmicki, B. Bulik-Sullivan, J. Grove, J. Maller, K. E. Samocha, S. J. Sanders, S. Ripke, *et al.*, "Genetic risk for autism

spectrum disorders and neuropsychiatric variation in the general population," *Nature genetics*, vol. 48, no. 5, pp. 552–555, 2016.

[18] C. DiGuiseppi, S. Hepburn, J. M. Davis, D. J. Fidler, S. Hartway, N. R. Lee, L. Miller, M. Ruttenber, and C. Robinson, "Screening for autism spectrum disorders in children with down syndrome: population prevalence and screening test characteristics," *Journal of developmental and behavioral pediatrics: JDBP*, vol. 31, no. 3, p. 181, 2010.

[19] S. R. Sharma, X. Gonda, and F. I. Tarazi, "Autism spectrum disorder: classification, diagnosis and therapy," *Pharmacology & therapeutics*, vol. 190, pp. 91–104, 2018.

[20] S. H. Baum, R. A. Stevenson, and M. T. Wallace, "Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder," *Progress in neurobiology*, vol. 134, pp. 140–160, 2015.

[21] M. L. Bauman, "Medical comorbidities in autism: challenges to diagnosis and treatment," *Neurotherapeutics*, vol. 7, no. 3, pp. 320–327, 2010.

[22] L. SE, "Schultz rt," *Autism Lancet*, vol. 374, pp. 1627–1638, 2009.

[23] D. L. Christensen, K. V. N. Braun, J. Baio, D. Bilder, J. Charles, J. N. Constantino, J. Daniels, M. S. Durkin, R. T. Fitzgerald, M. Kurzius-Spencer, *et al.*, "Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2012," *MMWR Surveillance Summaries*, vol. 65, no. 13, p. 1, 2018.

[24] B. Zablotsky, L. I. Black, and S. J. Blumberg, "Estimated prevalence of children with diagnosed developmental disabilities in the united states, 2014-2016," 2017.

[25] I. Ertem, G. Atay, D. Dogan, A. Bayhan, B. Bingoler, C. Gok, S. Ozbas, D. Haznedaroglu, and S. Isikli, "Mothers' knowledge of young child development in a developing country," *Child: care, health and development*, vol. 33, no. 6, pp. 728–737, 2007.

[26] P. L. Engle, L. C. Fernald, H. Alderman, J. Behrman, C. O'Gara, A. Yousafzai, M. C. de Mello, M. Hidrobo, N. Ulkuer, I. Ertem, *et al.*, "Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries," *The Lancet*, vol. 378, no. 9799, pp. 1339–1353, 2011.

[27] M. O. Bakare and K. M. Munir, "Excess of non-verbal cases of autism spectrum disorders presenting to orthodox clinical practice in africa–a trend possibly resulting from late diagnosis and intervention," *South African Journal of Psychiatry*, vol. 17, no. 4, pp. 118–120, 2011.

[28] Y. Liu, J. Li, Q. Zheng, C. M. Zaroff, B. J. Hall, X. Li, and Y. Hao, "Knowledge, attitudes, and perceptions of autism spectrum disorder in a stratified sampling of preschool teachers in china," *BMC psychiatry*, vol. 16, no. 1, p. 142, 2016.

[29] L. Dilly, *Autism Spectrum Disorder Assessment in Schools*. 10 2018.

[30] A. L. Reiss, "Childhood developmental disorders: an academic and clinical convergence point for psychiatry, neurology, psychology and pediatrics," *Journal of child psychology and psychiatry*, vol. 50, no. 1-2, pp. 87–98, 2009.

[31] M. S. Durkin, M. Elsabbagh, J. Barbaro, M. Gladstone, F. Happe, R. A. Hoekstra, L.-C. Lee, A. Rattazzi, J. Stapel-Wax, W. L. Stone, *et al.*, "Autism screening and diagnosis in low resource settings: challenges and opportunities to enhance research and services worldwide," *Autism Research*, vol. 8, no. 5, pp. 473–476, 2015.

[32] T. A. Lavelle, M. C. Weinstein, J. P. Newhouse, K. Munir, K. A. Kuhlthau, and L. A. Prosser, "Economic burden of childhood autism spectrum disorders," *Pediatrics*, vol. 133, no. 3, pp. e520–e529, 2014.

[33] M. W. Azeem, I. A. Dogar, S. Shah, M. A. Cheema, A. Asmat, M. Akbar, S. Kousar, and I. I. Haider, "Anxiety and depression among parents of children with intellectual disability in pakistan," *Journal of the Canadian Academy of Child and adolescent Psychiatry*, vol. 22, no. 4, p. 290, 2013.

[34] D. Fein, M. Barton, I.-M. Eigsti, E. Kelley, L. Naigles, R. T. Schultz, M. Stevens, M. Helt, A. Orinstein, M. Rosenthal, *et al.*, "Optimal outcome in individuals with a history of autism," *Journal of child psychology and psychiatry*, vol. 54, no. 2, pp. 195–205, 2013.

[35] A. L. Siu, K. Bibbins-Domingo, D. C. Grossman, L. C. Baumann, K. W. Davidson, M. Ebell, F. A. García, M. Gillman, J. Herzstein, A. R. Kemper, *et al.*, "Screening for autism spectrum disorder in young children: Us preventive services task force recommendation statement," *Jama*, vol. 315, no. 7, pp. 691–696, 2016.

[36] S. E. Levy, E. Giarelli, L.-C. Lee, L. A. Schieve, R. S. Kirby, C. Cunniff, J. Nicholas, J. Reaven, and C. E. Rice, "Autism spectrum disorder and co-occurring developmental, psychiatric, and medical conditions among children in multiple populations of the united states," *Journal of Developmental & Behavioral Pediatrics*, vol. 31, no. 4, pp. 267–275, 2010.

[37] A. C. Stahmer, N. Akshoomoff, and A. B. Cunningham, "Inclusion for toddlers with autism spectrum disorders: The first ten years of a community program," *autism*, vol. 15, no. 5, pp. 625–641, 2011.

[38] M. DeFilippis and K. Wagner, "Treatment of autism spectrum disorder in children and adolescents.," *Psychopharmacology bulletin*, vol. 46 2, pp. 18–41, 2016.

[39] M. G. Aman, C. A. Farmer, J. Hollway, and L. E. Arnold, "Treatment of inattention, overactivity, and impulsiveness in autism spectrum disorders," *Child and adolescent psychiatric clinics of North America*, vol. 17, no. 4, pp. 713–738, 2008.

[40] N. Brondino, L. Fusar-Poli, M. Rocchetti, U. Provenzani, F. Barale, and P. Politi, "Complementary and alternative therapies for autism spectrum disorder," *Evidence-Based Complementary and Alternative Medicine*, vol. 2015, 2015.

[41] M. Bang, S. H. Lee, S.-H. Cho, S. Yu, K. Kim, H. Y. Lu, G. T. Chang, and S. Y. Min, "Herbal medicine treatment for children with autism spectrum disorder: a systematic review," *Evidence-Based Complementary and Alternative Medicine*, vol. 2017, 2017.

[42] Y. Ooi, S. Weng, L. Jang, L. Low, J. Seah, S. Teo, R. Ang, C. Lim, A. Liew, D. Fung, *et al.*, "Omega-3 fatty acids in the management of autism spectrum disorders: findings from an open-label pilot study in singapore," *European journal of clinical nutrition*, vol. 69, no. 8, pp. 969–971, 2015.

[43] A. Horvath, J. Łukasik, and H. Szajewska, "ω-3 fatty acid supplementation does not affect autism spectrum disorder in children: a systematic review and meta-analysis," *The Journal of nutrition*, vol. 147, no. 3, pp. 367–376, 2017.

[44] V. Chaidez, R. L. Hansen, and I. Hertz-Picciotto, "Gastrointestinal problems in children with autism, developmental delays or typical development," *Journal of autism and developmental disorders*, vol. 44, no. 5, pp. 1117–1127, 2014.

[45] A. S. Chan, S. L. Sze, and Y. M. Han, "An intranasal herbal medicine improves executive functions and activates the underlying neural network in children with autism," *Research in Autism Spectrum Disorders*, vol. 8, no. 6, pp. 681–691, 2014.

[46] T. Miyaoka, R. Wake, M. Furuya, K. Liaury, M. Ieda, K. Kawakami, K. Tsuchie, T. Inagaki, and J. Horiguchi, "Yokukansan (tj-54) for treatment of pervasive developmental disorder not otherwise specified and asperger's disorder: a 12-week prospective, open-label study," *BMC psychiatry*, vol. 12, no. 1, p. 215, 2012.

[47] A. Kawicka and B. Regulska-Ilow, "How nutritional status, diet and dietary supplements can affect autism. a review," *Roczniki Państwowego Zakładu Higieny*, vol. 64, no. 1, 2013.

[48] M. Guo, J. Zhu, T. Yang, X. Lai, X. Liu, J. Liu, J. Chen, and T. Li, "Vitamin a improves the symptoms of autism spectrum disorders and decreases 5-hydroxytryptamine (5-ht): a pilot study," *Brain Research Bulletin*, vol. 137, pp. 35–40, 2018.

[49] J. B. Adams, T. Audhya, S. McDonough-Means, R. A. Rubin, D. Quig, E. Geis, E. Gehn, M. Loresto, J. Mitchell, S. Atwood, *et al.*, "Effect of a vitamin/mineral supplement on children and adults with autism," *BMC pediatrics*, vol. 11, no. 1, p. 111, 2011.

[50] K. Saad, A. A. Abdel-Rahman, Y. M. Elserogy, A. A. Al-Atram, J. J. Cannell, G. Bjørklund, M. K. Abdel-Reheim, H. A. Othman, A. A. El-Houfey, N. H. Abd El-Aziz, *et al.*, "Vitamin d status in autism spectrum disorders and the efficacy of vitamin d supplementation in autistic children," *Nutritional neuroscience*, vol. 19, no. 8, pp. 346–351, 2016.

[51] L. A. Mahmood, R. Al Saadi, L. Matthews, *et al.*, "Dietary and antioxidant therapy for autistic children: Does it really work?," *Archives of Medicine and Health Sciences*, vol. 6, no. 1, p. 73, 2018.

[52] Y.-J. Li, J.-J. Ou, Y.-M. Li, and D.-X. Xiang, "Dietary supplement for core symptoms of autism spectrum disorder: Where are we now and where should we go?," *Frontiers in psychiatry*, vol. 8, p. 155, 2017.

[53] R. G. Levy, P. N. Cooper, P. Giri, and J. Weston, "Ketogenic diet and other dietary treatments for epilepsy," *Cochrane database of systematic reviews*, no. 3, 2012.

[54] A. Evangeliou, I. Vlachonikolis, H. Mihailidou, M. Spilioti, A. Skarpalezou, N. Makaronas, A. Prokopiou, P. Christodoulou, G. Liapi-Adamidou, E. Helidonis, *et al.*, "Application of a ketogenic diet in children with autistic behavior: pilot study," *Journal of child neurology*, vol. 18, no. 2, pp. 113–118, 2003.

[55] A. S. Chan, S. L. Sze, Y. M. Han, and M.-c. Cheung, "A chan dietary intervention enhances executive functions and anterior cingulate activity in autism spectrum disorders: a randomized controlled trial," *Evidence-Based Complementary and Alternative Medicine*, vol. 2012, 2012.

[56] E. M. Alissa and G. A. Ferns, "Functional foods and nutraceuticals in the primary prevention of cardiovascular diseases," *Journal of nutrition and metabolism*, vol. 2012, 2012.

[57] S. Coghlan, J. Horder, B. Inkster, M. A. Mendez, D. G. Murphy, and D. J. Nutt, "Gaba system dysfunction in autism and related disorders: from synapse to symptoms," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 9, pp. 2044–2055, 2012.

[58] P. Trombley, M. Horning, and L. Blakemore, "Interactions between carnosine and zinc and copper: implications for neuromodulation and neuroprotection," *BIOCHEMISTRY C/C OF BIOKHIMIIA*, vol. 65, no. 7, pp. 807–816, 2000.

[59] S. Rose, S. Melnyk, O. Pavliv, S. Bai, T. Nick, R. Frye, and S. James, "Evidence of oxidative damage and inflammation associated with low glutathione redox status in the autism brain," *Translational psychiatry*, vol. 2, no. 7, pp. e134–e134, 2012.

[60] E. A. Mayer, D. Padua, and K. Tillisch, "Altered brain-gut axis in autism: comorbidity or causative mechanisms?," *Bioessays*, vol. 36, no. 10, pp. 933–939, 2014.

[61] S. A. Munasinghe, C. Oliff, J. Finn, and J. A. Wray, "Digestive enzyme supplementation for autism spectrum disorders: a double-blind randomized controlled trial," *Journal of autism and developmental disorders*, vol. 40, no. 9, pp. 1131–1138, 2010.

[62] K. Saad, A. A. Eltayeb, I. L. Mohamad, A. A. Al-Atram, Y. Elserogy, G. Bjørklund, A. A. El-Houfey, and B. Nicholson, "A randomized, placebo-controlled trial of digestive enzymes in children with autism spectrum disorders," *Clinical Psychopharmacology and Neuroscience*, vol. 13, no. 2, p. 188, 2015.

[63] A. Gill and C. N. Bell, "Hyperbaric oxygen: its uses, mechanisms of action and outcomes," *Qjm*, vol. 97, no. 7, pp. 385–395, 2004.

[64] D. A. Rossignol, L. W. Rossignol, S. J. James, S. Melnyk, and E. Mumper, "The effects of hyperbaric oxygen therapy on oxidative stress, inflammation, and symptoms in children with autism: an open-label pilot study," *BMC pediatrics*, vol. 7, no. 1, p. 36, 2007.

[65] D. A. Rossignol, L. W. Rossignol, S. Smith, C. Schneider, S. Logerquist, A. Usman, J. Neubrander, E. M. Madren, G. Hintz, B. Grushkin, *et al.*, "Hyperbaric treatment for children with autism: a multicenter, randomized, double-blind, controlled trial," *BMC pediatrics*, vol. 9, no. 1, p. 21, 2009.

[66] M. Sampanthavivat, W. Singkhwa, T. Chaiyakul, S. Karoonyawanich, and H. Ajpru, "Hyperbaric oxygen in the treatment of childhood autism: a randomised controlled trial," *Diving Hyperb Med*, vol. 42, no. 3, pp. 128–33, 2012.

[67] D. Dunleavy and B. A. Thyer, "Is hyperbaric oxygen therapy an effective treatment for autism? a review," *Journal of Adolescent and Family Health*, vol. 6, no. 1, p. 5, 2014.

[68] J. F. Risher and S. N. Amler, "Mercury exposure: evaluation and intervention: the inappropriate use of chelating agents in the diagnosis and treatment of putative mercury poisoning," *Neurotoxicology*, vol. 26, no. 4, pp. 691–699, 2005.

[69] J. B. Adams, M. Baral, E. Geis, J. Mitchell, J. Ingram, A. Hensley, I. Zappia, S. Newmark, E. Gehn, R. A. Rubin, *et al.*, "Safety and efficacy of oral dmsa therapy for children with autism spectrum disorders: Part a-medical results," *BMC Clinical Pharmacology*, vol. 9, no. 1, p. 16, 2009.

[70] D. A. Geier and M. R. Geier, "A clinical trial of combined anti-androgen and anti-heavy metal therapy in autistic disorders," *Neuroendocrinology Letters*, vol. 27, no. 6, p. 833, 2006.

[71] K. E. Bruscia, "Music in the assessment and treatment of echolalia," *Music Therapy*, vol. 2, no. 1, pp. 25–41, 1982.

[72] M. S. Solanki, M. Zafar, and R. Rastogi, "Music as a therapy: role in psychiatry," *Asian Journal of Psychiatry*, vol. 6, no. 3, pp. 193–199, 2013.

[73] A. Warwick and J. Alvin, *Music therapy for the autistic child*. Oxford University Press, 1991.

[74] A. K. Brandt, R. Slevc, and M. Gebrian, "Music and early language acquisition," *Frontiers in psychology*, vol. 3, p. 327, 2012.

[75] M. Sharda, R. Midha, S. Malik, S. Mukerji, and N. C. Singh, "Fronto-temporal connectivity is preserved during sung but not spoken word listening, across the autism spectrum," *Autism Research*, vol. 8, no. 2, pp. 174–186, 2015.

[76] M. Boso, E. Emanuele, V. Minazzi, M. Abbamonte, and P. Politi, "Effect of long-term interactive music therapy on behavior profile and musical skills in young adults with severe autism," *The journal of alternative and complementary medicine*, vol. 13, no. 7, pp. 709–712, 2007.

[77] G. Dawson, S. Rogers, J. Munson, M. Smith, J. Winter, J. Greenson, A. Donaldson, and J. Varley, "Randomized, controlled trial of an intervention for toddlers with autism: the early start denver model," *Pediatrics*, vol. 125, no. 1, pp. e17–e23, 2010.

[78] G. Mahoney and F. Perales, "Relationship-focused early intervention with children with pervasive developmental disorders and other disabilities: A comparative study," *Journal of Developmental & Behavioral Pediatrics*, vol. 26, no. 2, pp. 77–85, 2005.

[79] S. I. Greenspan and S. Wieder, "Developmental patterns and outcomes in infants and children with disorders in relating and communicating: A chart review of 200 cases of children with autistic spectrum diagnoses," *Journal of Developmental and Learning disorders*, vol. 1, pp. 87–142, 1997.

[80] M.-C. Lai, M. V. Lombardo, and S. Baron-Cohen, "Autism," *The Lancet*, vol. 383, no. 9920, pp. 896 – 910, 2014.

[81] K. Liddle, "Implementing the picture exchange communication system (pecs)," *International journal of language & communication disorders*, vol. 36, no. S1, pp. 391–395, 2001.

[82] L. K. Koegel, M. N. Park, and R. L. Koegel, "Using self-management to improve the reciprocal social conversation of children with autism spectrum disorder," *Journal of autism and developmental disorders*, vol. 44, no. 5, pp. 1055–1063, 2014.

[83] L. Green, D. Fein, C. Modahl, C. Feinstein, L. Waterhouse, and M. Morris, "Oxytocin and autistic disorder: alterations in peptide forms," *Biological psychiatry*, vol. 50, no. 8, pp. 609–613, 2001.

[84] S. Y. Lee, A. R. Lee, R. Hwangbo, J. Han, M. Hong, and G. H. Bahn, "Is oxytocin application for autism spectrum disorder evidence-based?," *Experimental neurobiology*, vol. 24, no. 4, pp. 312–324, 2015.

[85] E. Hollander, J. Bartz, W. Chaplin, A. Phillips, J. Sumner, L. Soorya, E. Anagnostou, and S. Wasserman, "Oxytocin increases retention of social cognition in autism," *Biological psychiatry*, vol. 61, no. 4, pp. 498–503, 2007.

[86] A. J. Guastella, S. L. Einfeld, K. M. Gray, N. J. Rinehart, B. J. Tonge, T. J. Lambert, and I. B. Hickie, "Intranasal oxytocin improves emotion recognition for youth with autism spectrum disorders," *Biological psychiatry*, vol. 67, no. 7, pp. 692–694, 2010.

[87] S. Wigham, J. Rodgers, M. South, H. McConachie, and M. Freeston, "The interplay between sensory processing abnormalities, intolerance of uncertainty, anxiety and restricted and repetitive behaviours in autism spectrum disorder," *Journal of Autism and Developmental Disorders*, vol. 45, no. 4, pp. 943–952, 2015.

[88] J. Case-Smith and T. Bryan, "The effects of occupational therapy with sensory integration emphasis on preschool-age children with autism," *American Journal of Occupational Therapy*, vol. 53, no. 5, pp. 489–497, 1999.

[89] L. M. McGarry and F. A. Russo, "Mirroring in dance/movement therapy: Potential mechanisms behind empathy enhancement," *The Arts in Psychotherapy*, vol. 38, no. 3, pp. 178–184, 2011.

[90] C. H. Yau, C. L. Ip, and Y. Y. Chau, "The therapeutic effect of scalp acupuncture on natal autism and regressive autism," *Chinese medicine*, vol. 13, no. 1, p. 30, 2018.

[91] T. Grandin, "Calming effects of deep touch pressure in patients with autistic disorder, college students, and animals," *Journal of child and adolescent psychopharmacology*, vol. 2, no. 1, pp. 63–72, 1992.

[92] C. C. Streeter, J. E. Jensen, R. M. Perlmutter, H. J. Cabral, H. Tian, D. B. Terhune, D. A. Ciraulo, and P. F. Renshaw, "Yoga asana sessions increase brain gaba levels: a

pilot study," *The journal of alternative and complementary medicine*, vol. 13, no. 4, pp. 419–426, 2007.

[93] S. Radhakrishna, "Application of integrated yoga therapy to increase imitation skills in children with autism spectrum disorder," *International journal of yoga*, vol. 3, no. 1, p. 26, 2010.

[94] M. E. O'Haire, S. J. McKenzie, S. McCune, and V. Slaughter, "Effects of classroom animal-assisted activities on social functioning in children with autism spectrum disorder," *The journal of alternative and complementary medicine*, vol. 20, no. 3, pp. 162–168, 2014.

[95] P. M. Barnes, B. Bloom, and R. L. Nahin, "Complementary and alternative medicine use among adults and children; united states, 2007," 2008.

[96] J. K. H. S. L. M. M. M. S. M. U. K. K. M. Vllasaliu, L and C. Freitag, "Diagnostic instruments for autism spectrum disorder (asd)," *Cochrane Database of Systematic Reviews*, no. 1, 2016.

[97] D. Skuse, R. Warrington, D. Bishop, U. Chowdhury, J. Lau, W. Mandy, and M. Place, "The developmental, dimensional and diagnostic interview (3di): a novel computerized assessment for autism spectrum disorders," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 43, no. 5, pp. 548–558, 2004.

[98] C. Lord, M. Rutter, and A. Le Couteur, "Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders," *Journal of autism and developmental disorders*, vol. 24, no. 5, pp. 659–685, 1994.

[99] C. Gillberg, C. Gillberg, M. Råstam, and E. Wentz, "The asperger syndrome (and high-functioning autism) diagnostic interview (asdi): a preliminary study of a new structured clinical interview," *Autism*, vol. 5, no. 1, pp. 57–66, 2001.

[100] J. Maljaars, I. Noens, E. Scholte, and I. van Berckelaer-Onnes, "Evaluation of the criterion and convergent validity of the diagnostic interview for social and communication disorders in young and low-functioning children," *Autism*, vol. 16, no. 5, pp. 487–497, 2012.

[101] L. Wing, S. R. Leekam, S. J. Libby, J. Gould, and M. Larcombe, "The diagnostic interview for social and communication disorders: Background, inter-rater reliability and clinical use," *Journal of child psychology and psychiatry*, vol. 43, no. 3, pp. 307–325, 2002.

[102] J. L. Matson, J. Wilkins, J. A. Boisjoli, and K. R. Smith, "The validity of the autism spectrum disorders-diagnosis for intellectually disabled adults (asd-da)," *Research in developmental disabilities*, vol. 29, no. 6, pp. 537–546, 2008.

[103] I. L. Cohen, V. Sudhalter, D. Landon-Jimenez, and M. Keogh, "A neural network approach to the classification of autism," *Journal of autism and developmental disorders*, vol. 23, no. 3, pp. 443–466, 1993.

[104] S. Baron-Cohen, S. Wheelwright, J. Robinson, and M. Woodbury-Smith, "The adult asperger assessment (aaa): a diagnostic method," *Journal of autism and developmental disorders*, vol. 35, no. 6, p. 807, 2005.

[105] C. Lord, S. Risi, L. Lambrecht, E. H. Cook, B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. Rutter, "The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism," *Journal of autism and developmental disorders*, vol. 30, no. 3, pp. 205–223, 2000.

[106] D. Neal, J. L. Matson, and M. A. Hattier, "Validity of the autism spectrum disorder observation for children (asd-oc)," *Journal of Mental Health Research in Intellectual Disabilities*, vol. 7, no. 1, pp. 14–33, 2014.

[107] B. J. Freeman, E. R. Ritvo, D. Guthrie, P. Schroth, and J. Ball, "The behavior observation scale for autism: Initial methodology, data analysis, and preliminary findings on 89 children," *Journal of the American Academy of Child Psychiatry*, vol. 17, no. 4, pp. 576–588, 1978.

[108] C. A. Vaughan, "Test review: E. schopler, me van bourgondien, gj wellman, & sr love childhood autism rating scale . los angeles, ca: Western psychological services, 2010," *Journal of Psychoeducational Assessment*, vol. 29, no. 5, pp. 489–493, 2011.

[109] M. Taj-Eldin, C. Ryan, B. O'Flynn, and P. Galvin, "A review of wearable solutions for physiological and emotional monitoring for use by people with autism spectrum disorder and their caregivers," *Sensors*, vol. 18, no. 12, p. 4271, 2018.

[110] M. V. Villarejo, B. G. Zapirain, and A. M. Zorrilla, "A stress sensor based on galvanic skin response (gsr) controlled by zigbee," *Sensors*, vol. 12, no. 5, pp. 6075–6101, 2012.

[111] O. Parlak, S. T. Keene, A. Marais, V. F. Curto, and A. Salleo, "Molecularly selective nanoporous membrane-based wearable organic electrochemical device for noninvasive cortisol sensing," *Science advances*, vol. 4, no. 7, p. eaar2904, 2018.

[112] J. Xie, W. Wen, G. Liu, C. Chen, J. Zhang, and H. Liu, "Identifying strong stress and weak stress through blood volume pulse," in *2016 International Conference on Progress in Informatics and Computing (PIC)*, pp. 179–182, IEEE, 2016.

[113] R. Luijcks, H. J. Hermens, L. Bodar, C. J. Vossen, J. van Os, and R. Lousberg, "Experimentally induced stress validated by emg activity," *PloS one*, vol. 9, no. 4, p. e95215, 2014.

[114] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *Journal of biomedical informatics*, vol. 73, pp. 159–170, 2017.

[115] S. Imani, A. J. Bandodkar, A. V. Mohan, R. Kumar, S. Yu, J. Wang, and P. P. Mercier, "A wearable chemical–electrophysiological hybrid biosensing system for real-time health and fitness monitoring," *Nature communications*, vol. 7, no. 1, pp. 1–7, 2016.

[116] S. Yoon, J. K. Sim, and Y.-H. Cho, "A flexible and wearable human stress monitoring patch," *Scientific reports*, vol. 6, p. 23468, 2016.

[117] C. Guo, Y. V. Chen, Z. C. Qian, Y. Ma, H. Dinh, and S. Anasingaraju, "Designing a smart scarf to influence group members' emotions in ambience: design process and user experience," in *International Conference on Universal Access in Human-Computer Interaction*, pp. 392–402, Springer, 2016.

[118] S. H. Koo, K. Gaul, S. Rivera, T. Pan, and D. Fong, "Wearable technology design for autism spectrum disorders," *Archives of Design Research*, vol. 31, no. 1, pp. 37–55, 2018.

[119] A. Serin, N. S. Hageman, and E. Kade, "The therapeutic effect of bilateral alternating stimulation tactile form technology on the stress response," *Journal of Biotechnology and Biomedical Science*, vol. 1, no. 2, p. 42, 2018.

[120] A. Muaremi, B. Arnrich, and G. Tröster, "Towards measuring stress with smartphones and wearable devices during workday and sleep," *BioNanoScience*, vol. 3, no. 2, pp. 172–183, 2013.

[121] S. Yoshimoto, R. Babygirija, A. Dobner, K. Ludwig, and T. Takahashi, "Anti-stress effects of transcutaneous electrical nerve stimulation (tens) on colonic motility in rats," *Digestive diseases and sciences*, vol. 57, no. 5, pp. 1213–1221, 2012.

[122] A. E. Kowallik and S. R. Schweinberger, "Sensor-based technology for social information processing in autism: A review," *Sensors*, vol. 19, no. 21, p. 4787, 2019.

[123] A. B. Dris, A. Alsalman, A. Al-Wabil, and M. Aldosari, "Intelligent gaze-based screening system for autism," in *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, pp. 1–5, IEEE, 2019.

[124] T. W. Frazier, E. W. Klingemier, M. Beukemann, L. Speer, L. Markowitz, S. Parikh, S. Wexberg, K. Giuliano, E. Schulte, C. Delahunty, *et al.*, "Development of an objective autism risk index using remote eye tracking," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 55, no. 4, pp. 301–309, 2016.

[125] A. G. Olivati, F. B. Assumpção Junior, and A. R. N. Misquiatti, "Acoustic analysis of speech intonation pattern of individuals with autism spectrum disorders," in *CoDAS*, vol. 29, SciELO Brasil, 2017.

[126] E. Marchi, B. Schuller, S. Baron-Cohen, A. Lassalle, H. O'Reilly, D. Pigat, O. Golan, S. Fridenson-Hayo, S. Tal, and S. Berggren, "Voice emotion games: Language and emotion in the voice of children with autism spectrum condition," 03 2015.

[127] C. Keenan, A. Thurston, and K. Urbanska, "Video-based interventions for promoting positive social behaviour in children with autism spectrum disorders: a systematic review and meta-analysis," *The Campbell Collaboration*, 2017.

[128] G. S. Young, J. N. Constantino, S. Dvorak, A. Belding, D. Gangi, A. Hill, M. Hill, M. Miller, C. Parikh, A. Schwichtenberg, *et al.*, "A video-based measure to identify autism risk in infancy," *Journal of Child Psychology and Psychiatry*, vol. 61, no. 1, pp. 88–94, 2020.

[129] S. Bellini and J. Akullian, "A meta-analysis of video modeling and video self-modeling interventions for children and adolescents with autism spectrum disorders," *Exceptional children*, vol. 73, no. 3, pp. 264–287, 2007.

[130] K. Dautenhahn and I. Werry, "Towards interactive robots in autism therapy: Background, motivation and challenges," *Pragmatics & Cognition*, vol. 12, no. 1, pp. 1–35, 2004.

[131] L. E. Libero, T. P. DeRamus, A. C. Lahti, G. Deshpande, and R. K. Kana, "Multimodal neuroimaging based classification of autism spectrum disorder using anatomical, neurochemical, and white matter correlates," *Cortex*, vol. 66, pp. 46–59, 2015.

[132] K. K. Hyde, M. N. Novack, N. LaHaye, C. Parlett-Pelleriti, R. Anden, D. R. Dixon, and E. Linstead, "Applications of supervised machine learning in autism spectrum disorder research: a review," *Review Journal of Autism and Developmental Disorders*, vol. 6, no. 2, pp. 128–146, 2019.

[133] L. Xu, X. Geng, X. He, J. Li, and J. Yu, "Prediction in autism by deep learning short-time spontaneous hemodynamic fluctuations," *Frontiers in Neuroscience*, vol. 13, 2019.

[134] J. Yang, M. N. Nguyen, P. P. San, X. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition.," in *Ijcai*, vol. 15, pp. 3995–4001, Citeseer, 2015.

[135] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, "Convolutional neural networks for human activity recognition using mobile sensors," in *6th International Conference on Mobile Computing, Applications and Services*, pp. 197–205, IEEE, 2014.

[136] A. Wijesinghe, P. Samarasinghe, S. Seneviratne, P. Yogarajah, and K. Pulasinghe, "Machine learning based automated speech dialog analysis of autistic children," in *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*, pp. 1–5, IEEE, 2019.

[137] O. Rudovic, Y. Utsumi, J. Lee, J. Hernandez, E. C. Ferrer, B. Schuller, and R. W. Picard, "Culturenet: A deep learning approach for engagement intensity estimation from face images of children with autism," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 339–346, IEEE, 2018.

[138] J. Matson, M. Matheis, C. Burns, G. Esposito, P. Venuti, E. Pisula, A. Misiak, E. Kalyva, V. Tsakiris, Y. Kamio, *et al.*, "Examining cross-cultural differences in autism spectrum disorder: a multinational comparison from greece, italy, japan, poland, and the united states," *European Psychiatry*, vol. 42, pp. 70–76, 2017.

[139] F. Chollet, "Deep learning with python," 2018.

[140] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT press Cambridge, 2016.

[141] T. M. Mitchell and M. Learning, "Mcgraw-hill science," *Engineering/Math*, vol. 1, p. 27, 1997.

[142] E. Stevens, L. Antiga, and T. Viehmann, *Deep Learning with PyTorch.* Manning Publications, 2020.

[143] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.

[144] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[145] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[146] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.